Journal of Universal Computer Science, vol. 15, no. 8 (2009), 1589-1606
submitted: 31/3/09, accepted: 24/4/09, appeared: 28/4/09 © J.UCS

Modelling Weblog Success: Case of Korea

Junseok Hwang
(Technology Management, Economics and Policy Program, College of Engineering
Seoul National University, Seoul, Korea
junhwang@snu.ac.kr)

Youngjin Lee[1]
(Technology Management, Economics and Policy Program, College of Engineering
Seoul National University, Seoul, Korea
happy07@snu.ac.kr)

Seunghyun Kim
(Technology Management, Economics and Policy Program, College of Engineering
Seoul National University, Seoul, Korea
octnine2@snu.ac.kr)

Abstract: Weblogs have received attention as new, personalized media. Yet, only few of them attract public attention and become successful. This study explores weblog success factors in three categories: content creation, content features and content diffusion. During the process of content creation in weblogs, we argue that weblog service providers (WSPs) support bloggers' resource collection. We also presume that the volume or the quality of posts in weblogs could be matter to gain visitors' attention when weblog content (i.e., post) is generated. During the process of content diffusion, we assume that use of blogging technologies (BTs) such as trackback or RSS would enhance content-sharing activities between weblogs. Based on the data from a sample of Korean individual weblogs, our analysis indicates that weblog success (in terms of the number of unique visitors per week) is related to the WSP's support level for content creation as well as content features.

Keywords: Weblog Success, WSPs' Support for Content Creation, Content Features, BTs' Contribution for Content Diffusion
Categories: H.3.5, H.4.3, H.5.1, M.6

1 Introduction

In recent years, weblogs have received attention as new, personalized media (see [Rosenbloom 04] and [Du, Wagner 06]). The expression, weblog, first originated from "log of the web" by Barger in December 1997 (see [Paquet 03] and [Safran, Kappe 08]). As the term implies, the web contents in a weblog, called posts, appear in reverse chronological order. In other words, the most recent post is presented at the top of a weblog (see [Paquet 03] and [Safran, Kappe 08]).

According to [OECD 06], the number of weblogs in the world has gone up to sixteen times and the number of post in weblogs has gone up to ten times between January 2004 and January 2006. From recent information from [OECD 08], this

[1] Corresponding author

blogging phenomenon seems to be maintained worldwide. The estimated number of bloggers and blog readers in Japan accounted for 8 and 35 million, each, by March 2007 (see [OECD 08] and [MIC 07]). In the United States, over 10 % of total Internet users experienced weblogs as their owners and about 30 % of Internet users were blog readers in 2007 (see [OECD 08] and [Pew 07]). By the end of June 2008, about 60 % of Internet users had visited weblogs or "mini-hompys" (mini homepages on some Korean social networking service platforms) and over 40% of Internet users owned either weblogs or "mini-hompys" in Korea [NIDA 08].

This prosperity of weblogs has led to rapid increase of internet traffic and attracted researchers' attention on the internet traffic management. Studies in the internet traffic management can be categorized into two fields: studies in the network layer level and studies in the application/contents layer level. Researches on the traffic at the network layer are focused on modelling the generation of traffic and finding solutions to the traffic crisis such as traffic jams. This is due to the function of network layer in the internet industry for providing and maintaining stable networks. [Li, Lim 08] proposed a network traffic model using generalized Cauchy (GC) process and [Li, Zhao 08] suggested a method to statistically measure local irregularity of traffic under distributed denial-of-service (DDOS) flood attacks. [Garcia, Hackbarth 08] presented an approximation method solving traffic aggregation problems. [Kozlak et al. 08] introduced an agent-based approach to solve city traffic crisis such as traffic jams. In the mean time, researches on traffic at the application/content layer are more interested in finding the origins of traffic and its use in terms of the internet services and web contents. That is, these studies classify the internet services or web contents which generate traffic and seek various methods to improve the quality of the internet services, transmission of online contents and profit generation, etc. [Guo et al. 08] mentioned that the majority of the weblog traffic comes from the minority of weblogs. [Guo et al. 08] also suggested that weblog traffic is affected by the internal, external and interface factors and it is difficult to expect the long term weblog traffic. This paper investigate the factors of weblog success in the latter's viewpoint.

