

An Introspection of Web Portals Quality Evaluation

VijaykumarMantri,
Research Schlor, CSE Dept,
K.L. University,
Vijayawada A.P. India

Dr. S. Kalaimagal,
Professor, CSE Dept,
Dean, R&D, K.I.T.S.,
KhammamT.S. India

Dr N.Srinivasu
Professor, CSE Dept,
K.L. University,
Vijayawada A.P. India

Abstract—Today, the internet has become the lifeline of people with loads of enormous & rich applications and information resources available at the click of a mouse. People and Organizations use Web Portals - which gives variety of data from different sources at one place in an easy to use way - to carry out their daily work & to take decisions on various situations. Web portals have become an integral part of industry which is supposed to provide reliable, high quality and useful information as per one's requirements. However the success of a web portal depends on the user satisfaction quotient. In the last decade a lot of research work is going on in the field of assessing and evaluating web portal quality. In our research we propose to use Six Sigma Methodology to develop a quality model that uses Analytic Hierarchical Processing (AHP) that can be used to evaluate the quality of an organization's web portal. To create such a model it becomes necessary that we study the existing models and their drawbacks, so that improvements can be made in our proposed model together with upgraded additions. This paper is a state of the art of existing web portal quality models and their drawbacks with suggested improvements.

Index Terms— data quality, information quality, quality models, users' satisfaction, web portals.

I. INTRODUCTION

We use the World Wide Web in various ways to make communication or information interchange with others. This motivates developers to develop number of applications like web sites, web portals, etc. A web portal is a web site that brings together information from diverse sources in a unified way. Apart from the standard search engine feature, web portals offer other services such as e-mail, news, stock prices, information, databases and entertainment. [13].

In the present scenario, massive number of organizations, people using portals as a gateway to web and ERP site with variety of services along with access to internal data of organization [3], [14]. There are number of ways people use portals like for business, education, technological progress, entertainment, latest happenings, etc. [2].

There are several examples of failures of web portal like boo.com, havenworks.com, flooz.com, kibu.com due to many reasons but few of the reasons are timing, user experience and satisfaction i.e. web portal developers and owners not taken care of quality of web portal in users perceptible.

The quality of data or information present on web portal is very much important factor which plays a major role in success or failure of web portals. Data Quality (DQ) or Information Quality (IQ) are interchangeably used terms

which means "fitness for use" [11], [13] or "ability of a collection of data to meet users' requirements" [12] or in terms of web portal "Quality of data contained in the web portal" [1]

The quality checking as well as controlling of such web portal which provides the information is very important for increasing usefulness and reliability as well as satisfaction of the user of the portal. For eg:- Consider a banking portal which provides the facility for transfer of the money from one account to another. If, while transferring amount if portal stops responding in the middle of transaction then it may create problems like money may be debited from sender account but not credited to receiver account. Also consider the portal which provides online shopping. During shopping if the money is debited from the users account but the transaction fails to complete the procedure. In such situation the quality of web portal to overcome such problems is needed.

In last decade a lot of research work is going on in the field of assessing and evaluating Data Quality for Web Portal. The prominent work in this area is done by M Angeles, Coral Calero, Angelica Caro, Mario Piattini, Carmen Moraga and other researchers. There is lot of work done in other countries but in India still the work on Web Portal Quality is in its starting stage. There are so many models that exist to check and evaluate the quality of web portal. It is important to have a comprehensive model to evaluate the web portal. For defining the comprehensive model along with some standard, we need to study the existing model for their merits and demerits. This paper gives the summary of study of such existing web portal quality models. The rest of paper is organized as follows. Section 2 gives the detailed discussion of existing web portal quality models. In section 3 we summarize these models. The section 4 shows our conclusion as well as future work towards quality of web portals.

II. STATE OF ART – WEB PORTAL DATA QUALITY MODELS

Various types of Web portals available on internet and are used by most of the people daily but much attention is not given by developers towards the quality of web portal. There are very few research papers available on this topic. This section summarizes the different models available for controlling and checking quality of web portals.

