Big Data Case Study: A Web-Log Based Framework for Analyzing the Use-Quality of a Website

Hyun-Ho Lee [1], Jin-Chul Park [2], Jong-Min Lee [2], Nam-Hun Park [3], Won-Suk Lee [4]

[1.] Yeonsung University, Anyang-si Gyeonggi-do, South Korea
[2.] Hanyang University, Ansan-si Gyeonggi-do, South Korea
[3.] Anyang University, Anyang-si Gyeonggi-do, South Korea
[4.] Yonsei University, Seoul, South Korea
hhlee@yeonsung.ac.kr

Abstract: One key element of assessing a website is to measure how much it is used by users. This paper proposes a framework of measuring the use-quality of a website by analyzing its web-log. A web-log is a representative type of big data, which is spotlighted as an important emerging issue. With the web-log of a website, metadata for its users and contents allows use-quality analysis from various viewpoints. As a big data case study, this paper describes the overview of how to process the web-log and metadata under the proposed framework. It describes web-log standardization, web-log DW model, knowledge repository, and the overall analysis workflow. In addition, it shows the actual case that analyzes the use-qualities of a few websites under the proposed framework.
[Hyun-Ho Lee, Jin-Chul Park, Jong-Min Lee, Nam-Hun Park, Won-Suk Lee. **Big Data Case Study: A Web-Log Based Framework for Analyzing the Use-Quality of a Website**. *Life Sci J* 2014;11(7):588-596] (ISSN:1097-8135). http://www.lifesciencesite.com. 81

Keywords: big data; web-log; meta-data; use-quality; analysis framework; log DW

1. Introduction

Nowadays, in the age of mass information, most organizations manage websites that publicize their contents on the internet regardless of their organizational characteristics and scales. Websites are categorized into several classes in respect of their characteristics. Some organizations merely provide practical information through their websites. Others actively advertise their products and services using the websites. Since the actual and perceived importance of on-line marketing by a website is increasing, there is now a growing trend for most organizations to strengthen their provision of on-line services. In addition, it is considered more and more important to analyze websites in various respects, including brand power and marketing.

From a common-sense standpoint, a website should be useful to users. In other words, the website that is ignored by users is useless, despite having good contents and performance. Thus, it is important to analyze a website in respect of how much its contents are utilized by users. Such data indicates the *use-quality* of a website.

Until now, various methods have been proposed in order to estimate the use-quality of a website. However, as shown in Madu (2002), Liu (2000) and Misic (1999), they have almost exclusively measured the service performance of the website or judged the service completeness qualitiatively through user surveys. This paper proposes a new method that measures the use-quality of a website by analyzing its web-log.

Recently, big data has been spotlighted as an important emerging issue. As one of the representative types of big data, a web-log keeps track of users' click streams (i.e., where they go in the website and how long they stay). Measuring the use-quality of a website based on its web-log can be more precise than any other method, because it is based on the actual actions of website users. If the log information of a website can be associated with the meta information of users and web contents, the scope of web-log analysis is remarkably extended by enabling to administrator to identify which contents are accessed by which users. The proposed method provides a novel framework that measures the use-quality of a target website, from the site level to the individual contents level, by monitoring and analyzing its web-log for the metadata about users and contents.

The remainder of this paper is organized as follows: The next section reviews the related works. Subsequently, we define the use-quality of a website. Section 4 describes the standardization of web-log & metadata and the model of a web-log data warehouse. The next section describes the analysis workflow under the proposed framework. Subsequently, we introduce the case studies for the proposed methodology. The final section summarizes and concludes the present paper.

2. Related Works

Liu (2000) identifies critical factors associated with website success. It explores the factors in the context of electronic commerce (EC). The research framework was derived from information

systems and marketing literature. Webmasters from Fortune 1000 companies were used as the target group for a survey. Four factors that are critical to website success in EC were identified: 1) information and service quality, 2) system use, 3) playfulness, and 4) system design quality. Misic (1999) presents a case study on how benchmarking is used to determine how one organization's website compared to websites of related schools and professional organizations. The results of the benchmarking study provided a measure of how a given website compares to the sites of related organizations, ideas on how it may be further enhanced, and be also evaluated regularly. Juan (2007) introduces a Knowledge Base (KB), which consists of a DB-type repository, for maintaining the patterns and rules as an independent program that consults the pattern repository. Using the proposed architecture, either an artificial system or a human user can consult the KB in order to improve the relation between the website and its visitors. Bonchi (2001) introduces intelligent web caching algorithms that employ predictive models of web requests; the general idea is to extend the least recently used (LRU) policy of web and proxy servers by making it sensitive to web access models extracted from web-log data using data mining techniques. Two approaches to this end have been studied in particular: frequent patterns and decision trees. The experimental results of the new algorithms show substantial improvement over existing LRU-based caching techniques, in terms of *hit rate*. However, unlike this paper, Bonchi (2001) focuses on algorithm development for improving the performance of a web-server rather than use-quality.