As more traffic is generated as the more weblogs flourish, weblog traffic in terms of readership is important issue for the industry as well as researchers. Here, the readership refers to the group of audience participating in exchange of information or personal opinions and weblogs with wider readership often receive financial support from companies for the purpose of advertising. When the readership of weblog is defined as a possible measure to judge weblog success, to the best of our knowledge, there exist only two academic studies to identify the factors affecting the success of weblogs in readership perspective: one of them emphasized the importance of blogging technologies on weblog success [Du, Wagner 06]; the other focused on the type of weblog contents, level of community activities or writing style [Safran, Kappe 08]. The past studies made great contributions to weblog success researches: especially, they illuminated the weblog success researches with the emphasis on weblog readership. However, they only considered limited aspects of blogging aspects. Hence, this paper tries to broaden the weblog success analysis in terms of weblog readership.

Generally, activities related to weblog contents (i.e., posts) involve three components: content creation, content features and content diffusion. Figure 1

illustrates how these components may affect the weblog success. During the process of content creation in weblogs, weblog service providers (WSPs) support bloggers' resource collection: they provide weblog contents, Internet news, dictionaries, etc. on the same weblog domain and we argue that WSPs' support level could affect weblog success. Following [Du, Wagner 06] and [Safran, Kappe 08], we presume that content features or content diffusion would be matter considering weblog success: the volume or the quality of posts in weblogs may have influence on obtaining visitors' attention when a weblog content (i.e., a post) is generated; during the process of content diffusion, use of blogging technologies (BTs) such as trackback or RSS would enhance content-sharing activities between weblogs.

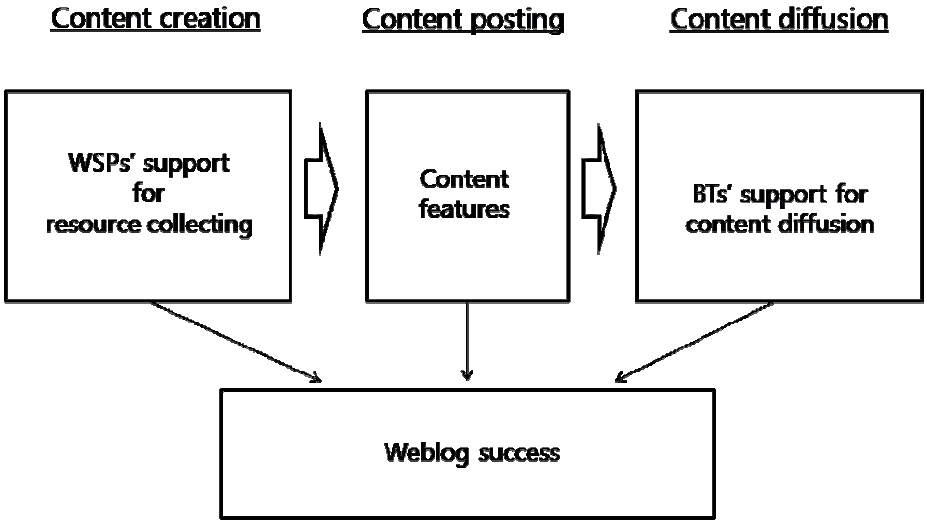

Figure 1: Activities related to Weblog Contents and Weblog Success

In this perspective, past studies are considerably focused on the impact of content features or content diffusion technologies. But, we propose that weblog success needs to be analyzed in more comprehensive perspective and content creation process is also important in weblog success. Then, the key research question is: regarding weblog content-related activities, what makes weblogs succeed? For analysis, we explore weblog success factors in three categories: WSPs' support in content creation (resource collection), content features and BTs' support for content diffusion. The rest of this article is organized as follows: Section 2 reviews related previous studies. Section 3 proposes our study design for weblog success analysis which is useful for conceptualizing activities associated with weblog contents. The results of the empirical analysis are presented in section 4 and discussion, further research suggestions and conclusions are followed in section 5.

2 Related Studies

2.1 Weblog Success

According to the study of [Du, Wagner 06], weblogs are created for people to freely express and exchange their personal opinions, knowledge and information. That is, weblogs serve their owners and users as highly social media: on one side, audience read and comment the weblog contents; on the other side, authors correspond to the reaction of the audience. In this perspective, ensuring a large number of blogging fellows (either visitors or blog readers) is essential to maximize the benefits of weblog utilization. [Guo et al. 08] described this large number of blogging fellows as readership and mentioned that many companies pay huge amount of money to bloggers with wide readership doe to the advertising impact obtained from readership. Consequently, regarding the aim of the creation of weblogs, we can presume that socially successful weblogs are defined as weblogs with wide readership. Considering the characteristics of weblogs as social media, various criteria would be available (e.g., specialty in provision of certain topics, or meeting public interests, etc.). However, this paper confines the measure of weblog success to the wide readership of weblogs.

