International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:8, 2009

A Framework for Ranking Quality of Information on Weblog

Mohammad Javad Kargar, and Fatemeh Azimzadeh

Abstract**—**The vast amount of information on the World Wide Web is created and published by many different types of providers. Unlike books and journals, most of this information is not subject to editing or peer review by experts. This lack of quality control and the explosion of web sites make the task of finding quality information on the web especially critical. Meanwhile new facilities for producing web pages such as Blogs make this issue more significant because Blogs have simple content management tools enabling non-experts to build easily updatable web diaries or online journals. On the other hand despite a decade of active research in information quality (IQ) there is no framework for measuring information quality on the Blogs yet. This paper presents a novel experimental framework for ranking quality of information on the Weblog. The results of data analysis revealed seven IQ dimensions for the Weblog. For each dimension, variables and related coefficients were calculated so that presented framework is able to assess IQ of Weblogs automatically.

Keywords**—**Information Quality, Weblog, Web Ranking, Web-Quality.

I. Introduction

THE World Wide Web (WWW) has become one of the fastest growing electronic information sources. Meanwhile new facilities for producing web pages such as Blogs have considerably increased the rate. Because Blogs have simple content management tools enabling non-experts to build easily updatable web diaries or online journals. In May 2007, blog search engine Technorati tracking more than 70 million blogs. Every day 120,000 new blogs are created and 1.5 million posts are made, it found during its quarterly survey [28].

Blog became a popular media for publishing information on the internet [3] and has come into the spotlight in the World Wide Web [23]. Ohmukai [23] called these frequently-posted contents as small contents. A vast number of the small contents and citations among Weblog communities are increasing day by day. Some efforts such as topic discovery, trend analysis and content ranking are applied to these large amounts of information.

A Weblog, sometimes written as web log or Weblog, is a Web site that consists of a series of entries arranged in reverse chronological order, often updated on frequently with new information about particular topics. The information can be written by the site owner, gleaned from other Web sites or other sources, or contributed by users. Weblog articles contain various topics such as on personal activities, technology, politics, international problems, and so on. By browsing Weblog articles, we can find frank and up-to-date opinions on various topics such as computer software, poem and literature, social problems, and so on.

Despite a decade of research and practice, only piece meal, ad hoc techniques are available for measuring, analyzing and improving information quality on the web. Unfortunately there is not any framework for measuring IQ in Weblogs. We believed that Weblog can be a suitable application for evaluating quality of information because Weblogs use common templates, so that quality of content of a Weblog is almost equal to quality of Weblog. Weblog owners usually are not computer and information technology specialist. They take advantage of prearranged templates, concentrate on content of Weblog prefer to think about its template and appearance.

In the research we are going to develop a quality of information model for Weblogs. In order to accurately define and measure the concept of information quality, it is not enough to identify the common elements of IQ frameworks as individual entities in their own right. In fact, information quality needs to be assessed within the context of its generation [27] and intended use [13]. This is because the attributes of information quality can vary depending on the context in which the information is to be used.

II. Related Works

Information quality frameworks have been developed over the past few years by various authors in different areas. In one of our earlier works[11], we classified IQ research into four categories; first, literatures which only have listed some of IQ criteria. For instance Collins Memorial Library[26] and Virtual Case [31] have listed some criteria. Second, research which propose information quality models. These models are general purpose or special purpose. In general purpose model criteria are examined in a most general way. In the other word criteria selection and definition is independent of environment and information framework. The aim of such models is that everybody can match the model to their applications. TDQM [32], Naumann [22] and AIMQ [16] are most popular general purpose models.

Unlike general purpose models special purpose models develop the criteria according to their requirements in a

M.J. Kargar is a faculty member in Islamic Azad University- Maybod Branch in Iran (e-mail: showkaran@ hotmail.com).

