QUALITY MODELLING OF WEB PORTALS

Nitasha Bansal, Dr Harish Mittal
Vaish College of Engineering, India.

ABSTRACT
Web Portals appeared in the late 1990s as a new type of Internet website architecture specifically designed to provide personalized online services. The success of Web portals depends on their ability to provide accurate content and useful services specifically tailored to individual users according to their requirements. Data Quality is a critical issue in today's interconnected society. From the software point of view there is a widely accepted standard proposed by ISO/IEC (the ISO/IEC 9126) which proposes a quality model for software products. However, until now a similar proposal for data quality of web portals has not existed. In order to assess the success of a website, a quality model highlighting relevant properties of the website and specifying how to measure them is needed.In this paper a brief overview of some proposals of web portal quality is presented.

Keyword
Data Quality, Information quality, web portal.

1. INTRODUCTION
There are many different definitions of what quality is and what it represents, depending what services it pertains to. Reeves and Bednar (1994) suggest that quality has different root definitions like quality is excellence,quality is value, Quality is Conformance to Specifications, Quality is Meeting and /or Exceeding Customers' Expectations etc. There are those however that contest and argue that quality means that there is no compromising with anything second rate. (Tuchman, 1980). The DQ(Data Quality) is often defined as "fitness for use", i.e., the ability of a data collection to meet user requirements (Strong et al., 1997; Cappiello et al., 2004), and in the DQ/IQ(Data Quality/Information Quality) literature several frameworks providing categories and dimensions as a way of facing DQ/IQ problems. A Web portal is defined as "a Web site or service that offers a broad array of resources and services, such as e-mail, forums, search engines, and on-line shopping malls (Webopedia, 2005)". According to (Marshak and Seybold, 2003) portals can be divided into generations like first generation,second generation, other generations. The quality of a website is a property that is difficult to define and capture in an operational way, yet everybody feels it when it is missing. In fact, for a website there can be as many views of its quality as there are usages. According to Giorgio Brajnik(2001), quality may depend on task-related factors, performance-related factors, development-related factors. The life cycle of a website is determined by the processes of analysis, design, implementation, validation and maintenance involving a variety of persons, resources, methods and tools. When these processes are not based on well defined frameworks, it is likely that they will be neither effective nor efficient, leading to products whose success is difficult to achieve.

2. DQ/IQ FRAMEWORKS ON WEB PORTALS
The notion of DQ has been widely studied in literature and is commonly approached as a multi-dimensional concept (Wang and Strong 1996; Redman 2000; Cappiello et al. 2004; Gertz et al. 2004). We can, furthermore, observe that various DQ attributes have been proposed, according to an author's philosophical view-point (Knight and Burn 2005). With the idea of taking advantage of work already carried out and applying it to Web portals, we also decided to recompile DQ attributes proposed in literature for Web and/or the context of Web portals. Works for different domains in the Web context were selected. Among these are: data integration (Naumann and Rolker 2000; Bouzeghoub and Peralta 2004), e-commerce (Katerattanakul and Siau 2001), Web information portals (Yang et al. 2004), cooperative e-services (Fugini et al. 2002), decision making (Graefe 2003), organizational networks (Melkas 2004) and data quality on the Web (Katerattanakul and Siau 1999; Eppler et al. 2003; Gertz et al. 2004; Moustakis et al. 2004).As a result of this review, it was possible to define a basic set of one hundred DQ attributes proposed for different domains in the Web. Table shows the research works used as sources of Web DQ attributes; the author, the Web domain and the number of DQ attributes obtained from the model/framework are shown for each of them.

Table: Research Works Used As Source of Web Data Quality Attributes

Autor	Domain	No. of DQ attributes obtained from the model/framework
Katerattanakul and Siau (1999)	Personal web sites	6 DQ attributes
Katerattanakul and Siau (2001)	e-Commerce	
Naumann and Rolker (2000)	Data integration	22 DQ attributes
Pernici and Scannapieco (2002)	Web information systems (data evolution)	4 DQ attributes
Fugini et al. (2002)	e-Service cooperative	8 DQ attributes
Graefe (2003)	Decision making	8 DQ attributes
Eppler et al. (2003)	Web sites	16 DQ attributes
Gertz et al. (2004)	DQ on the web	5 DQ attributes
Moustakis et al. (2004)	Web sites	4 DQ attributes
Melkas (2004)	Organizational networks	20 DQ attributes
Bouzeghoub and Peralta (2004)	Data integration	2 DQ attributes
Yang et al. (2004)	Web information portals	5 DQ attributes

2.1 Proposals of Data Quality dimensions

Although there are many DQ proposals, the most widely approached proposals are discussed below:

In (Wang and Strong ,1996), nearly 179 data quality dimensions have been collected from the user's point of view by means of surveys. Out of those, the authors selected 15 different dimensions. Further, these dimensions are grouped under four different categories such as Intrinsic, Accessibility, Contextual, and Representational.