There are only two past studies on weblog success: [Du, Wagner 06] and [Safran, Kappe 08]. But, they also find the origin of weblog success from the readership of weblogs: this adds validity of our viewpoint on weblog success. Especially, this paper analyzed weblog success by introducing more accurate measure of readership to complement the weakness of two past studies. The summaries and limitations are as follows. [Du, Wagner 06] explored the success of weblogs regarding the level of blogging technologies used. As criteria of weblog success in readership point of view, the popularity ranks and growth rates of weblogs were used. By combining these two criteria, A-list weblogs were divided into six categories: rising stars, core, falling stars, emerging, fringe and dropout. Following the definition of [Herring et al. 05], "A-list" weblogs refer to "the most popular blogs as determined by number of inbound links, a practice that also underlies the identification of top-100 lists of blogs posted to the web." The study found that the level of blogging technologies has a significant relationship with weblog success, which was measured as weblog popularity.

[Safran, Kappe 08] conducted a quantitative analysis on a community of approximately 15,000 weblogs. The researchers considered the number of weblog visitors as an indicator of weblog success in readership perspective. Nine hypotheses in three categories (weblog content types, community activities and writing style) were tested and all of them found to be correlated to the number of weblog visitors.

Even though past two studies reflect the social features of weblogs to weblog success by introducing readership viewpoint, the measures they used have some shortcomings. First, popularity 'ranking' has a weakness to be a proxy for weblog success: the difference in ranking between the top-first weblog and the tenth popular weblog is nine. But, the gap in the weblog rankings does not accurately proportional to the weblog success. In other words, one cannot say that the most popular weblog is ten times as successful as the tenth popular one. On the other hand, the number of visitors to a weblogs is an analogue, continuous value. And it can reflect how successful the weblogs are more precisely. In addition, this study suggests the use of the number of unique visitors to a weblog as a measure for its success. Weblog visits result from (a) access to the posts or (b) personal networks with the weblog owner. To

focus on the weblog success from the view point of weblog content (as presented in Figure 1.), the impact of (b) needs to be removed. Therefore, the number of unique visitors is introduced to eliminate multiple visits of a user for a week; that is, all hits by one user were counted as one unique visit. We propose that the number of unique visitors is an appropriate indicator of weblog success because it excludes repeated visits of a single user who may have a personal connection with the blog owner.

2.2 WSPs' (Weblog Service Providers') Role in Content Creation and Weblog Success

We argue that the level of WSPs' support for weblog content creation is differentiated by the type of weblogs. In this section, we first look at the classification of weblogs types then explain the relationship between WSPs' role in post creation and weblog types.

Korean IT columnist, Joongtae Kim, categorized weblogs into membership and installation types [Kim 04]. A person can use a membership-type weblog service by simply joining the membership of weblog sites. In installation-type weblogs, the weblog owner buys server space and installs blogging tools and software. [Kim 04] then divided membership-type weblogs into three types: specialized; portal; corporate weblogs. Specialized weblog services are offered by WSPs that deal only with weblog services while portal weblogs are serviced by WSPs that deal with other kinds of services as well. A corporate weblog is defined a blog operated and owned by a company, but users can join the membership and create contents related to the company's products and services.

In this paper, we try explore the success factors of weblogs which are owned and operated not by firms but by individual bloggers. Hence, we modified the weblog types defined by [Kim 04] and categorized individual weblogs into three types: portal; specialized; self-type. Portal weblogs and specialized weblogs are defined as mentioned earlier, following [Kim 04]. A self-type weblog corresponds to an installation-type weblog defined by [Kim 04].

Portal weblogs are provided through portal web service providers, such as www.naver.com or www.daum.net. A portal WSP provides information and/or resources (which are required for its authors' weblog content creation) via its portal web service (for example, via news and dictionaries). This type of WSP also enables its bloggers to access other weblogs or weblog contents in the same domain. Unlike portal weblogs, specialized weblogs are provided by WSPs exclusively involved in weblog services rather than other web services. Self-type weblogs are created and operated by an individual blogger who does not receive any support help from either portal or specialized weblog companies.

To create a weblog without any help from WSPs, a blogger needs to make considerable efforts: he/she may struggle to purchase server space and a domain address, for instance. In addition, the blogger either writes code or uses software to design the weblog. Finally, he/she cannot make use of the resources on other WSP's domain. On the other hand, a blogger of a specialized weblog simply joins the membership and receives support: he/she can get access to the contents of other weblogs in the same WSP domain and can modify those resources directly in his/her posts. The author gets free server space and free blogging tools, too. In addition to the supports received from a specialized WSP, the blogger can obtain various kinds of

online resources (e.g., news, the portal encyclopaedias, search results, etc.) by joining a portal-site membership.