A. Data Quality in Web Applications: A State of The Art

M. Angélica Caro et.al introduces data quality and assessment of work going in the area of World Wide Web. Number of researchers done their research work for evaluation of data quality available on different web applications. The systematic study of these works done by M. Angélica Caro et.al is discussed in the paper [8]. The data quality (DQ) mainly classified into four groups using customers' perception. These are Intrinsic, Accessibility, Contextual and Representational which are subdivided into various characteristics. Based on this survey they defined various important features along with percentage in web applications like Accuracy (Accurate), Completeness, Timeliness (Timely), Concise (Concise representation), Consistent (Consistent representation), Currency (Current), Interpretability, Relevance, Secure (Security), Accessibility (Accessible), Amount of data, Availability, Credibility, Objectivity, Reputation, Source Reliability, Traceability (Traceable), etc. Most of these properties are defined in International Standards of Software Quality as ISO9126.

B. A First Proposal of a Portal Quality Model

The quality of data on web portal is important in customers' perceptive. One has to have a case study work to check quality of web portal. Such a case study of Castilla-La Mancha University, Spain's portal given by M^a Ángeles Moraga et.al [1] using various characteristics one need to consider for Portal Quality model. They defined a model that uses iterative method to collect the opinion of different types of users of the portal. This model considers various dimensions for portal quality like Tangible, Reliability, Responsiveness, Assurance and Empathy based on SERQUAL Model (by Parasuraman et al 1998). Along with these parameters M^a Ángeles Moraga et.al defined other dimensions like Intrinsic DQ, Representation DQ, Accessibility DQ and Contextual DQ. For testing these dimensions Questionnaire consisting of 42 questions prepared and given to Castilla-La Mancha portal users to rate them on the scale of 1-5 where 1 means 'Strongly Disagree' and 5 means 'Strongly Agree'. The results of survey are analyzed and proposed the steps & corrective measures need to be taken to improve the quality of web portal [1].

C. Portal Data Quality Model

Angélica Caro et.al [2], [5], [13] introduced Portal Data Quality Model (PDQM) which uses Bayesian Networks for checking quality of web portal. A Bayesian Network is a directed acyclic graph where nodes represent variables (factor) and arcs represent dependence relations between variables. The PDQM model emphasizes on data in users' view as the data users don't control over the data quality as well as need to make sure their requirements are fulfilled. The developers & owners of such web portal should check quality of data to meets its users' requirements. The PDQM is based on three key points – Data Consumer perspective Web data quality attributes and Web portal functionalities. This model proposes mainly two sub models – theoretical definition model and operational model. In the first sub model have four phases which are used for collection and classification of

various Data Quality parameters into matrix for portal quality evaluation. The result of last phase of first sub model is used to convert theoretical model to operational model in second sub model. The second sub model divided into four phases which organizes parameters defines structure and prepares PQDM for DQ evaluation followed by Validation using Bayesian Network. The first phase identifies 41 DQ parameters, out of these 41 attributes 33 attributes are used in PDQM for DQ evaluation in portals in data consumers' perspective. For using these attributes in the BN for PDQM classified into four main categories namely Intrinsic, Operational, Contextual and Representational. The different attributes which comes under these main categories given in following table.

DQ Category	DQ Attributes
Intrinsic: It denotes that data have quality in their own right.	Accuracy, Objectivity, Believability, Reputation, Currency, Duplicates, Expiration, Traceability
Operational: It emphasizes the importance of the role of systems; that is, the system must be accessible but secure to allow the personalization and collaboration among other aspects.	Accessibility, Security, Interactivity, Availability, Customer support, Ease of operation, Response time
Contextual: It highlights the requirement which states that DQ must be considered in the context of the task in hand.	Applicability, Completeness, Flexibility, Novelty, Reliability, Relevancy, Specialization, Timeliness, Validity, Value-Added
Representational: It denotes that the system must present data in such a way as to be interpretable and easy to understand, as well as concisely and consistently represented.	Interpretability, Understandability, Concise Representation, Consistent Representation, Amount of Data, Attractiveness, Documentation, Organization

Based on these attributes are classified a BN graph is prepared by putting these attributes at different levels as shown in fig 1 for PDQM model using relationship between different categories at Level 2 & 3 BN graph obtained by Phase 2.

