

A Content Management System for Building Quality Blogs

Mohammad Javad Kargar

Islamic Azad University, Maybod Branch, Department of Computer Engineering, Iran

showkaran@hotmail.com

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 a content management system (CMS) to facilitate measuring information quality on the Blogs yet. This paper presents a novel CMS for ranking quality of information on the Blog. The CMS includes appropriate criteria for ranking Blogs according to IQ parameters. The developed CMS collect and calculate IQ scores of Blogs automatically.

Keywords—Information Quality, Blog, Content Management System

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 Blog 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 Blog, sometimes written as web log or Blog, 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. Blog articles contain various topics such as on personal activities, technology, politics, international problems, and so on. By browsing Blog articles, we can find frank and up-to-date opinions on various topic 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 Blogs. We believed that Blog can be a suitable application for evaluating quality of information because Blogs use common templates, so that quality of content of a Blog is almost equal to quality of Blog. Blog owners usually are not computer and information technology specialist. They take advantage of prearranged templates, concentrate on content of Blog prefer to think about its template and appearance. In the research we are going to develop a content management system for Blogs. 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 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.

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

The first academic research on the Blogging community appeared WWW conference in 2003 [4]. Since then, the first bloggers' conference has also occurred, 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 METHODLOGY

The general aim of the research is to develop a CMS for building quality Blogs. Developing a CMS for producing Blog and then assessing quality of information in the Blog is a multi-layer process.

The first phase in methodology was identifying information quality criteria for Blogs. Evaluating information quality on Web and especially on Blog 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 Blog server and Blog management system. The Blog management system is heart of our framework as system test-bed. The Blog management system comprises all the software modules and components which are employed in the whole framework.

After implementation of the Blog management system, information quality modules were added. In the end of this stage the Blog system is ready for Blog creation and data entry. Thus participants managed to create their Blogs and start data entry and other activities in the Blog 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 Blogs and complete other parts of their Blogs. At the same time, the data entered by the users and their activities were registered in system database which is located in Blog management system.

After data entry and Blog construction stage, all the data were saved into the database system. The measured criteria could evaluate quality of Blog's content. The score of each Blog shows in each Blog. Also visitors could observe score of each Blog, if Blog owner set permissions for visitors.

A. *Blog Management System*

Blog management system is core of the framework. To develop the Blog 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]. Figure 1 shows structure of the CMS in the current research.

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].

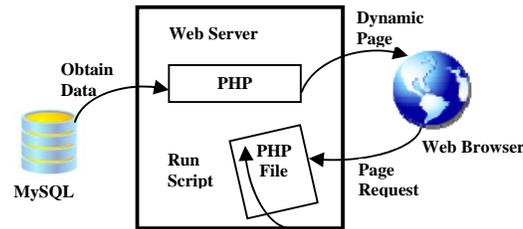


Figure 1: Structure of CMS in the Quality Blog System

Typically, a CMS consists of two elements: the content management application (CMA) and the content delivery application (CDA). The CMA element allows the content manager or author, who may not know Hypertext Markup Language (HTML), to manage the creation, modification, and removal of content from a Web site without needing the expertise of a Webmaster. The CDA element uses and compiles that information to update the Web site

In the simplest case, there are three components designed in a CMS. In addition to the host computer where everything is stored: a database to organize the information; a script language to provide an interface to the information; and a web server to present the information as an HTML file for a web browser. The current Blog management system includes several technologies commonly used in the modern web applications. The system was developed by PHP, MySQL, HTML, CSS, JavaScript, and Ajax. Figure 2 shows Blog content management system architecture.

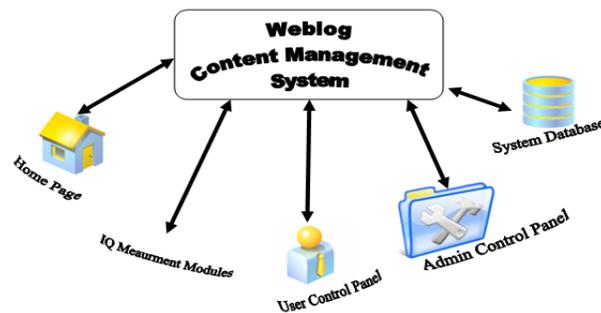


Figure 2: Blog Content Management System Architecture

The Blog 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 Blog management system. User control panel is an interface designed for users to produce and manage contents of their Blogs. This panel provides features which user needs to manage a Blog such as links management, edit and create template, add new article, manage comments and configuration of Blog. IQ modules measure information quality for each Blog based on considered IQ criteria. System database records users' activities on the Blogs. Administrator control panel is an interface that was designed for administrator to control, manage, and monitor the Blog management system. This panel comprises two major panels. A panel manages content structure of the website. So this panel, controls contents of homepage and other pages that are linked to the homepage. Menu management, publish new subject, subjects management, news and announcement management, and page management are the most important features in this panel. Another panel controls and monitors user activities on the Blogs by accessing to the system database and log files. In the panel, administrator can observe and monitor many activities that are performed by Blog owners. Whenever a user logs in to the system, administrator can observe details of the login in the panel. As soon as user posts an article, the panel enables the administrator to control details of the posted article. Overall status, observation of members and entries, edition of templates and HTML, observation of log files, configuration of specifications and registration are the most important features in this panel. In sum, administrator control panel facilitated content management for administrator; provide an interface for monitoring the Blog system. The panel makes

easier extension and maintenance of the Blog management system because it is designed according to properties of a content management system. User control panel is an interface designed for users to produce and manage contents of their Blogs. A user is able to publish and manage his/her Blog by logging to the user control panel. This panel provide features which user needs for managing a Blog. Links management, adding new article, observing posted article, template selection and edition, general configuration of Blog, comments configuration, friends' management, and sending files are the principle features of the user control system.