In brief, a portal WSP provides the widest and deepest support for weblog content creation to its users among the three types of weblogs. A self-type weblog receives the least support. Table 1 summarizes our definition and the characteristics of those three types of weblogs.

Type	Description	Degree of WSP support
Portal-type	– service provided by portal weblog service providers – other portal services are also provided by service providers	very high
Specialized-type	– service provided by weblog-only service providers – relatively limited support is offered (compared to portal-type) by service providers	high
Self-Type	– blogger manages all process of content creation and weblog operation	low or none

Table 1: Weblog Types and Level of Support in Content Creation

2.3 Content Features and Weblog Success

Weblog success in terms of the number of visitors may be affected by the quantity or the quality of weblog content. Suppose that a blog reader wants to get some information on restaurants for an important business meeting. When he/she searches for possible choices, the quantity or the quality of weblog content may give signals to a blog reader that a weblog with more volume or better quality is more trustworthy. In fact, many researchers also have noticed the importance of weblog content features such as quantity and quality of posts in weblog success.

In terms of the quantity of weblog content, [Safran, Kappe 08] noted that the number of posts in a weblog is important in determining weblog success. From a quantitative analysis on a community of approximately 15,000 weblogs, the study found a correlation (0.60) between the quantity of weblog content and weblog success. The number of posts generated during their six-week studies in each weblog was used to reflect the quantity of weblog content. And the number of visitors to each weblog during their study period was used as a measure for weblog success.

Regarding the quality of weblog content, the number of the comments attached on a certain post is often used. [Marlow 04] argued that comments are reader responses to certain posts in websites and the attached comments characterize basic forms of the social use of weblogs. The study discussed that comments could be indicators of usefulness. Highly popular web sites have considerable amount of responses (i.e., comments of uncountable pages) while less popular web sites have fewer comments implying that a readership does not exist for those weblogs. For this reason, comments show reader evaluation of certain posts as indicators of the quality of weblog contents. More comments on a post show the enthusiasm of other users for the post. Large number of attached comments also implies that the quality of a weblog contents is high. [Safran, Kappe 08] found a correlation (0.77) between the number of comments received and weblog success measured as the number of visitors to each weblog during their study period.

2.4 BTs' (Blogging Technologies') Role in Content Diffusion and Weblog Success

Several studies on weblogs have emphasized the role of blogging technologies such as trackback and RSS in weblog content diffusion and weblog success.

According to [Blood, 04], a trackback is an automated reverse link from a reference post/weblog to a referrer post/weblog. [Marlow 04] also noted that the trackback system is a form of automatic communication. Thanks to this automatic referencing system, when a blogger refers to another blog, the writer of the cited blog is notified that the post is being referenced. In addition, readers of either blog can see the linkages to all related trackbacked blogs. [Lee et al. 07] explained that the trackback system enhances the diffusion of weblog contents: it notifies both the owner of the original contents and readers of the modified contents when a modified weblog post is written and refers original ones to it. As a result, adoption of a trackback system gives opportunities for diffusing the weblog contents widely.

RSS is an abbreviation for Really Simple Syndication. This blogging technology allows automatic, regular subscription of new posts in a weblog and thus helps diffusion of posts. [Lee et al. 07] claimed that "Combined with personal relationship-oriented features of Korean portals, RSS system contributed bloggers who have similar tastes or perspectives to diffuse the contents selectively within themselves."

[Du, Wagner 06] classified weblogs by type I, II and III according to evolution levels of weblogs, then, examined the relationship between the level of blogging technologies used in weblogs and weblog success. The researchers found that the level of blogging technologies has a significant positive relationship with weblog success.

3 Our Approach

3.1 Model

Figure 2 illustrates our model for analysis. Based on the discussion in two previous sections, we argue that weblog success is a function of WSPs' support in content creation, content features, and BTs' support in content diffusion.