3. Use-quality

In order to measure the use-quality of a website, four perspectives are proposed, as follows:

- User (Who): How many users visit the website?
- Pageview (How many): How many web pages do users visit?
- Contents (What): What content do users view frequently?
- Residence time (How long): How long do users reside?

By combining these perspectives, several indicators can be defined as shown in Table 1.

Based on the candidate indicators of Table 1, a set of final indicators are determined by considering their respective effectiveness. Table 2 shows whether the candidate indicators are accepted together with their respective measurement method. Additionally, if they are not accepted, the reason behind the rejection is described. As shown in Table 2, nine indicators are finally adopted.

Table 1. Candidate indicators based on four perspectives

	User	Pageview	Contents	R. Time
User	-visits(#) -visitors(#) -foreign visitors(%)	-pageviews(#) per a visit	-content views(#) per a visit	-residence time per a visit
Pageview	-	-pageviews(#)	-content views(%) -pageviews(#) per a content	-residence time per a pageview
Contents	-	-	-content views(#)	-residence time per a content view
R. Time	-	-	-	-residence time

Table 2. The acceptance of candidate indicators

indicator	Measuring method	Accept?	Rejecting reason
visits(#)	[1]transactions(#)	Y	
visitors(#)	different clients(#)	Y	
foreign visitors(%)	transaction(%) by foreign clients	Y	
pageviews(#)	[2]total effective log records(#)	Y	
content views(#)	effective contents view log records(#)	Y	
residence time	[3]the difference between log-times	N	meaningless if a visit is ignored
pageview(#) per a visit	effective log records per a transaction	Y	
content views(#) per a visit	effective contents view log records(#) per a transaction	Y	

[1] logical visit unit extracted by analyzing the log-times of a web-log
[2] *see* Section 6.
[3] The time difference between the first and the last among target logs

residence time per a visit	the difference between log-times per a transaction	Y	
content views(%)	effective contents view log records(%)	Y	
residence time per a pageview	the difference between log-times per an effective log record	N	not effective because a user behavior is unknown
pageviews(#) per a content	effective log records(#) per a content	N	full volume of contents is unknown
residence time per a content view	the difference between log-times per an effective log record	N	not effective because a user behavior is unknown

4. Analysis Framework

Most websites manage their web access logs. In addition, the information about users and their contents is stored in their database. In order to effectively estimate the use-quality of a website, including comparing with other websites, various types of web-logs and metadata should be standardized to be easily shared/cross-referenced.

4.1. Analysis Framework

Presently, there are several commonly-used web-log formats. They differ slightly according to web-servers in which they are operated. An IIS web-server selectively uses one of four types: MS IIS, W3C Extended, ODBC Logging, and NCSA Common Log Format (CLF). On the other hand, Apache and WebToB web-servers only use NCSA CLF (see Table 3).