In (Naumann,2002), the author introduces four different categories of DQ dimensions that play an important role in integrated web information system. Those four categories are as follows: content related, technical, intellectual, instantiation-related criteria. Out of several criteria, some criteria will be more important for a particular application domain than other criteria to attain high quality of data.

In (Bovee et al ,2001), this model has been adapted from Wang and Strong and the attributes have been compared with that model to assess DQ with all its essential dimensions or attributes that determine the quality, in any domain. It consists of four attributes, namely: accessibility, interpretability, relevance, integrity. The attribute Integrity is further classified into four sub attributes: accuracy, completeness, consistency and existence where existence is found absent in many studies.

In (Kahn et al.,2002), this proposal developed a methodology for information quality assessment and improvement called as AIMQ(Assessment and Improvement Quality) . The main objective of this methodology is to assist organizations in achieving high quality of information, so the authors designed a model called PSP/IQ(Product and Service Performance/Information quality) model of what DQ means to information consumers and an assessment instrument for measuring the DQ among the users of different organizations. The model has four quadrants, where the information is considered to be a product or a service, and on whether the improvements can be assessed towards a formal specification or customer expectations. The dimension in the model has been classified into four categories: sound, dependable, useful and usable information.

In (Lee et al ,2002), in this proposal , the authors consider a number of DQ dimensions that can be used to assess the DQ. Although this proposal does not contain any model or framework, the dimensions discussed here are shown to be of particular interest and importance to many organizations. The proposed dimensions are as follows: free of error, completeness, consistency, believability, appropriate amount of data, timeliness and accessibility.

2.2 Need of quality models

A quality model can be used to understand, control, improve a product or a process. For example,

To determine usability problems; or performance bottlenecks

1.To determine a baseline for comparison (for example, determining the current levels of usability of a product before moving to the subsequent release)

2. To assess the progress (for example, benchmarking the adoption of a new technology in the product)

3. To predict certain attributes from others (for example, predicting reliability -- mean-time-to-failure -- from complexity).

DeMarco's [1982] statement in the context of (early 80's) software engineering applies equally well to nowadays website engineering: "you cannot control what you cannot measure."

2.3 Quality models

Different quality models are discussed below:

In (GQM: Goal-Question-Metric - Basili and Weiss ,1984),this is a useful framework to guide the definition of a quality model which is based on three steps: list the major goals of the development or maintenance process ,derive from each goal the questions that need to be answered in order to determine if the goals have been met, decide what must be measured in order to answer the questions and how. The goal and questions determine the quality factors that are more important and those that should be discarded.

In (SERVQUAL: Service Quality - Parasuraman et al. ,1998), This model contains five dimensions and 22 items used to measure the different elements of service quality across a broad spectrum of services. The SERVQUAL model is composed of five dimensions: tangibles, reliability, responsiveness, assurance and empathy.

In (PQM: Portal Quality Model – Moraga et al. ,2004), this model has been made using as a basis the SERVQUAL model and the GQM model. This model adapted the different dimensions of the SERVQUAL model to the portal context and split some of them up into sub-dimensions, in order to create a more specific model. As a final result, the dimensions identified for the PQM model were: tangibles, reliability, responsiveness, assurance , empathy and data quality.

In (An instrument to measure user perceived service quality – Yang et al. ,2004), The objective of Yang et al (2004) is to develop and validate an instrument to measure user-perceived overall service quality of IP Web portals (Information Presenting Web portal). This information is useful for researchers and for portal managers. The following conceptual foundations are taken into account: information quality (IQ) and system quality (SQ). Under IQ, the dimensions are classified into Usefulness of content, Adequacy of information. Under SQ factors are categorized into four dimensions namely usability,accessibility,privacy/security,interaction.

In (A flexible evaluation framework for web portals – Sampson and Manouselis ,2004), present an evaluation framework for addressing the multiple dimensions of Web portals that can affect user satisfaction. As a first step, the authors defined several dimensions related to the main satisfaction factors, namely: Web portal content, Web portal design, Web portal personalization and Web portal community support. For each one of these, different dimensions were identified.

In (Drivers of Web portal use – Telang and Mukhopadhyay ,2004), they tried to explore how Internet users choose portals.In order to do so, they relied on the cognitive psychology and human computer interaction (HCI) literature, along with marketing literature, in an attempt to understand the drivers of Web portal use. Firstly, the authors affirmed that a successful portal needs users who repeatedly come back to the same portal on a frequent basis for extended periods of use. This is because portal services are given away for free, and users have access to multiple providers. So ,they defined three measures of portal use: repeat use, stickiness, frequency,.Secondly, they developed a conceptual model of portal use made up of three components: model of repeat use, model of stickiness, model of use frequency.