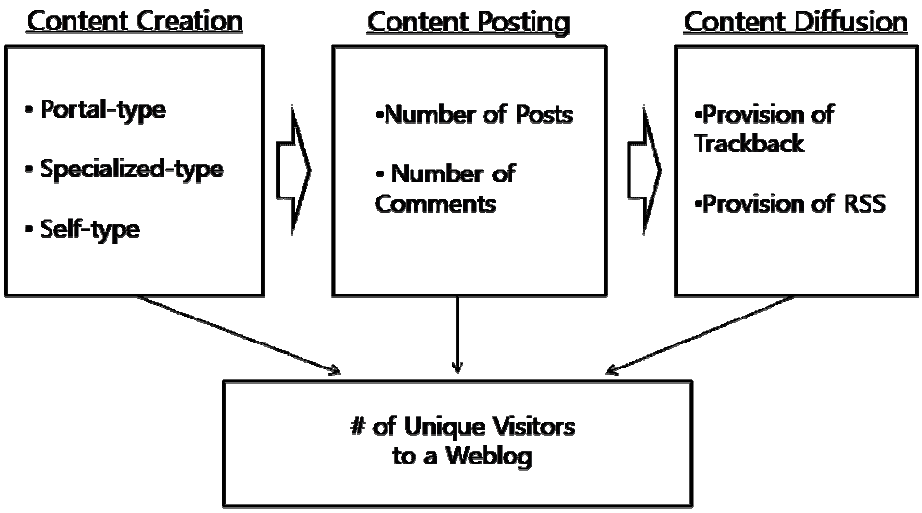

Figure 2: Weblog Success Model

As mentioned in section 2.1, the number of unique visitors to a weblog is chosen to represent its success. Three weblog types are used to reflect WSPs' level of support in content creation process as discussed in section 2.2. Similarly, the number of posts per week and the number of comments attached to those posts are proxies for weblog features (i.e., the quantity and the quality of weblogs). And two blogging technologies, trackback and RSS are indicators of BTs' support for content diffusion. Hence, weblog success model is summarized as:

$$\begin{aligned} \#\,of\ \ Unique\ \ Visitors = f(&Web\log\ Type,\\ &\#\,of\ \ Posts, \#\,of\ \ Comments\ \ Attached,\\ &Trackback,\ Rss) \end{aligned} \quad (1)$$

3.2 Methodology

In this paper, the Heckman two-step estimation is used to handle possible sample selection bias in the data (lack of the number of unique visitors in some of the sample). Sample selection bias occurs when the selected samples are non-randomly censored (i.e., missing data problem exists). [Heckman 76] and [Heckman 79] proposed a two-step estimation procedure to remove specification error resulted from the missing values of omitted variables. Equation (2) and (3) represent the first and the second step of estimation procedure.

Stage 1:

$$y_{1i} = x_{1i}^{'}\beta_1 + \varepsilon_{1i} \qquad \begin{cases} y_{1i} = 1 \ \text{ if } y_{1i}^{*} > 0 \\ y_{1i} = 0 \ \text{ if } y_{1i}^{*} \le 0 \end{cases} \quad (2)$$

Stage 2:

$$y_{2i} = x_{2i}^{'}\beta_2 + \varepsilon_{2i} \qquad y_{2i} = y_{2i}^{*} \text{ only if } y_{1i}^{*} > 0 \tag{3}$$

Using this methodology, one can obtain unbiased and consistent estimates when sample selection bias may exist as [Heckman 76] and [Heckman 79] noted. In the first stage, the probit choice model is used: the dependent variable in this stage is the availability of the positive dependent variable in the second stage, which is the real dependent variable researchers are interested. During the first stage one can estimate the inverse Mills ratio (λ). The inverse Mills ration (λ) is for eliminate the selection bias for censored dependent variable. In the second stage, Ordinary Least Squares (OLS) estimation for positive dependent-variable samples is conducted with the additional explanatory variable λ (i.e., the inverse Mills ratio which was derived from former stage).

3.3 Data and Variables

To concentrate on the factors affecting weblog success, we limited our sample to the top 200 weblogs owned by individual bloggers. Each week, [Daum Directory Search 08] provides rankings of approximately 1,400 Korean weblogs counted by 1,300,000 anonymous panels and this ranking was used to identify our sample. To decide the ranking of weblogs, [Daum Directory Search 08] integrates the number of visits of each panel to each weblog, time spent on each blog. Based on this ranking, data were collected every week for four weeks from April 6 to May 3, 2008. Collected items include: the number of unique visitors of the top 200 weblogs; type and age of each weblog; the number of created posts for each week during study period; the number of comments attached to those posts and categories in each weblog. In addition, we observed whether advertisement program, trackback or RSS systems were adopted. Except for the number of unique visitors offered by [Daum Directory Search 08], the data were collected by eight hired students.

After four weeks' observation, we could collect data of 800 weblogs (top 200 weblogs*4 weeks). Some of them appeared from two to four weeks while the rest appeared only once during our study period. According to [Du, Wagner 06], the maturity of weblogs can affect weblog success. In that sense, the use of the cumulative number of unique visitors to weblogs may reflect how maturity works for weblog success. However, the information on the cumulative number of unique visitors to weblogs was not included in our data collection due to the data unavailability from [Daum Directory Search 08]: we could only get information on the weekly unique visits to weblogs in the sample. To resolve this maturity problem, the number of weekly unique visitors was selected as a dependent variable.