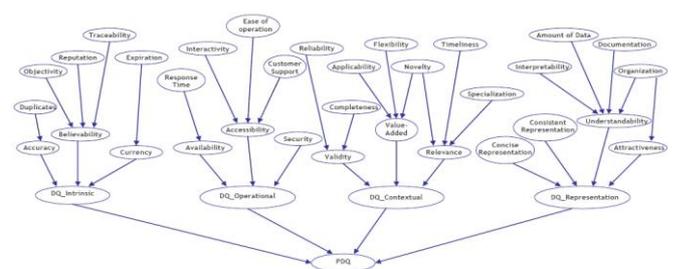


Fig 1 : Graph of the BN that represents the PDQM.

The BN model simplified for studying four fragments separately. The model is validated using three ways like

validating with experts who uses different testing methods, using data collected from experiments where each data point is used as evidence and from experiments used to automatically build another BN. The author(s) used third method for validation. Data was collected by 79 undergraduate students studying in Castilla-La Mancha University who are using web portal as consumers by using a Questionnaire consist of nine questions. The collected data was used as input to BN tool and results are used for analysis. Differences between the BNs were identified and suggested corrections for improvement of portal.

D. Portal Data Quality Assessment Tool (PoDQA)

The PoDQA tool (Portal Data Quality Assessment Tool) which is used to implement PDQM [5], [11] is a 3-tier architecture tool. The main purpose of this tool is to estimate level of DQ for specified Web portal domain and finding ranking for DQ. This tool is based on different previous methods and work done in previous paper. By using definition of PDQM different indicators are obtained. These are Level of Consistent Representation (LCsR), Level of Concise Representation (LCcR), Level of Documentation (LD), Level of Amount of Data (LAD), Level of Interpretability (LI), Level of Organization (LO) and the consistent combinations of these various components.

Using these indicators a tool is developed that can be used by customers to evaluate DQ for given portal. The tool developed is only for representational DQ and cannot be used for real time because we need to download & analyze all pages of portal. This tool can be used on different domains as DQ is based on domain of portal and uses probability tables for each domain. This tool can be used for finding DQ and also gives suggestions to improve web portal DQ. The tool uses 3-tier architecture to separate presentation, application and storage components. The presentation layer gives the user interface for evaluation process. The storage layer consists of a database which is used to store analyzed and evaluated results. The application layer calculates the indicators, stores the results and generate input for second sub-application which executes appropriate BN. The tool calculates the values for different indicators LCsR, LCcR, LD, LAD, LI, LO which takes values between 0 & 1. These values are converted into probabilities and used in BN. This evaluation gives the level of DQ and suggestions for improvements.

The results of DQ calculated using PoDQA were compared with manual results and refined the tool to get better results which matches with manual process by adjusting node probability tables. But this method is not efficient as it takes longer time for evaluation. The limitation of this tool is that it is useful only for Representational DQ & not for other DQs.

E. SQuaRE Portal Data Quality Model (SPDQM)

The SPDQM (SQuaRE Portal Data Quality Model) [4], [7], [12], is a data quality model which is based PDQM (Portal Data Quality Model) and the ISO/IEC 25012, the Data Quality standard, part of the SQuaRE (Software product Quality Requirements and Evaluation) family. The ISO/IEC 25012 defines set of general characteristics for data quality. The quality model uses ISO/IEC 25012 standards specifically

for web portal based on PDQM, which has 33 DQ characteristics. The ISO/IEC 25012 defines 15 characteristics into different category like Inherent data quality (Accuracy, Completeness, Consistency, Credibility and Currentness), System dependent data quality (Availability, Portability and Recoverability) and those which come under both (Accessibility, Compliance, Confidentiality, Efficiency, Precision, Traceability and Understandability).