F. Azimzadeh is a PhD student at University Putra Malaysia (e-mail: f.azimzadeh@gmail.com).

specific application such as Data Warehouse Quality (DWQ) [9], IQIP for information retrieval purposes [14] and intranet application [17], quality of information in Wikipedia [29, 30].The aim of such models mainly hasn't been identifying criteria for information quality. Instead the models have been employed for efficiency improvement in considered application. Third, research which have tackled a few of criteria and have attempted to find methods for computing and measuring the criteria. Measuring timeliness in [34, 35], cohesiveness in [36, 37], frequency analysis in [1, 25] are examples of these works.

Forth, studies which propose frameworks for evaluating the quality of conceptual models. The aim of these researches is to identify worth and validity of information quality models. For instance, in [21] was conducted an empirical analysis of the conceptual model quality framework proposed by Lindland et al [18]. Although literature in information quality proposes several different techniques for measuring information quality, none have addressed the issue of measuring and evaluating information quality in Blogs. There are studies such as [7] which analyzed Blogs and studied Blog comments [15, 20], without entering to information quality issue.

Weblogging has emerged in the past few years as a new grassroots publishing medium. Although some work on analysis of Weblog's components have been pursued, to date no study has specifically addressed the constructing an information quality model for Weblogs. Meanwhile there are some researches which indirectly have studied some of information quality criteria or components of Weblog which influence quality of information.

The first academic research on the Weblogging community appeared WWW conference in 2003 [4]. Since then, the first bloggers' conference has also occured, BloggerCon 2003, allowing bloggers to meet face-to-face, united both by technological interests.

The limited quantitative research on blogs has primarily focused on determining the size and usage of blogspace [19] as well as some explorations on dynamics [15]. Also we established a prioritization of IQ criteria and gap analysis in our previous research [12].

III. Research Design and Methodology

The general aim of the research is to develop a framework for evaluating information quality on the Weblogs. Developing a framework for assessing quality of information is a multi-layer process.

The first phase in methodology was identifying information quality criteria for Weblogs. Evaluating information quality on Web and especially on Weblog requires selecting appropriate criteria. Moreover, criteria selection is one of the most important stages when evaluation of something is intended. In second stage was developed a Weblog server and Weblog management system. The Weblog management system is heart of our framework as system test-bed. The Weblog management system comprises all the software modules and components which are employed in the whole framework.

After implementation of the Weblog management system, information quality modules were added. In the end of this stage the Weblog system is ready for Weblog creation and data entry. Thus participants managed to create their Weblogs and start data entry and other activities in the Weblog environment. The period for data entry and users' activities was two months. In this period, the users were able to post articles, write comments and modify their Weblogs and complete other parts of their Weblogs. At the same time, the data entered by the users and their activities were registered in system database which is located in Weblog management system.

After data entry and Weblog construction stage, all the data were saved into the database system. Since the system had collected information quality scores for each Weblog, the data analysis could be applied on the collected data. The aim of data analysis was finding probable correlation between criteria and sub-criteria. The output of data analysis was calculating information quality scores for each Weblog, finding IQ dimensions in Weblog, and ultimately gaining quality of information for each Weblog.

A. Weblog Management System

Weblog management system is the most important part of the framework. To develop the Weblog management system, it was decided to design a Content Management System (CMS). A content management system is a computer software system for organizing and facilitating collaborative creation of documents and other content. A content management system is a system used to manage the content of a Web site [33].

Many organizations have turned to CMS to publish data with the speed and freedom provided by the Web [24]. Many of modern applications have been developed by CMS. For example, the software for the website Wikipedia is based on a wiki, which is a particular type of content management system [33]. Wiki systems such as wikipedia.org are similar to blogs in principle as they are based on user participation to add content [10].

The current Weblog management system includes several technologies commonly used in the modern web applications. The system was developed by PHP, MySQL, HTML, CSS, JavaScript, and Ajax.

The Weblog management system contains four major components; Administrator control panel, user control panel, IQ modules, system database. Administrator control panel is an interface for system's administrator to control, manage, and monitor the Weblog management system. User control panel is an interface designed for users to produce and manage contents of their Weblogs. This panel provides features which user needs to manage a Weblog such as links management, edit and create template, add new article, manage comments and configuration of Weblog. IQ modules measure information quality for each Blog based on considered IQ criteria. System database records users' activities on the Weblogs.