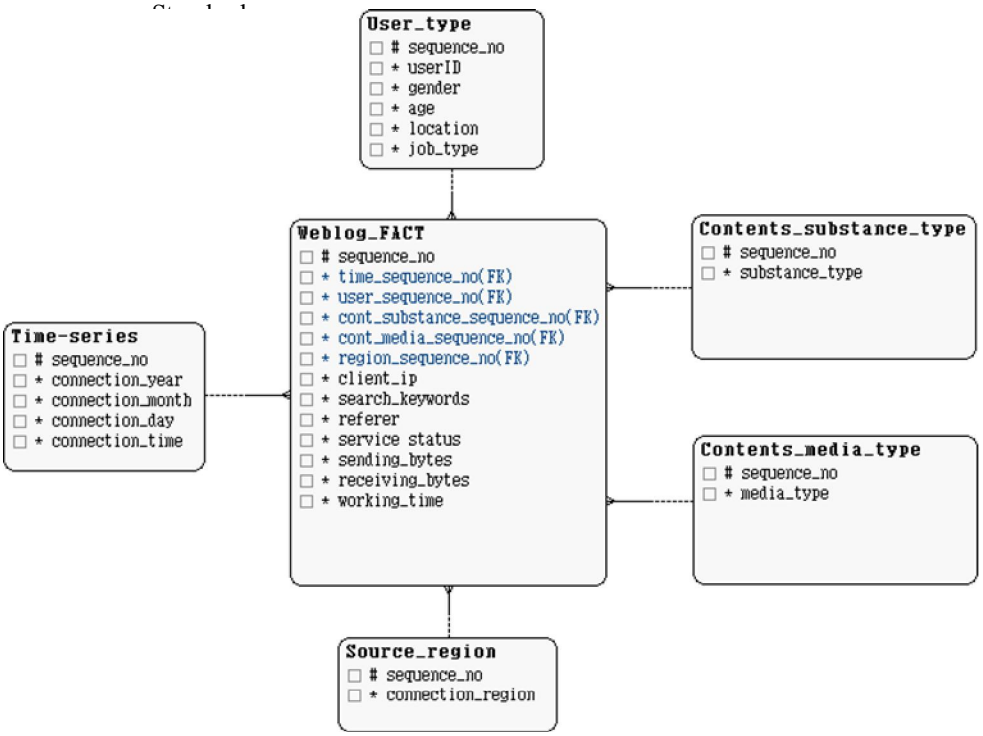

Figure 1. Web-log DW(DataWarehouse) Schema

Table 3. Web-log formats

Operating System	Web Server	Log format
WINDOWS	IIS	NCSA CLF, MIS IIS, W3C Extended, ODBC Logging
UNIX/LINUX	Apache, WebToB	NCSA CLF

By analyzing the four formats shown in Table 3, we propose a standard web-log format consisting of 10 items. Table 4 shows the individual item and its usage of the proposed standard format. Since the top four items of the standard format are basically managed commonly in the four formats of Table 3, they are defined as mandatory items of the standard log format. In particular, the 4th item, URL, records the information on which URL a user visits including search keywords. Also, if user ID or contents ID are contained in the URL parameter, it is possible to acquire user or contents information by analyzing the parameter. This enables the use-quality of a website to be measured in various respects. Table 5 shows whether each of four log formats of Table 3 support each item of the proposed standard format.

Table 4. Proposed standard web-log format (not show).

Table 5. The present condition for standard format

No	Standard item	NCSA CLF	MS IIS	W3C Extended	ODBC Logging
1	Client-IP	O	O	O	O
2	Log time	O	O	O	O
3	Method	O	O	O	O
4	URL	O	O	O	O
5	Cookie			O	
6	Referer			O	
7	Service Status	O	O	O	O
8	Sending Bytes		O	O	O
9	Receiving Bytes	O	O	O	O
10	Working time	O		O	O

4.2. Metadata standardization

The analysis viewpoint for the use-quality of a website can be classified into user and contents viewpoints. If the type information about user and contents is managed in a meta-DB, it is possible to analyze various aspects for users' contents utilization – the contents utilization patterns by user types, the favorite contents types of a specific user type, and the major user type using a specific contents type, etc. In case of a user meta-DB, detailed user-specific data – userID, SSN, address, and telephone number, etc. - are excluded from managed criteria because of the issue of personal privacy protection. In fact, it is enough to analyze user's utilization patterns to only provide user type information – gender, age band, residential district, and job category, etc. A userID is needed to identify a specific user. However, since it is necessary to judge whether or not target users are the same person in the proposed metadata, a userID is converted to a corresponding virtual sequence (virtual userID). The types of contents metadata can be categorized into two criteria: contents substance and contents media. The former indicates the logical classification of individual contents. For example, in an on-line bookstore site, it can contain contents categorized as literature, economy, examination, computer, and so on. The latter indicates the physical classification for which media individual contents are delivered with. For example, it could be categorized as text, sound, image, moving picture, and so on.