After tracking same weblogs that appeared more than twice (maximum four times), we could not see any significant changes in weblog type as well as in the adoption or elimination of new blogging tools such as trackback or RSS. Considering the short sample period, we assumed that the impacts of time or changes in weblog environment are negligibly small. Hence, we assumed that the data collected from identical weblogs in the four weeks are independent of each other. Therefore, we conducted data pooling. Let us assume that Kim's weblog was put on the top 200

weblogs list three times during the observation period: the first, third and fourth week. The assumption of independence implies that the data obtained from Kim's weblog during the first week is assumed to be independent of the data obtained from Kim's weblog during the third week or the fourth week. The three sets of data obtained from Kim's weblog in three different weeks are regarded as three independent data sets in the sample. To exclude irrelevant weblogs from our sample, we removed weblogs not owned by individual bloggers as well as several simple web pages for advertisement purposes only and not for blogging purposes. Our final sample included 781 weblogs.

Table 2 illustrates the definition of variables. As discussed in Section 2.1, the number of unique visitors per week is selected as an indicator of weblog success. If we continue to make use of Kim's weblog example above, due to the assumption of independence of weblogs and data pooling process, the number of unique visitors during the first week is independent of the number of unique visitors during the third week or the fourth week. Three observations of the number of unique visitors to Kim's weblog are handled as three different data sets in the sample. By introducing the weekly unique visits to weblogs as a dependent variable and conducting data pooling, we tried to reduce the possible effects of the weblog maturity on weblog success. We used the logged value of the number of unique visitors per week, LNUV, because collected data figures vary from 5,280 to 308,357 in the sample while the values of independent variables do not vary largely. When a weblog is self-type, both weblog type variables (PORT, SPEC) are zero. By introducing these dummy variables, we can capture the relative effects of support from two weblog types on LNUV while self-type weblog is set to be the reference: no specific support from WSP is offered to self-type weblogs.

For analysis, three control variables are included: age of each weblog in months (AGE), the number of posting categories in a blog (CATE) and a dummy variable for advertising program (ADS). [Du, Wagner 06] used the age of weblogs as a control variable and conducted chi-square tests between age groups (a group of weblogs four years old and older vs. a group of those younger than four years). To exclude influence of maturity on weblog success, we also included blog age (AGE) as a control variable in the first stage of the Heckman two-step estimation. To eliminate the possible impact of subject diversity of weblog contents on the number of unique visitors (LNUV), the number of posting categories (CATE) in a weblog was included. In other words, CATE intends to remove the likely effects of segmented and itemized presentation of weblog contents on weblog success. Authors can classify their posts into several themes and present them into various categories. When the subjects in a weblog are specific with little variation, the content variety is small. However, if the subjects of a weblog are not limited to a certain area, the variety in contents subjects is wide. Advertisement programs such as Ad Sense of Google are attachable to weblogs. When a blog reader visits one's weblog, the existence of advertisement may make such weblog less attractive. The value of ADS is 1 if a weblog have installed advertisement program and 0 otherwise.

	Variables	Description	Related Studies
Weblog Success	LNUV	– dependent variable – logged value of the number of unique visitors per week	[Du, Wagner 06], [Safran, Kappe 08]
WSP's Support Level in Weblog Content Creation	PORT	– explanatory dummy variable – 1 if a blog is portal type, 0 otherwise	[Kim 04]
	SPEC	– explanatory dummy variable – 1 if a blog is specialized type, 0 otherwise	[Kim 04]
Content Features	QUAN	– explanatory variable – the number of created posts per week	[Safran, Kappe 08]
	QUAL	– explanatory dummy variable – the number of attached comments to 'QUAN'	[Safran, Kappe 08], [Marlow 04]
BT's Support in Weblog Content Diffusion	TRAC	– explanatory dummy variable – 1 if a blog offers trackback to its visitors, 0 otherwise	[Du, Wagner 06] , [Lee et al. 07]
	RSS	– explanatory dummy variable – 1 if a blog offers RSS to its visitors, 0 otherwise	[Du, Wagner 06] , [Lee et al. 07]