B. IQ Criteria and Sub-criteria

As mentioned before, 18 sub-criteria could be calculated automatically while 9 qualitative criteria were obtained by

voting were selected for information quality assessment on the Weblog. The 9 qualitative criteria were cohesiveness, concise, believability, understandability, completeness, objectiveness, accuracy, informativeness, and presentation. Subjective measures depend on the environment in which they are made [2]. On the other hand, quality is a matter of perception, and is often difficult to measure objectively. These criteria were obtained by voting. The voting module collects and calculates users' votes for each Weblog. When users intended to leave a comment for a Weblog's post, in addition to writing comments could participate in the voting. There were 9 statements in voting division according to 9 criteria. Users could select scores between 1, as the lowest score, to 9 as the highest score. The results of voting were stored automatically in the system database.

Meanwhile following 18 quantitative sub-criteria were implemented in the system; Last update, last login, meta information checking, ratio of visited links to failed links, number of written comments, Weblog age, number of posted articles, number of external links, number of internal links, initial load time (time for loading first component of Weblog), full load time (time for loading all the components of Weblog), number of received comments, average of received comments (total number of received comments/total number of entries), number of visitors, number of referred links, customer support link, Weblog size, Ratio of multimedia elements to the overall information.

While many of the sub-criteria are common on the Web and are implemented in a common way, some of the sub-criteria depend on the Weblog context such as number of received comments, average of received comments, number of referred links, number of written comments.

C. Data Entry and Weblogs Construction

After developing the Weblog management system and incorporating the information quality criteria to the system, it was time for the Weblog construction and data entry phase. Since a considerable number of participants were needed, a group of computer engineering students in Azad University of Maybod in Iran were asked to create their Weblogs (Persian or Farsi, the official language of Iran, is the newcomer to the top 10 blogging languages [28] on the world). These students had enrolled for either programming languages course, multimedia course, or internet engineering course. A total of 294 students had registered for these three courses. For more contribution, Weblog construction and completion was as a part of assignments that were assigned to the students. As soon as the Weblog system became ready, the system was uploaded with homepage http://www.iranweb3.com, and the students were invited to contribute.

The period of data entry, Weblog construction and Weblogs activity was from 20/11/2007 until 20/1/2008. In the two months period, the students could make their Weblogs, post articles, write comments, add friends, insert links and other activities which are usual in popular Weblog service providers. All the activities were stored in system database. In the two months period, 473 Weblogs were created by the students. The number of Weblogs was more than the number of students because some of the students have made more than one Weblogs supposing that making more Weblogs has more scores. To encourage the students to have more activity, the record of the most active students was frequently updated in terms of some criteria such as the number of visitors, the number of received comments, and the number of friends in the homepage.

IV. Result and Discussion

Real-world data tend to be incomplete, noisy, and inconsistent. Data cleaning routines attempt to fill in missing values, smooth noise while identifying outliers, and correct inconsistencies in the data.

During data collection, students created a total of 473 Weblogs. After verifying the data, 157 Weblogs were found to be without entry. Since Weblogs without entries were void of content, records of these Weblogs were deleted. As a result, the population size was reduced to 316 Weblogs.

The next stage was data cleaning, wherein the data was analyzed in order to find outliers. An outlier is an observation that is numerically distant from the rest of the data. Dirty data can cause confusing for the mining procedure, resulting in unreliable output [6].

Also, outlier analysis revealed that three Weblogs had the highest number of outliers. Because these Weblogs were suspected to have obtained high score in a defrauding way, they were removed. Thus, the final Weblog population size was reduced to 313.

Factor analysis is a multivariate statistical method whose primarily purpose is to define the underlying structure in a data matrix. There are some requirements for deploying factor analysis. The variable for factor analysis should be measured at least at the ordinal level and the sample size should be 100 or larger [8]. Population size for this project was 313; all the data were as of scale type, therefore factor analysis was conducted

There were 18 variables which were measured automatically and 9 variables which were obtained by voting. Thus, 27 variables on Weblog IQ were employed for factor analysis. Principal component analysis was used to reduce the number of variables and detect linear relationships in the 27 variables.