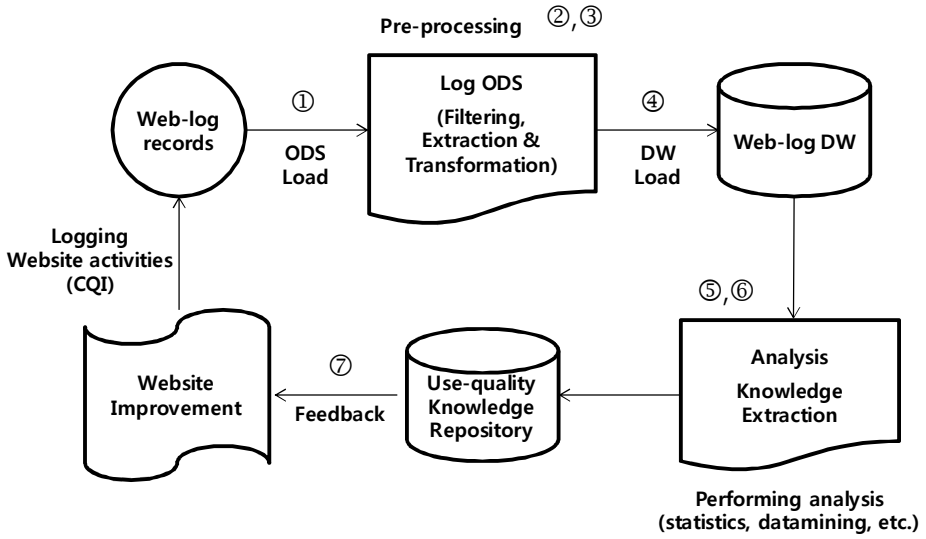

Figure 2. The framework for the use-quality analysis

4.3. Web-log data warehouse

In order to apply a web-log and a meta-DB to the use-quality analysis of a website, a log DataWarehouse (DW) should be constructed like Bonifati (2001). In the DW data model (Star Schema), a web-log is mapped to a fact table and the metadata for users and contents are mapped to dimension tables. A fact contains analysis target attributes. On the other hand, a dimension means analysis viewpoints. Beside users and contents viewpoints, it can include various viewpoints such as Time-series, indicating when log tuples are recorded, and Source_region, indicating where client connections are established, and so on. Figure 1 shows the example of a web-log DW that can be constructed from a web-log and a meta-DB. In the meta-DB, the dimension table User_type is concerned with users metadata. In addition, two dimension tables, Contents_substance_type and Contents_media_type, are each concerned with contents metadata.

5. Analysis Workflow

The fact that a user visits a website is recorded in the corresponding web-log. Generally, since an individual web-page in a website contains more than one object, the record of a web-log does not exactly correspond to one page view. In other words, a page view can produce more than one log record because each object composing a web-page can independently produce a log record. Many such log-records are not useful to the web-log analysis, since the number of pages visited by users is important in respect of estimating use-quality of a website. Accordingly, in order to enhance the analysis accuracy, it is necessary to filter-out useless log-records. The attributes {virtual userID, access contentID, keywords} of the fact table in Figure 1 should be extracted from the items {URL, Cookies} of the proposed standard web-log format. In addition, the values of foreign keys referring to dimension tables should be searched from standard log-items. The process that extracts and transforms the log-items to load them into the fact table is referred to as a pre-processing.

In order to measure the use-quality of a website, various analysis tasks can be performed for the data loaded into a web-log DW through the pre-processing, such as methods proposed in Velasquez (2003). They include statistical analysis, such as page view count by user types, and data mining analysis such as association for contents utilization. The results of a specific analysis task are manufactured into corresponding use-quality knowledge items to be stored in a use-quality knowledge repository like Juan (2007) and Debenham (2001). Use-quality knowledge items are defined to reflect the analysis requirements. Subsequently, a knowledge repository is designed based on the defined knowledge items. The knowledge items stored in the repository are used to improve the use-quality of a target website. They are feedbacked to re-design a web-log format, a meta-DB, and a web-log DW. Subsequently, based on the re-designed log and databases, new use-quality knowledge items are generated. Thereby, a CQI (Continuous Quality Improvement) chain is established. Figure 2 shows a series of use-quality analysis activities for a specific website.

① ODS Load (pre-processing)
The task that loads the record of a web-log into an Object Data Store (ODS) for the purpose of pre-processing the log records

② Filtering (pre-processing)
The task that filters out useless log records

③ Extraction & Transformation (pre-processing)
The task that extracts the attributes of a DW fact table from log records and searches the values of foreign keys referring to DW dimension tables.

④ DW Load (pre-processing)
The task that loads the pre-processed log records into a DW fact table.