Table 2: Summary of the Variables

3.4 Estimation

During our study period, [Daum Directory Search 08] yielded the unique visitor numbers for only 381 weblogs among the 781 blogs in the sample. Missing unique-visitor data for lower-ranked weblogs in the sample results from the ranking calculation method of the [Daum Directory Search 08]. In case that only small number of their panel visits a weblog, the company's software is less likely to estimate reliable statistics and [Daum Directory Search 08] did not provide unique-

visitor data for those weblogs. Hence, our data is censored, and we used a Heckman two-step estimation technique to handle possible selection bias [Heckman 76], [Heckman 76]. The estimation of the study is based on the two equations stated below:

Stage 1 :

$$UVSE = C + \alpha_1 PORT_i + \alpha_2 SPEC_i + \alpha_3 QUAN_i + \alpha_4 QUAL_i + \alpha_5 TRAC_i + \alpha_6 RSS_i + \alpha_7 CATE_i + \alpha_8 ADS_i + \alpha_9 AGE_i + \varepsilon_s \quad (4)$$

$$where\ UVSE = 1\ if\ LNUV > 0,$$

$$UVSE = 0\ otherwise$$

Equation (3) is used in all models. It was used in all of the four models to calculate the inverse Mills ratio, which was included in the second stage of all models.

Stage 2 :

$$LNUV_i = C + \beta_1 PORT_i + \beta_2 SPEC_i + \beta_3 QUAN_i + \beta_4 QUAL_i + \beta_5 TRAC_i + \beta_6 RSS_i + \beta_7 CATE_i + \beta_8 ADS_i + \beta_9 \lambda + \varepsilon_i \quad (5)$$

where λ is the inverse mills ratio from the Probit *estimation at the first stage*

4 Results of the Analysis

4.1 Basic Statistics

In previous section, the data, the definition of variables and our model for the empirical study were explained. Table 3 presents the basic statistics of the variables in the sample. According to [Hardy 93], the use of the descriptive information could be valuable to observe the distribution of dummy variables. [Hardy 93] explains that the mean value of a dummy variable accounts for the number of cases with value 1. Following the definition of a dummy variable, the value of PORT is 1 and SPEC is 0 if a weblog is portal-type. Similarly, the value of PORT is 0 and SPEC is 1 if a weblog is specialized-type. When a weblog is self-type, the value of PORT and SPEC are both zero. In our sample, 118 weblogs were portal-type and 170 weblogs were specialized-type while 93 weblogs were turned out to be self-type among 381 weblogs. And this distribution of the weblog types also appears via the mean value of two dummy variables on weblog types: the mean values of PORT and SPEC inform us that approximately 31% and 45% of weblogs among 381 samples are portal-type and specialized-type. The rest (i.e., around 24% of weblogs) are self-type. So, the distribution of weblog-type seems relatively even. Similarly, the mean values of TRAC and RSS show the distributions of the weblogs which adopted trackback and RSS technologies. Among 381 weblogs, 356 weblogs (approximately 93% of the sample) provided trackback features while 318 weblogs (approximately 83% of the

sample) provided RSS features to their visitors. As Table 3 shows, the mean values of TRAC and RSS are very close to 1, which imply that most of top-ranked weblogs (top 200 ranked weblogs during the study period were selected as samples) offer blogging technologies supporting content diffusion. Hence, these two variables may not be very effective to explain common success factors in our sample. In the mean time, 240 weblogs among 381 weblogs (approximately 63% of the sample) adopted advertising programs and the mean value of ADS represent this distribution clearly.

Variables	Mean	Max	Min	St.dev
LNUV	9.69	12.64	8.57	0.73
PORT	0.31	1	0	0.46
SPEC	0.45	1	0	0.50
QUAN	22.72	336	0	48.73
QUAL	96.05	2920	0	194.55
TRAC	0.93	1	0	0.25
RSS	0.83	1	0	0.37
AGE	18.05	131	2	16.62
CATE	21.08	279	0	40.70
ADS	0.63	1	0	0.48

Table 3: Basic Statistics obtained from 381 Weblogs with Unique Visitor Data

4.2 Estimation Results

Estimated from the first stage of the Heckman two-step estimation (probit model), the inverse Mills ratio was significant. This implies that the data used in the estimation may show sample selection bias. We controlled this bias by using Heckman two-step estimation method.

Table 4 presents the estimates of the regression in the second stage of the Heckman two-step procedure. The two dummy variables, PORT and SPEC are both significant and positive. The dummy variable PORT captures the 'relative' effect of the portal weblog service providers' support in content creation on weblog success: the base level for comparison is the self-type. In similar way, SPEC captures the 'relative' effect of the specialized weblog service providers' support in content creation on weblog success. The value of SPEC is 0.160: specialized weblogs in the sample have approximately 16% more unique visitors than the self-type weblogs. The value of PORT is 0.244, which shows that portal weblogs have around 24% more visitors than the self-type weblogs in our sample. This suggests that more support from WSPs translates to more unique visitors, which in turn indicates weblog success. Hence, the importance of WSPs' support in weblog success is proved.