The SPDQM model consists of five steps. In the first step the DQ characteristics are identified which consists of 33 from PDQM, 15 from ISO/IEC 25012 and 39 characteristics from survey. Once these characteristics identified these were further refined to make sure all are appropriate for web portal and there is no ambiguity or conflict between these characteristics. After refining the SPDQM developed using 30 characteristics from PDQM, 5 from survey and 7 from ISO/IEC 25012 to a total 42 DQ characteristics. These 42 characteristics were classified into four categories as

Intrinsic: Data that have quality by their own right

Operational: Highlights the important role of information systems in areas such as accessibility, security, customization, collaboration, etc.

Contextual: Data quality must be considered in the context of use

Representational: Indicates that the system must present data in a way that is interpretable, easy to understand and is represented concisely and consistently

F. Quality in Use Model for Web Portals (QiUWeP)

A quality model based on ISO/IEC 25010 standard part of SQuaRE (Software product Quality Requirements and Evaluation) family developed to assess quality in use of web portals using users' opinion [6]. The software product quality model is used to find quality of product using internal and external measures. The ISO/IEC 25010 defines three major attributes usability, safety and flexibility which are used in QiUWeP model. The usability defines the use of product that meets users need with effectiveness, efficiency and satisfaction. The property safety defines the loss to people, business, data, software or environment in context of use. This may cause due to invalid or incomplete input. The flexibility defines the way portal can be used by different users with different needs, culture, etc. The flexibility means making changes in the portal according to users' need. The satisfaction of web portal user is the main property to verify the quality of web applications. The model proposed for web portal quality evaluation along with characteristics and sub characteristics is shown in fig 2.

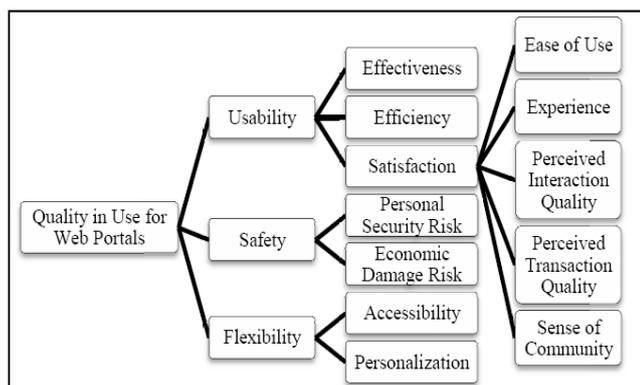


Fig. 2 : Quality in use model for Web Portals

G. Website Quality Assessment using AHP

The various web portal quality models use different technique for evaluation of quality of portal. The Vassilis S. Moustakis et.al [10] developed the model for testing quality of website by using Analytical Hierarchy Process (AHP) to enhance the weightage of each parameter which plays very important role in website quality. The main parameters used as starting point are Content, Navigation, Design and structure, Appearance and multimedia, Uniqueness. These attributes further divided into sub criteria to capture users’ perception and used AHP for quality assessment of these criteria along with their sub criteria. The authors used these criteria and sub criteria (fig 3) to analyze web site quality from web users on three different cellular phone companies’ sites. The model uses AHP to analyze complex issue by dividing into smaller parts in hierarchical manner where top level indicates the overall purpose of the task, the criteria are mentioned in the next level and last level indicates the alternatives. The attributes at particular level are assigned the priority weight and compared with each other to check their impact on overall system.

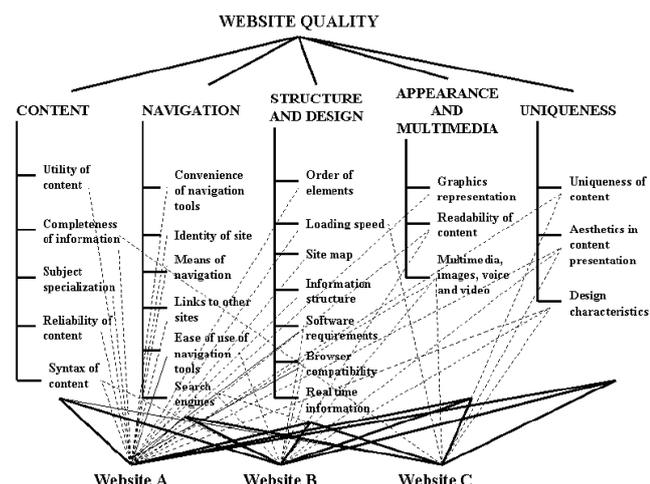


Fig 3 : Criterion and sub-criterion assessment space.