Eigenvalue represents the amount of variance accounted for by a factor. Only factors with eigenvalues ≥ 1 were considered and then corroborated by a scree test (Fig. 1) because according to [5], only factors having eigenvalues greater than 1 are considered significant. A criterion loading of 0.3 was used to determine which scale statements were included in a given factor because according to [5, 8], when sample size is greater than 100, a factor loading greater than .3 is considered to meet the minimum level of significance. Thus, factor loading less than .3 was considered insignificant.

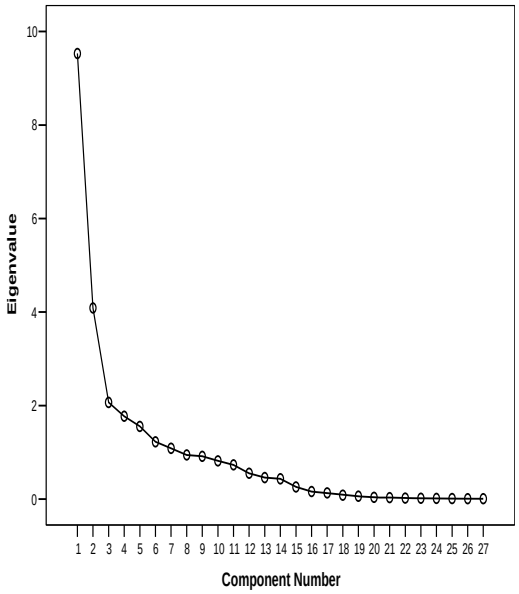

Fig. 1 Screen Plot for Factor Analysis

Considering the above criteria, a total of seven factors were obtained from the principal components, accounting for 79% of the total variance, as shown in Table I.

These seven factors are termed 'subjective score', 'authority', 'link popularity', 'timeliness', 'latency', 'oldness' and 'redundancy'.

The factor loadings for each factor are specified in Table I. Detailed explanations for each factor is as follows: Factor 1 (subjective score) is composed of cohesiveness, concise, believability, understandability, completeness, objectiveness, accuracy, informativeness, and presentation**.** All the subjective criteria which were obtained by voting were loaded on factor 1. Because these criteria tend to be based mostly on the perception of the users, factor 1 was labeled as the subjective score and accounted for 33% of the total variance. Thus,

$$\begin{aligned} Subjective\ Score = {} & .988 * Cohesiveness + .987 * Concise + \\ & .987 * Believability + .985 * Understandability + \\ & .985 * Completeness + .984 * Objectiveness + \\ & .983 * Accuracy + .98 * Informativeness + .977 * Presentation \end{aligned} \quad (1)$$

TABLE I
RESULTS OF FACTOR ANALYSIS

Rotated Component Matrix(a)							
	Component						
Variables	1	2	3	4	5	6	7
Cohesiveness	.988						
Concise	.987						
Believability	.987						
Understandability	.985						
Completeness	.985						
Objectiveness	.984						
Accuracy	.983						
Informativeness	.980						
Presentation	.977						
ReceivedComment		.889					
WrittenComments		.732					
Entries		.619					
Referred		.602					
Visitors		.391					
Links			.935				
VisitedLinks			.925				
Friends			.648				
LastLogin				.760			
LastUpdate				.759			
CommentPerEntry		.323					
Availability				.542			
FirstLoadTime					.952		
FullLoadTime					.938		
Age						.748	
METATag						.647	
MultimediaRate							-.808
WeblogSize							.638
Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization.							
Rotation converged in 9 iterations.							