In (PORTAL DATA QUALITY MODEL -PDQM-Caro et al.,2006), PDQM is a data quality model for web portals that evaluates data as a data consumer would and is based on the user's perspective.The development of PDQM was divided into two stages: the theoretical definition and the operational definition of the model. The goal of the theoretical definition was to determine a set of DQ characteristics that are relevant to data consumers when evaluating the DQ in a web portal. A final set of 33 characteristics was selected and empirically validated. To obtain the operational version of PDQM, the characteristics were organized into four categories:intrinsic,operational,contextual,representational. In each category, the characteristics were organized as Bayesian networks(BN), implementing the DQ assessment with a probabilistic approach to represent the user's subjectivity.

In (SPDQM -SQUARE Aligned Portal Data Quality Model-Moraga et al.,2009), In order to eliminate lack of PDQM, a Portal Data Quality model has been created. This model will allow users and developers to know whether a web portal has an appropriate data quality level or not. Based on that, they will decide to use it or not. Even, it will guide the designers to improve it and to obtain more data quality. This model is based on three different items: PDQM , a systematic literature review carried out to identify new data quality attributes and the ISO/IEC 25012 standard.. SQuaRE (Software product Quality Requirements and Evaluation) is a set of International Standards which consists of different divisions .The final set of characteristics and subcharacteristics that we considered relevant to SPDQM was grouped in function of: Categories the same as PDQM, in which the characteristics and subcharacteristics are classified into one of these categories (Intrinsic, Operational, Contextual, Representational).Points of view that consider the classification that the ISO/IEC 25012 defines to classify the DQ characteristics (Inherent, System Dependent).

In(Model of Usability Evaluation of Web Portals Based on the Fuzzy Logic-Hub et al.,2010), The model of fuzzy usability evaluation is a multi-layer process of obtaining a usability score as a result of an evaluation provided by user of the system. The structure of the model is decomposed into three layers, where each of them consists of procedures that need to be executed before moving down to the next layer. The first layer describes the evaluated system, its users and the criteria. The system is more specifically defined as an object that interacts with human being. The key aspect of the accuracy and significance of the fuzzy usability evaluation model is the proper determination of the criteria that affect the usability of the system. The goal of second layer is to perform a usability evaluation with desired amount of users and group of systems. Third layer of the fuzzy usability evaluation model processes obtained results.

3. QUALITY ATTRIBUTES FOR WEB PORTALS

From the studied works, we have identified several quality characteristic. The most considerated are:

Attractiveness	Documentation	Customer Support
Accessibility	Duplicates	Reliability
Accuracy	Ease of Operation	Reputation
Amount of Data	Expiration	Response Time
Applicability	Flexibility	Security
Availability	Interactivity	Specialization
Believability	Interpretability	Timeliness
Completeness	Novelty	Traceability
Concise Representation	Objectivity	Understandability
Consistent Representation	Organization	Validity
Currency	Relevancy	Value added

It is important to point that most of these characteristics are included on international standards of software quality as ISO9126. However, some of them are not considered in the same level in such standards. As a conclusion of all these antecedents we think that there is necessary a general quality framework for data on the web that include all the possible characteristics related to them.

4. CONCLUSION

Web portals are applications which have, over the last decade, established their position as information sources and/or as a means of accessing information. Of course those who look for information by means of these portals need some means by which to ensure that this information is indeed suitable for the use that they require. In other words, they really need to assess the level of the quality of the data obtained. However, within the literature studied we have found no specific proposals for data quality models for Web portals. Conventional approaches are based on crisp set,which are useful for precise data. But data attributes of web portals are imprecise and uncertain .Conventional approaches can't deal with uncertainity. Fuzzy approach allows for representation of imprecise data. A better approach is needed using fuzzy membership function. Fuzzy techniques can manage the vagueness and ambiguity efficiently. Fuzzy Logic is a mathematical tool for dealing with uncertainty and also it provides a technique to deal with imprecision and information granularity. Fuzzy logic deals with fuzzy sets. A fuzzy set or subset is a generalization of an ordinary or crisp set. The domain membership function is fixed, usually the set of real numbers. There are total 11 mf available in matlab.