The weight of criteria and sub criteria and website preferences are taken from participant of survey. The author has done case study on three different web sites and shown the various results obtained in survey by 122 participants. The authors found nine criteria with weightage of each are Relevance (18%), Usefulness (15%), Reliability (13%), Specialization

(9%), Architecture (9%), Navigability (7%), Efficiency (6%), Layout (5%) and Animation (3%) which are important for quality evaluation of web site.

H. Quality of Web Portal using AHP

TheMingjingGuo and Yang Zhao [15] described importance & method of information service quality evaluation for high-tech information centers web portal (HIIC web portal) which provides information services for high-tech enterprises by using an Analytical Hierarchy Process (AHP) approach. The quality model use hierarchical structure evaluation criteria model by assigning global priority weight for each criterion assigned. This model used for evaluating the information service quality of ten primary high-tech industry information center web portals in China using comprehensive score method which is useful managers who want to enhance their web portal’s service performance. The results of evaluation using this model gave suggestions for high-tech industry information centers to improve their web portals qualities and to provide better information services for users. The model consists of five major criteria for measuring the information service quality of HIIC web portals along with sub criteria. The major criteria are usefulness of content, adequacy of information, specialization, ease of use and interaction. The AHP model shows the relationship between different factors more clearly. As discussed above the common AHP process involves three phases: construction of a hierarchical structure of the AHP model to present the problem, performing pairwise comparison of the criteria at the same level and determining their weights and synthesis to obtain the global weights for the criteria. The AHP model for evaluation of HICC web portal with 5 criteria and 19 sub criteria is as shown in fig 4.

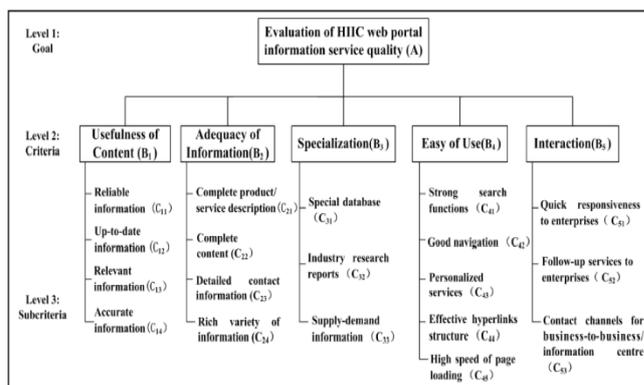


Fig 4 : The evaluation criteria model.

For all these 5 criteria and 19 sub criteria, the weights are assigned by taking geometric mean of weightage collected from experts. A questionnaire of 19 questions is prepared on 5 point scale and shared with participants(CIOs of high tech industry) to evaluate 10 Chinese websites belonging to government departments or industry associations, which are the most authoritative web portals of each high-tech sub-industry. The results were analyzed by using AHP method using global weight of each criterion and given suggestions for improvement of quality.

III. SUMMARY OF COMPARISONS

This paper provides systematic study and summary of existing web portal quality models. We considered various models available on web. These various models proposed and developed by various researchers considered different criteria to evaluate quality of web portals. These models have limitations as these were useful for particular domains. We found these models were using so many attributes which are very difficult to be tested by customers'. In those models many characteristics are irrelevant and many really not useful for web portal quality evaluation. The many attributes add complexity to the model. Many of these models useful only for Web Sites. There are many models which solve the quality related issues to evaluate and give suggestions for improvement of the Web Portal, but still there is uncertainty in quality evaluation by other models. Also these models don't use the techniques like Analytic Hierarchy Process (AHP) or Six Sigma. The AHP gives a structured technique based on mathematics for organizing and analyzing complex decisions using weight for each attribute. It uses hierarchical structure for evaluating different criteria with various alternatives and decision makers systematically evaluate its various elements by comparing them. AHP has been successfully used in qualifying the overall quality of software systems. The Six Sigma is an effective systematic quality improvement approach to enhance the organizations performance based on the adoption of various statistical analytic techniques. Many software organizations have tried to adopt Six Sigma and initiated Six Sigma projects to improve the software development processes from requirement to design, implementation and quality evaluation of end products with the aim to achieve the goal of high customer satisfaction.