Factor 2 (authority) had cross-correlation with the following variables: received comments, written comments, visitors, entries, comment per entry and referred. Because these criteria imply the authority of the Weblog, factor 2 was then labeled as authority and accounted for 11% of the total variance. Thus,

$$\begin{aligned} Authority = {} & .889 * RecivedComments + .732 * WrittenComments + \\ & .619 * Entries + .602 * Referred + .391 * Visitors + \\ & .323 * CommentPerEntry \end{aligned} \quad (2)$$

Factor 3 (link popularity) consists of three variables: links, visited links, and friends. This is logical because friends are counted as internal links and links as external links. Visited links show how many people have visited the links. Since the variables point to the popularity related to links, factor 3 was labeled as link popularity and accounted for 9% of the total variance. Thus,

$$Link\ Popularity = .935 * Links + .925 * VisitedLinks + .648 * Friends \quad (3)$$

Factor 4 (timeliness) is composed of last login, last update, and availability. This shows which Weblogs were updated more, had more available links. This is logical because old links may be inactive after a long time, availability is consequently reduced. Factor 4 was labeled timeliness because the variables of the factor were related to currency and accounted for 9% of the total variance. Thus,

$$Timeliness = .76 * LastLogin + .759 * LastUpdate + .542 * Availability \quad (4)$$

Factor 5 (latency) consists of first load time and full load time. Because the variables show latency and response time of Weblogs, factor 4 was entitled latency and accounted for 8% of the total variance. Thus,

$$Latency = .952 * FirstLoadTime + .938 * FullLoadTime \quad (5)$$

Factor 6 (maturity) is composed of age and Meta tag. This may seem abnormal at a glance because age is usually considered as a sub-criterion for authority but the data shows that older Weblogs have more Meta tags than young Weblogs. This is logical because older Weblogs have obtained more experience and understanding on the importance of Meta tags. In effect, older Weblogs have more Meta tags in their Weblogs. The factor was labeled maturity and accounted for 5% of the total variance.

$$Maturity = .647 * MetaTag + .748 * Age \quad (6)$$

Factor 7 consists of multimedia rate and Weblog size. Since multimedia rate and size of Weblog induce redundancy to the Weblogs, the factor was labeled redundancy and accounted for 4% of the total variance. Weblog size has a positive coefficient while multimedia has a negative coefficient because the variables have negative correlation and multimedia rate was calculated as the size of multimedia elements divided by the size of the Weblog. Thus,

$$Redundancy = .638 * Weblog Size - .808 * MultimediaRate \quad (7)$$

There are strong evidences that validate the results of this experiment; High correlation coefficients between many variables, justifiability and logicality of the high correlations show the validity of the results. Results of factor analysis identified seven appropriate factors, as explained in the previous section. In turn, the factor analysis presented interesting new dimensions of information quality on the Weblogs that were obtained by experiment and not merely based on theory. Statistics in the factor analysis show the validity of the experiment. Seven factors extracted by factor analysis covered 79% total variance which is adequately high. In many other experiments, the factors covered less than 70% of the total variance. Also Bartlett test of sphericity, a statistical test for the presence of correlations among the variables [5] was applied. If the test value is large and significant level is small (<0.05), the hypothesis that the variables are independent can be rejected [8]. In the present analysis, the Bartlett test of sphericity yielded a value of 12457 and an associated level of significance less than 0.001. Thus, the hypothesis that the variables are independent is rejected. In other words, the results of factor analysis are valid.

Another measure used to quantify the validity of the factor analysis is the measure of adequacy. This index ranges from 0 to 1, reaching 1 when each variable is perfectly predicted without error by the other variables. If the measure is below .50 it is unacceptable; .6 or above is mediocre; .7 or above is middling and .80 or above is meritorious [5]. Since the overall measure of adequacy for the current research is .857, results of factor analysis can be considered as strongly valid.

V. Conclusion

The aim of this research was to develop a framework for evaluating information quality on Weblogs. In order to develop the framework, appropriate information quality criteria for Weblogs were first identified**.** The next stage included the implementation of the Weblog management system as a test bed of the research. The Weblog management system contained all the facilities for content production on Weblog. Moreover, all the activities carried out by participants, as well as their information quality scores were saved in the system database. After developing the Weblog server, participants were invited to create Weblogs and add contents, in a period of two months. The last step of this research project was data analysis and the calculation of information quality scores for the created Weblogs. The role of each IQ parameter and sub-parameter was analyzed and overall information quality scores for each Weblog were calculated.