5. FUTURE SCOPE

As future work, we will present the development of a process through which to obtain of a set of attributes for web portal DQ, which focuses on the data consumer's point of view. The process will have been built by using three basic elements: a set of Web DQ attributes found in the relevant literature, DQ expectations of data consumers on the Internet, and the functionalities which a Web portal may offer its users. Our set is composed of some DQ attributes which, from the data consumer's perspective, can be used to assess the DQ in Web portals. We consider the proposed set of attributes to be an important step towards achieving a DQ model for Web portals, as they can be used in the definition of a quality assessment process. Some case studies will have been carried out to show the validity of these attributes within a concrete domain of Web portals (the bank portal domain), thus demonstrating that the set of DQ attributes is correct and complete.Fuzzy approach will be used to analyze and validate attributes.

6. REFERENCES

[1] M Ángeles Moraga, Coral Calero, and M. Piattini,2004, A First Proposal of a Portal Quality Model, IADIS International Conference.

[2] M. Angélica Caro, Coral Calero, Ismael Caballero, Mario Piattini,2005, DATA QUALITY IN WEB APPLICATIONS:A STATE OF THE ART, IADIS International Conference.

[3] M Ángeles Moraga, Coral Calero,Mario Piattini,2006, COMPARING DIFFERENT QUALITY MODELS FOR PORTALS,Online Information Review,vol 30,iss 5,pp 555-568.

[4] C.A. Calero, C. Caballero, I. Piattini, M., 2007, Towards a Data Quality Model for Web Portals, Springer-Verlag Berlin Heidelberg.

[5] Giorgio Brajnik ,2001, Towards valid quality models for websites, 7th Conference on Human Factors and the Web, Madison, Wisconsin.

[6] Mayte Herrera, Ma Ángeles Moraga, Ismael Caballero ,Coral Calero,2010, Quality in Use Model for Web Portals (QiUWeP) , Current Trends in Web Engineering, Vol 63,iss 85, pp 91-101.

[7] Angélica Caro, Coral Calero, Ismael Caballero and Mario Piattini,2007, Towards a Data Quality Model for Web Portals(Research in Progress),Vol 1, Part II, pp 228-237.

[8] Enrique Herrera-Viedma , Eduardo Peis , Jose´ M. Morales-del-Castillo , Sergio Alonso, Karina Anaya ,2007, A fuzzy linguistic model to evaluate the quality of Web sites that store XML documents, International Journal of Approximate Reasoning, vol 46 , pp 226–253.

[9] Ange´lica Caro , Coral Calero , Ismael Caballero ,Mario Piattini,2008, A proposal for a set of attributes relevant for Web portal data quality, Software Qual J, Springer Science+Business Media, LLC 2008.

[10] Abdullah S. Al-Mudimigh, Zahid Ullah,2011, Effective Implementation of Portals: Best Practice Model, www.ccsenet.org/ijbm, International Journal of Business and Management Vol. 6, No. 2.

[11] Zhilin Yanga, Shaohan Caib, Zheng Zhouc, Nan Zhoua,2004, Development and validation of an instrument to measure user perceived service quality of information presenting Web portals, Information & Management,vol 42 ,pp 575–589

[12] Offutt, A. J., 2002, Quality attributes of web software applications. IEEE Software.Vol. 19(2), pp. 25-32.

[13] Carmen Moraga, M Ángeles Moraga, Angélica Caro, Coral Calero,2009, Modeling the intrinsic quality of web portal data, Online Information Review.

[14] Bovee M., Srivastava R. R., Mak B.R ,2001, A Conceptual Framework and Belief Function Approach to Assessing Overall Information Quality,Proceedings of the 6th International Conference on Information Quality.

[15] Sivanandam S.N.,Sumathi S.,Deepa S.N.,2007,Introduction to fuzzy logic using MATLAB,Springer.

[16]Y.W.Lee, D.M.Strong, B.K.Khan, R.Y.Wang,2002,AIMQ-A methodology for information quality assessment , Information and Management, pp: 133-146

[17] Caro, A., Calero, C. and Piattini, M. ,2007, A Portal Data Quality Model for Users and Developers, In ICIQ2007, The 12th International Conference on Information Quality.

[18] Miloslav Hub, Michal Zatloukal,2010, Model of Usability Evaluation of Web Portals Based on the Fuzzy Logic, WSEAS TRANSACTIONS on INFORMATION SCIENCE and APPLICATIONS, ISSN: 1790-0832, Vol 7, Iss 4, pp-522-531.

[19] D. Dhyani, Ng.W. Keong, S.S. Bhowmick,2002, A survey of Web metrics, ACM Computing Surveys,vol 34(4),pp 469–503.

[20] Moraga C., Moraga M., Calero C., Caro A. ,2009,SQuaRE-Aligned Data Quality Model for Web Portals",In 9th International Conference on Quality Software,pp 117-122.