⑤ Analysis
The task that analyzes the newly-inserted tuples of a DW fact table with target-oriented methods.

⑥ Knowledge Extraction
The task that manufactures the use-quality knowledge items from the analysis results and stores them in a knowledge repository.

⑦ Knowledge Feedback
The task that feedbacks the newly produced knowledge to improve the use-quality of a target website.

6. Analysis Case Studies

Under the proposed framework, we have analyzed the web-logs of 19 websites in order to measure each's use-quality. The targeted websites have been chosen from among the public institutions that provide cultural information services. By comparing them with one another, the utilization characteristics of the websites can be analyzed. In order to analyze the more effectively, we have classified them by service field and type. Service fields are categorized into nine classes including eight national standard classes (MCST (2014)) and one miscellaneous class, as shown in Table 6. The service types were defined as belonging to one of five classes according to the goals of the services that the websites provided, as shown in Table 7. 19 target websites are listed in Table 8; their web-logs have been collected for a certain period — from 1 month to 1 year.

Table 6. The classification of service fields

Categories	Explanations
Cultural Industry **(CID)**	Websites associated with cultural industries — movies, performance, music, design, craft, broadcast, newspapers, internet content, etc.
Cultural Inheritance **(CIH)**	Websites associated with cultural inheritance — national museums, folk museums, etc.
Culture Art **(CA)**	Websites associated with cultural arts — literature, plays, art, dance, traditional art, etc.
Tourism **(T)**	Websites providing tourism content — tourism organizations, etc.
Adolescent **(A)**	Websites associated with adolescent living culture — adolescent activities, adolescent research, adolescent support, etc.
Sports **(S)**	Websites associated with sports — sport for all, sports games, sports events, sports industries, etc.
Library **(L)**	Websites associated with libraries — libraries, bookstores, etc.
Information Shaping Operations **(ISO)**	Websites associated with Information shaping operations — informationization, copyrighting, national languages, national archives & image (video), etc.
Etc. **(E)**	Cultural information websites that do not belong to the above eight categories

Table 7. The classification of service types

Categories	Explanations
Organization Advertisement Category (OA)	Websites that mainly provide formal information for a specific organization including notices, news, and simple reference BBS; most organizations' official websites belong to this category.
Contents Providing Category (CP)	Websites that are operated to provide content and data; they mainly serve their own content.
Search Service Category (SS)	Websites that are operated to provide search portal services; generally, they possess an integrated search engine.
Contents Producing Participation Category (CPP)	Websites that are operated so that users can produce content themselves and post it such as the cultural UCC website.
User Community Category (UC)	Websites that provide a platform for user communication, such as an internet café or society page.

In general, for a user pageview, several log records are generated in corresponding web-logs because a log record is generated for each individual object that comprises the webpage. Since no analysis tasks were performed for a web-log for most websites, we have performed a pre-processing task that filters out useless log records for all the web-logs that are generated by a webserver. The web-log that results from the pre-processing task is called an *effective web-log* (see Table 8). Accordingly, the number of effective log records approximately corresponds with the number of actual user pageviews.

Table 8. Analyzed websites list

No	Abbr.	Organization	URL	Classifi-cations		Effective Log Re- cords(#)
				Field	Type	
1	*e-museum*	Museum Portal of Korea	www.emuseum.go.kr	CIH	SS	7,746,166
2	*museum*	National Museum of Korea	www.museum.go.kr	CIH	CP	30,786,795
3	*kcisa*	Korea Culture Portal	www.culture.go.kr	CA	SS	21,729,891
4	*nfm*	The National Fork Museum of Korea	www.nfm. go.kr	CIH	CP	2,549,230
5	*nfm-k*	The National Fork Museum of Korea for Kids	www.kidsnfm.go.kr	CIH	CP	1,284,474