One of the key advantages of the current framework is revealing seven dimensions, along with respective coefficient of variables, which is used to evaluate information quality on Weblogs. These dimensions were tested and validated practically, unlike many preceding studies on information quality wherein dimensions were either theoretically selected or based on subjective criteria. Moreover, because of the special nature of Weblogs, three special variables were considered and measured. These variables have not been considered in previous information quality research; namely the number of written comments, number of received comments and comment per entry, all of which were calculated automatically. Interestingly, these three variables fell in the same dimension that was labeled authority.

References

[1] D. Dhyani, W.K. Ng and S.S. Bhowmick, A Survey of Web Metrics, ACM Computing Surveys (CSUR), , 34 (2002) 469–503.
[2] N.E. Fenton and S.L. Pfleeger Software Metrics: A Rigorous & Practical Approach. International Thomson Computer Press, 1997.
[3] T. Fukuhara, T. Murayama and T. Nishida, Analyzing concerns of people from Weblog articles AI & Society, Springer London, 22 (2007) 253-263.
[4] N. Glance, M. Hurst and T. Tomokiyo. BlogPulse: Automated Trend Discovery for Weblogs WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics, New York, 2004.
[5] J.F. Hair, B. Babin, R.E. Anderson and R.L. Tatham Multivariate Data Analysis, Prentice Hall, 2005.
[6] J. Han and M. Kamber Data Mining: Concepts and Techniques. Elsevier Science & Technology, 2006.