IV. CONCLUSION & FUTURE SCOPE

Research Literature elaborates data quality or information quality as a very important factor and is depends on the framework where it is used. There is drastic increase in number of web portals and their uses in day-to-day life by users for important work in last decade. The success of portal depends on number of customers using the portal satisfactorily. The quality of information available on these web portals is very much important. The various models available used different criteria, different techniques for quality evaluation of web portal which are complex in nature. There are few models which are good and also there is ambiguity present in various models available. There is need of good web portal quality evaluation model which is really related to good quality characteristics of web portal.

We propose a web portal quality evaluation model using Six Sigma & Analytical Hierarchy Process (AHP). The Six Sigma follows five phases methodology Define, Measure, Analyze, Design & Validate (DMADV) to develop a quality evaluation model for web portal quality evaluation. The AHP helps decision maker to find the best technique that suits their goal and their understanding of the problem. It provides a comprehensive and rational framework for structuring a decision problem, for representing its qualifying elements, for relating those elements to overall goals, and for evaluating alternative solutions for web portal quality evaluation. Our proposed model is defined using few attributes which plays

very important role in quality of information available on Web portal. The model is comprehensive and easy way of measurement to be used for quality evaluation of Web Portal.

REFERENCES

- [1] M Ángeles Moraga, Coral Calero, Mario Piattini, "A first proposal of a portal quality model", Proceedings of the IADIS International Conference, E-society (Avila, Spain), Berlin: IADIS Press. pp. 630-638, 2004.
- [2] Angélica Caro, Coral Calero, Mario Piattini, "Assessment of Web Portal Data Quality using Bayesian Networks", 33 Conferencia Latinoamericana de Informática (CLEI07), pp. 57-66, 2007.
- [3] TsuangKuo, Iuan-Yuan Lu, Chiung-Hui Huang, Guo-Chiang Wu, "Measuring users' perceived portal service quality: An empirical study", Total quality management and business excellence 16.3, pp309-320, Aug, 2006.
- [4] Carmen Moraga, M Ángeles Moraga, Angélica Caro, Coral Calero, "Modeling the intrinsic quality of web portal data", 13th Workshop on Quantitative Approaches on Object-Oriented Software Engineering and Related Paradigms, Málaga, Spain, pp. 1-12, July 2010.
- [5] Angélica Caro, Coral Calero, Mario Piattini, "A Portal Data Quality Model For Users And Developers", ICIQ, pp. 462-476, Nov 2007.
- [6] Mayte Herrera, M Ángeles Moraga, Ismael Caballero, Coral Calero, "Quality in use model for web portals (QIUWeP)", Springer Berlin Heidelberg, pp. 91-101, July 2010.
- [7] Carmen Moraga, M Ángeles Moraga, Angélica Caro, Coral Calero, "SPDQM: SQuaRE-aligned portal data quality model", Ninth International Conference on Quality Software, QSIC, pp. 24-25, Aug 2009.
- [8] M. Angélica Caro, Coral Calero, Ismael Caballero, Mario Piattini, "Data quality in web applications: A state of the art", IADIS International Conference WWW/Internet. Vol. 2, pp. 364-368, 2005.
- [9] Nitasha Bansal, Dr Harish Mittal, "Quality Modelling of Web Portals", International Journal of Computer Applications (IJCA) proceedings on National Workshop-Cum-Conference on Recent Trends in Mathematics and Computing (RMTC) 2011.
- [10] Vassilis S. Moustakis, Charalambos Litos, Andreas Dalivigas, Loukas Tsironis, "Website Quality Assessment Criteria", Ninth International Conference on Information Quality (ICIQ-04), pp. 59-73, Nov 2004.
- [11] Angélica Caro, Coral Calero, Juan Enriquez de Salamanca, Mario Piattini, "Refinement of a tool to assess the data quality in Web portals", Seventh International Conference on Quality Software - QSIC'07, IEEE, pp. 238-243, Oct 2007.
- [12] Carmen Moraga, M^a Ángeles Moraga, Coral Calero, Angélica Caro, "SQuaRE-aligned data quality model for web portals", Ninth International Conference on Quality Software QSIC'09, IEEE, pp. 117-122, Aug 2009.
- [13] Angélica Caro, Coral Calero, Emilia Mendes, Mario Piattini, "A probabilistic approach to web portal's data quality evaluation", Sixth International Conference on the Quality of Information and Communications Technology QUATIC 2007, IEEE, pp. 143-153, Sept 2007.