6	*arch-nfm*	The Center of Folklife Archives	archive.nfm.go.kr	CIH	CP	1,137,126
7	*moca*	National Museum of Modern and Contemporary Art	www.moca.go.kr	CA	OA	17,318,857
8	*visitkorea*	Korea Tourism Organization	www.visitkorea.or.kr	T	SS	26,658,138
9	*kocis*	Korean Culture and Information Service	www.korea.net	E	CP	21,017,152
10	*gugak*	National Gugak Center	www.gugak.go.kr	CA	OA	3,156,355
11	*naa*	The National Academy of Arts of Korea	www.naa.go.kr	CA	OA	13,667,894
12	*kocca*	Korea Creative Contents Agency	www.kocca.kr	CID	OA	2,156,054
13	*kmdb*	Korean Film Archive	www.kmdb.or.kr	CID	CP	1,972,744
14	*kofic*	Korean Film Council	www.kofic.or.kr	CID	OA	848,860
15	*tour*	Tourism Knowledge & Information System	www.tour.go.kr	T	CP	4,524,850
16	*korean*	The National Institute of Korean Language	www.korean.go.kr	CIH	OA	3,162,776
17	*mcst*	Ministry of Culture, Sports and Tourism	www.mcst.go.kr	E	OA	29,978,196
18	*sports*	Korean Olympic Commitee	www.sports.or.kr	S	CP	5,942,877
19	*kspo*	Korea Sports Promotion Foundation	www.kspo.or.kr	S	OA	5,567,658
The total number of effective log records						201,206,093

For each effective web-log of the 19 websites, we have analyzed its use-quality by measuring the proposed nine indicators of Section 3. Table 9 shows the results of analyzing the use-quality of the target websites with respect to six indicators: Visits (#), visitors (#), foreign visits (%), pageviews (#), pageviews (#) per a visit, and residence time (min) per visit. Unfortunately, three indicators associated with content were measured for only three websites —*tour, visitkorea, kocis* — because the other web-logs did not have any information that could identify their content.

Table 9. Use-quality analysis results for 19 websites.

No	website	Use-quality indicators (day average)					
		visits(#)	visitors (#)	foreign visits (%)	page-views(#)	Pageviews (#) per a visit	residence time(min) per a visit
1	*e-museum*	3,090	1,226	46.1	27,387	8.9	2.9
2	*museum*	22,081	7,532	18.6	174,532	7.9	1.8
3	*kcisa*	4,773	1,783	26.0	42,191	8.8	3.0
4	*nfm*	3,103	1,464	17.2	23,657	7.6	2.8
5	*nfm-k*	2,869	1,420	6.0	19,682	6.9	3.4
6	*arch-nfm*	303	82	18.3	2,760	9.1	50.3
7	*moca*	4,610	2,862	13.3	28,460	6.2	4.6
8	*visitkorea*	182,652	49,226	66.9	1,777,165	9.7	1.5
9	*kocis*	11,498	5,274	88.2	86,805	7.5	10.3
10	*naa*	1,239	329	47.8	12,574	10.1	7.3
11	*gugak*	4,768	1,742	34.2	44,997	9.4	3.2
12	*kocca*	5,755	2,312	25.6	38,167	6.6	2.9
13	*kmdb*	2,587	964	22.2	32,316	12.5	1.8
14	*kofic*	3,059	1,854	20.9	22,942	7.5	29.8
15	*tour*	1,679	666	35.4	15,710	9.4	3.3
16	*korean*	3,479	1,379	11.0	26,637	7.7	1.6
17	*mcst*	18,480	5,447	34.0	163,815	8.9	1.7
18	*sports*	3,444	2,551	5.7	18,830	5.5	18.9
19	*kspo*	5,917	2,589	33.4	16,702	2.8	1.9

Figure 3 compares the use-qualities of the top ten sites from the 19 websites with respect to *visits (#), visitors (#),* and *pageviews (#)*. As shown in this figure, *visitkorea* can be designated as the most utilized website with respect to these three indicators, followed by *museum* and *mcst*. In addition, the value distributions of these indicators display a similar story even though the indicators' rankings for the websites are slightly different.

Figure 4 compares the indicator *foreign visitors (%)* for each website. *kocis*'s value is the highest in this regard. This means that the utilization characteristics of *kocis* accords with its stated goals. *visitkorea*'s value is next highest, reflecting foreigners' interest in touring Korea. *visitkorea* maintains a broad international membership because it serves their online contents with ten different national languages.

Figure 5 shows more or less interesting results, compared with those of Figure 2. From these two figures, it can be inferred that the high ranked websites of Figure 4 — *kmdb*, *naa* and *tour* — retain highly loyal memberships even though the volume of visits and visitors is relatively small.