[7] S.C. Herring, L.A. Scheidt, S. Bonus and E. Wright, Bridging the Gap: A Genre Analysis of Weblogs, Proceedings of the 37th Annual Hawaii International Conference on System Sciences (HICSS'04), IEEE Computer Society 2004, pp. 101-104
[8] R. Ho Handbook of univariate and multivariate data analysis and interpretation with SPSS. Chapman & Hall/CRC, Boca Raton, 2006.
[9] M. Jarke and Y. Vassiliou, Data warehouse quality design: A review of the DWQ project, Proceeding of the International Conference on Information Quality (IQ), Cambridge, MA, 1997.
[10] M. Jazayeri, Some Trends in Web Application Development, Future of Software Engineering (FOSE '07), IEEE, 2007, pp. 199-213.
[11] M.J. Kargar, A.A. Ramli, H. Ibrahim and S.B. Noor. Assessing Quality of Information on the Web Towards a Comprehensive Framework 14th IEEE International conference on Internet Communication Technology (ICT/MICC), IEEE, Malaysia, May 2007.
[12] M.J. Kargar, A.R. Ramli, H. Ibrahim and F. Azimzadeh, Formulating Priority of Information Quality Criteria on the Blog, World Applied Sceince Journal, Accepted for World Applied Science Journal (2008).
[13] P. Katerattanakul and K. Siau, Measuring information quality of web sites: Development of an instrument, Proceedings of the 20th international conference on Information Systems, Charlotte, North Carolina, United States, 1999, pp. 279–285.
[14] S.a. Knight and J. Burn, Developing a Framework for Assessing Information Quality on the World Wide Web, Informing Science Journal, 8 (2005) 159-172.
[15] R. Kumar, J. Novak, P. Raghavan and A. Tomkins, On the bursty evolution of blogspace, Proceedings of the 12th international conference on World Wide Web, New York, NY,USA, ACM Press, 2003, pp. 568–576.
[16] Y.W. Lee, D.M. Strong, B.K. Kahn and R.Y. Wang, AIMQ: a methodology for information quality assessment, Information & Management, 40 (2002) 133–146.
[17] H.K.N. Leung, Quality metrics for intranet applications, Information & Management, 38 (2001) 137-152.
[18] O.I. Lindland, G. Sindre and A. Sølvberg., Understanding Quality in Conceptual Modeling, IEEE Software, 3 (1994).
[19] C.D. Manning and H. Schutze, Foundations of Statistical Natural Language Processing, MIT Press, Cambridge, MA (1999).
[20] G. Mishne and N. Glance. Leave a reply: An analysis of weblog comments In Third annual workshop on the Weblogging ecosystem, Edinburgh, Scotland, 2006.
[21] D.L. Moody, G. Sindre, T. Brasethvik and A. Solvberg. A (2003), .. Evaluating the Quality of Information Models: Empirical Testing of a Conceptual Model Quality Framework presented at 25th IEEE International Conference on Software Engineering (ICSE'2003), Portland, Oregon, 2003.
[22] F. Naumann and C. Rolker, Assessment methods for information quality criteria, Proceedings of 5th International Conference on Information Quality, 2000, pp. 148–162.
[23] I. Ohmukai, K. Numa and H. Takeda. Egocentric Search Method for Authoring Support in Semantic Weblog Workshop on Knowledge Markup and Semantic Annotation (Semannot2003), Held in conjunction with the Second International Conference on Knowledge Capture (K-CAP2003), 2003.
[24] S. Rainville-Pitt and J.-M. D'Amour, Using a CMS to create fully accessible websites Proceedings of the 2007 international cross-disciplinary conference on Web accessibility (W4A), Banff, Canada ACM, 2007, pp. 130-131.
[25] M. Recker and J. Pitkow, Predicting document access in large multimedia repositories, ACM Transactions on Computer-Human Interaction (TOCHI), 3 (1996) 352-375.
[26] L. Ricigliano. Criteria for Evaluating Information on the Web, Collins Memorial Library, 2006.
[27] G. Shanks and B. Corbitt, Understanding data quality: Social and cultural aspects, Proceedings of the 10th Australasian Conference on Information Systems, 1999.
[28] D. Sifry. The State of the Live Web, 2007.
[29] B. Stvilia, M. B. Twidale, L. Gasser and L. C. Smith Information quality discussions in Wikipedia, Proceeding of the International Conference on Knowledge Management (ICKM05), 2005, pp. 1--20.
[30] B. Stvilia, M.B. Twidale, L.C. Smith and L. Gasser, Assessing information quality of a community-based encyclopedia, Proceedings of the International Conference on Information Quality (ICIQ), Cambridge, MA, 2005, pp. 442-454.
[31] G. Tyburski. Criteria for Quality in Information, 2006.
[32] R.Y. Wang and D.M. Strong, Beyond accuracy: what data quality means to data consumers, Journal of Management Information Systems, 12 (1996) 5-34.
[33] Wikipedia. Content Management System, 2008.
[34] Y. Zhang, H. Zhu and S. Greenwood, Empirical Validation of Website Timeliness Measures, Proceedings of the 29th Annual International Computer Software and Applications Conference (COMPSAC'05), IEEE, 2005, pp. 313-318.
[35] Y. Zhang, H. Zhu, Q. Huo and S. Greenwood, Measurement of Timeliness of Web-based Information Systems, Proceedings of the 6th World Multi-Conference on Systemic, Cybernetics and Informatics (SCI 2002), 2002.
[36] X. Zhu and S. Gauch, Incorporating quality metrics in centralized/distributed information retrieval on the World Wide Web, Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval, Athens, Greece, ACM, 2000, pp. 288–295.
[37] X. Zhu, S. Gauch, L. Gerhard, N. Kral and A. Pretschner, Ontology-based web site mapping for information exploration, Proceedings of the eighth international conference on Information and knowledge management ACM Press, 1999, pp. 188 - 194

Mohammad Javad Kargar received MSc degree in computer architecture engineering from University of Science and Research, and PhD Degree in Information Technology and Multimedia system from University Putra Malaysia in 2008. He is currently a faculty member in Islamic Azad university –Maybod Brach in Iran. His research interests include web and information quality.

Fatemeh Azimzadeh received the B.S. degree in Applied Mathematics from University of Tehran, Iran in 1999 and MSc Degree in information Technology and Multimedia System from University Putra Malaysia in 2008. She is currently working toward the PhD degree at the University Putra Malaysia.