- [14] Jung Hyun Nam, "Web portal quality", IEEE/INFORMS International Conference on Service Operations, Logistics and Informatics, pp 163 – 168, 2009.
- [15] Mingjing Guo, Yang Zhao, "AHP-Based Evaluation of Information Service Quality of Information Center Web Portals in High-Tech Industry", International Conference on Management and Service Science MASS'09, IEEE, pp. 1-4, Sept 2009.
- [16] Areej Al-Wabil, Rend Al-Khalifa, "A framework for integrating usability evaluations methods: the Mawhiba web portal case study", International Conference on the Current Trends in Information Technology (CTIT), IEEE, pp. 1-6, Dec 2009.
- [17] Zhang Juan, Tong Weiqin, Cailizhi, "An Evaluation Model in Software Testing Based on AHP", IEEE/ACIS 11th International Conference on Computer and Information Science (ICIS 2012), Shanghai, China, IEEE, pp. 601 – 604, May, 2012.
- [18] S.Christy, S.Brintha Rajakumari, M.Suryakala, "Quality data representation in web portal - A case study", 2nd International Conference on Trendz in Information Sciences & Computing (TISC), Chennai, IEEE, pp. 230 – 232, Dec, 2010.
- [19] Angélica Caro, Coral Calero, M^a Ángeles Moraga, "Are Web Visibility and Data Quality Related Concepts?", IEEE Internet Computing Journal, IEEE Computer Society, pp. 43 – 49, April, 2011.
- [20] Dahlan Nariman, "Evaluating User Expectancy and Satisfaction of e-Government Portals", The Fifth International Conference on Complex, Intelligent and Software Intensive Systems, (CISIS 2011), Seoul, Korea, IEEE CS, pp. 685 – 690, July, 2011.

Khammam, TS, India. Her research interest includes Software Engineering, Software Components and Software Quality and Education Technology.



Dr. N.Srinivasu earned **MCA** from Andhra University, Visakhapatnam, AP, India. **M.Tech.** in Computer Science and Engineering from Acharya Nagarjuna University, Guntur, AP, India. **Ph.D.** in Computer Science and Engineering from Acharya Nagarjuna University, Guntur,

AP, India. Presently working as Professor of Computer Science and Engineering at the K L University, AP India. His active research interests include Cloud Computing, Big Data Analysis, Real time Computation, Distributed Computation, Networks and Data Mining.

Authors Profile



Vijaykumar Mantri received the **B.E.** degree in Computer Science and Engineering from the TPCT's College of Engineering, Osmanabad, MH, India in 1999. **M.Tech** degree in Computer Science and Engineering from the KBN College of Engineering, Gulbarga, Kar,

India in 2008. Currently pursuing **Ph. D.** in Computer Science and Engineering from K.L. University, Vijayawada, AP, India. Presently working as Associate Professor of Information Technology Department at BVRIT, Narsapur, TS, India. His research interest includes Software Engineering, Database Management Systems, Distributed Databases and Unified Modeling Language.



Dr. S. Kalaimagal received her **B.E.** degree in Computer Science and Engineering from Vellore Engg College, Madras University, TN, India in 1997. **M.Tech** degree in Computer Science and Engineering from Pondicherry Central University in the year 1999. **Ph.D.** from

Anna University, Chennai, TN, India in 2010. She has currently nineteen years of teaching and research experience. She is presently working as Professor of Computer Science and Engineering Department and Dean, R&D at K.I.T.S.,