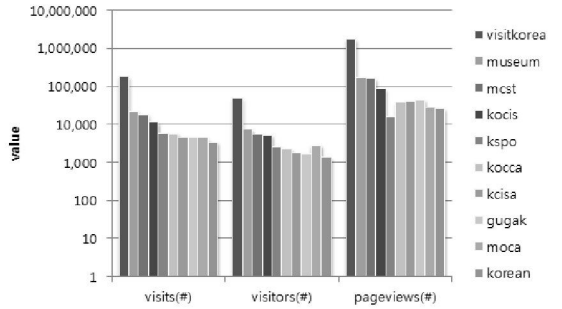

Figure 3. Comparison among the use-quality indicators of top-10 websites - *visits(#), visitors(#), and pageviews(#)*

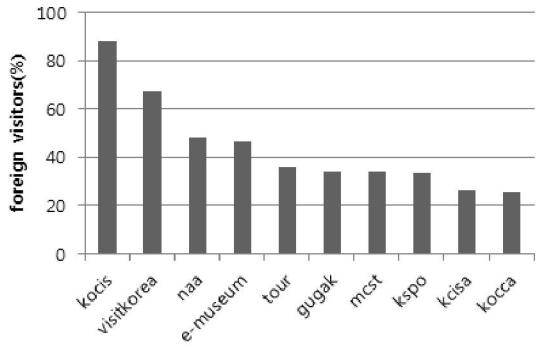

Figure 4. Comparison among the use-quality indicators of top-10 websites - *foreign visitors(%)*

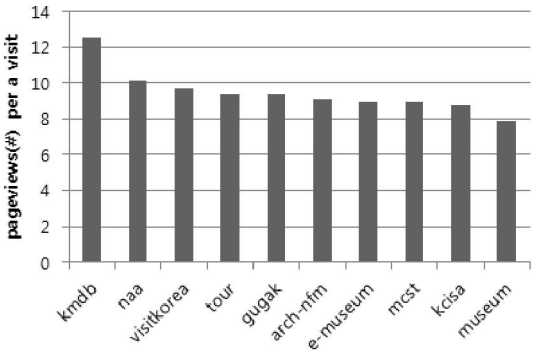

Figure 5. Comparison among the use-quality indicators of top-10 websites – *pageviews(#) per a visit*

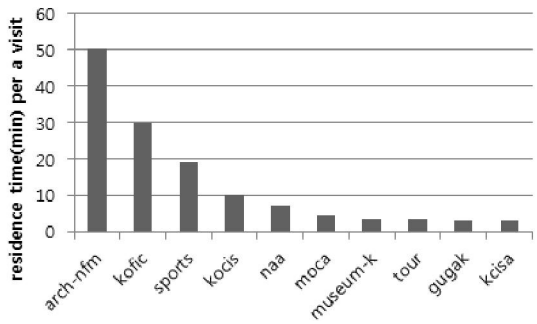

Figure 6. Comparison among the use-quality indicators of top-10 websites – *residence time(min) per a visit*

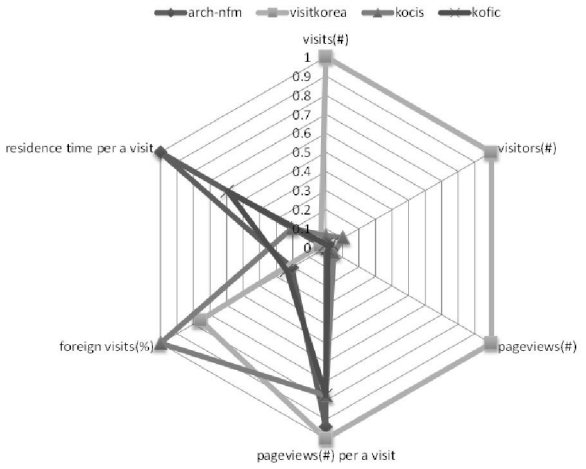

Figure 7. Comparison among the relative value distributions of use-quality indicators

The results of Figure 6 can also be interpreted similarly to those of Figure 5. In addition, in the case of *arch-nfm* and *kofic*, they can have an effect on the residence time as they contain large amounts of video content.

Figure 7 compares the relative value distributions of the use-quality indicators for four websites — *arch-nfm*, *visitkorea*, *kocis* and *kofic*. *visitkorea* can be estimated to be the most popular website because its three indicators — visits (#), visitors (#) and pageviews (#) — are outstandingly higher than the other websites. However, since its residence time per visit is much shorter relative to the other websites, user loyalty seems low. *kocis* has a relatively higher percentage of *foreign visitors* than the other websites; the goal of *kocis* is to inform foreigners of Korean culture, and from these statistics, it can be estimated that it satisfies this goal. The use-quality indicators of *arch-nfm* and *kofic* take on a similar aspect. The degree of their utilization is relatively low, while their residence time is markedly longer than the other websites. In this regard, it can be

estimated that their user loyalty is very high even though their number of users is small. In addition, it can be estimated that they host plenty of video content that requires a long play time.