B. IQ Criteria and Sub-criteria

Identifying criteria is often the initial part of an evaluation process. For evaluating information quality on Web, and especially on Blog, it is required to select appropriate criteria. The criteria must be integrated to cover evaluation of information quality on Blog. On the other hand, redundancy and complexity issues must be taken into account.

in the current research 18 criteria were selected based on earlier research [12]. Table 1 shows selected IQ criteria for calculating quality of Blog's content. The selected criteria have some sub-criteria. From this, 18 sub-criteria could be calculated automatically while 9 qualitative criteria were obtained by voting were selected for information quality assessment on the Blog. 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 Blog. When users intended to leave a comment for a Blog'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, Blog age, number of posted articles, number of external links, number of internal links,

Table 1: selected IQ criteria for calculating quality of Blog's content

Criteria	Sub-criteria
Understandability	-
Informativeness	-
Representation	-
Accuracy	-
Completeness	-
Timeliness	-Last update -Last login
Believability	-
Concise	-
Cohesiveness	-
Maintainability	-Meta information checking
Availability	-Ratio of visited links to failed links
Authority	-Number of written comments -Blog Age -Number of posted articles -Number of external links -Number of internal links
Latency	-Initial load time - Full load time
Popularity	-Number of received comments -Average of received comments -Number of visitors -Number of referred links

Customer support	Customer support link
Amount of Data	-Blog size
Objectivity	-
Redundancy	Ratio of multimedia elements to the overall information

initial load time (time for loading first component of Blog), full load time (time for loading all the components of Blog), 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, Blog 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 Blog context such as number of received comments, average of received comments, number of referred links, number of written comments.

C. Data entry and Blogs construction

After developing the Blog management system and incorporating the information quality criteria to the system, it was time for the Blog 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 Blogs (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, Blog construction and completion was as a part of assignments that were assigned to the students. As soon as the Blog 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, Blog construction and Blogs activity was from 20/11/2009 until 20/1/2010. In the two months period, the students could make their Blogs, post articles, write comments, add friends, insert links and other activities which are usual in popular Blog service providers. All the activities were stored in system database. In the two months period, 473 Blogs were created by the students. The number of Blogs was more than the number of students because some of the students have made more than one Blogs supposing that making more Blogs 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. ADVANTAGES OF THE CMS

The aim of the project was a building a content management system which consider quality criteria. As mentioned earlier, after developing the system and incorporating the selected criteria, a user interface was designed; We monitored users' behavior in the test period. The results of our observations showed the developed system worked and collected data in the Blog. According to our initial aim, each Blog's owner could observe its scores in the Blog. Also our feedbacks shows users could observe the scores.

The advantage of the current CMS is that as an experimental framework which evaluates and rank Blogs in a real environment. The current framework is able to rank IQ on Blogs automatically. The implemented Blog management system automatically calculates and collects values of 18 variables for ranking the Blogs. Data collection and score calculation were done automatically (and without interposing), and thus had minimum noise. However, because of importance of subjective criteria, nine subjective criteria were obtained by voting.

An important advantage of the current system is a rich data set collected by system log files in the CMS. The data set determine our road map for next stages of the research. We are going to conduct data analysis and data mining technique such as cluster analysis to find a model for evaluation of Blog content. We expect that the model reveal new dimensions in evaluation of Blog content, because Blogs have especial nature and social networking criteria.

V. CONCLUSIONS

The aim of this research was to develop a content management system for evaluating information quality on Blogs. In order to develop the content management system, appropriate information quality criteria for Blogs were first identified. The next stage included the implementation of the Blog management system as a test bed

of the research. The Blog management system contained all the facilities for content production on Blog. Moreover, all the activities carried out by participants, as well as their information quality scores were saved in the system database. After developing the Blog server, participants were invited to create Blogs and add contents, in a period of two months.

One of the key advantages of the current CMS is that as an experimental framework which evaluates and rank Blogs in a real environment. The current framework is able to rank IQ on Blogs automatically. The implemented Blog management system automatically calculates and collects values of 18 variables for ranking the Blogs. The produced CMS is a specific content management system with content oriented aspect.

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 Blog articles *AI & Society*, Springer London, 22 (2007) 253-263.
- [4] N. Glance, M. Hurst and T. Tomokiyo. *BlogPulse: Automated Trend Discovery for Blogs WWW 2004 Workshop on the Blogging 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 Blogs, *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 Science 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. Sølvsberg., Understanding Quality in Conceptual Modeling, *IEEE Software*, 3 (1994).
- [19] C.D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA (1999).
- [20] G. Mishne and N. Glance. Leave a reply: An analysis of Blog comments *In Third annual workshop on the Blogging 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 Blog *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 an Associated Prof and faculty member in Islamic Azad University –Maybod Branch in Iran. His research interests include web and information quality.