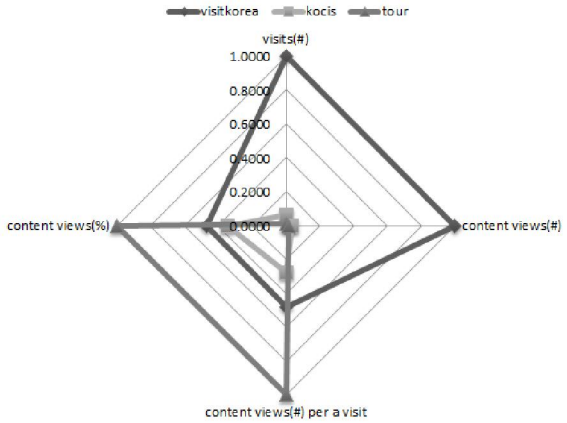

Figure 8. Comparison among the relative value distributions of use-quality indicators associated with contents

Figure 8 compares the relative value distributions of the use-quality indicators associated with content for three websites — *visitkorea*, *kocis*, and *tour*. Both *visitkorea* and *tour* are websites associated with tourism. Even though *visitkorea* has more content views *(#)* than *tour*, *tour* has higher content views (%) and content views (#) per a visit. In this regard, it can be inferred that *visitkorea* is utilized more often by general users and *tour* is utilized in more depth by experts and associated workers.

6. Conclusions

The present paper proposes a framework that estimates the use-quality of a website based on its web-log. In order to enhance the accuracy and variety for the use-quality, it uses a meta-DB containing the metadata of users and contents together with the web-log. Specifically, it proposes the standard web-log format to easily share the various types of web-logs. With such a standard web-log and meta-DB, a web-log DW is constructed. In addition, this paper illustrates the procedure for the use-quality analysis for a given website under the proposed framework. It consists of three phases – pre-processing, analysis, and knowledge extraction & feedback – and makes CQI possible by applying the analysis results to the improvement of a target website and, again, estimating the use-quality of the improved website.

In future works, more detailed and systematic methodologies should be utilized to enhance the accuracy and variety of the use-quality analysis under the future proposed frameworks, including the standardization of analysis methods, the effective design of a use-quality knowledge repository, and a comparison of methods for multiple similar websites with respect for their use-qualities. In addition, it will be important to design the framework and workflow in such a way that enables real-time analysis by distributed computing environments such as Hadoop.

Acknowledgements:

Foundation item: This research was supported by Basic Science Research Programs (NRF-2012R1A1A2009170 and NRF-2011-002530) through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and Future Planning.

Corresponding Author:
Prof. Lee
Department of Military Science
Yeonsung University
34, Yanghwa-ro37beon-gil, Manan-gu, Anyang-si
E-mail: hhlee@yeonsung.ac.kr

References

1. Madu, Christian N. and Madu, Assunpta A., Dimension of e-Quality, International Journal of Quality & Reliability Management 2002; 19(3), 246-258.
2. Liu, C. and Arnett, K. P., Exploring the Factors Associated with Web Site Success in the Context of Electronic Commerce, Information and Management 2000; 38(1), 23-33.
3. Misic, M. M. and Johnson, K., Benchmarking, A Tool for Web Site Evaluation and Improvement, Internet Research 1999; 9(5), 383-392.
4. J.D.Velasquez, H.Yasuda, T.Aoki, R.Weber, A genetic data mart architecture to support web mining, Proceedings of the 4th International Conference on Data Mining 2003; 389-399.
5. Juan D. Velasquez, Vasile Palade, A knowledge base for the maintenance of knowledge extracted from web data, Knowledge-Based System 2007; 20, 238-248.
6. J. Debenham, Knowldege base maintenance through knowledge representation, Proceedings of the 21th International Conference on Database and Expert Systems Applications 2001; 599-608.
7. Ministry of Culture, Sports and Tourism(MCST), http://www.culture.go.kr, 2014.

5/26/2014