The coefficient of QUAN is negative (-0.003) and significant. The number of posts generated weekly denotes the quantity of weblog contents. Hence, an increase in quantity of weblog contents per week makes weblog less successful. [Safran, Kappe 08] reported a positive correlation between the number of posts generated and weblog visits; however, our estimation results suggest that the number of posts generated each week does not contribute to an increased number of unique visitors. In the mean time,

the coefficient of QUAL is positive (0.001) and significant. Earlier, [Safran, Kappe 08] reported a positive correlation between the number of comments received and the number of visitors. Our estimation results further suggest that the quality of weblog content contribute to weblog success.

TRAC and RSS have positive signs as expected but insignificant. Although [Lee et al. 07] mentioned that trackback and RSS help the diffusion weblog contents, our results show that one cannot definitely say that these two systems contribute to weblog success.

Variables	**Coefficients**
PORT	0.244 (2.07) *
SPEC	0.160 (1.68) †
QUAN	-0.003 (-2.94) **
QUAL	0.001 (2.55) *
TRAC	0.098 (0.56)
RSS	0.134 (1.30)
CATE	-0.006 (-0.47)
ADS	0.080 (0.87)
Λ (Inverse Mills Ratio)	-0.597 (-2.99)**
Sample size	381
Wald $\chi^2_{(14)}$	126.16

Table 4 : Estimation Results
(Note: The t values are in parentheses. Variables are defined in the text.
*†p<0.1; * p < 0.05; ** p < 0.01; *** p < 0.001)*

Figure 3 summarizes the estimation results which correspond to our weblog success model. When analyzed the success factors in top-ranked weblogs, WSPs' support in content creation process proved to be effective. In terms of content features, only the quality of weblog content was helpful to bring success while the quantity of weblog content disturbs weblog success. And BTs' contribution in content diffusion is not fruitful to weblog success.

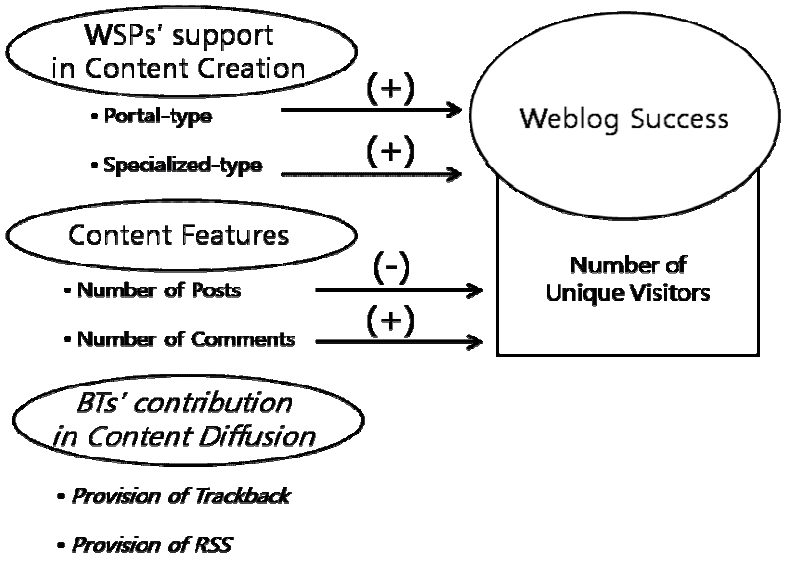

Figure 3: Summary of the Model Estimation Results

5 Discussion and Conclusions

[Du, Wagner 06] mainly considered success factors in term of blogging technologies. [Safran, Kappe 08] focused on success factors in terms of content feature (e.g., the number of posts created or comments attached). Our study investigated both sides together; in addition, we introduced WSPs' support in content creation as weblog success factors.

The estimation results indicate that the WSPs' support is more effective to bring success to weblogs, compared to no or very low support represented by self-type weblogs. In addition, the value of PORT is considerably larger than that of SPEC. This implies that the support of portal WSPs attracts more unique visitors than that of specialized WSPs. Most specialized weblog companies currently provide support in content creation by allowing their users compatibility of others' posts in same domain. Most portal WSPs offer broader range of resource pools for content creation: accessibility and compatibility with news, dictionaries, and online shopping information are also provided to the users. The big difference in coefficients between the two dummy variables (PORT and SPEC) seems to reflect the level of support offered by WSPs in two weblog types. Hence, WSPs will have a greater chance to make their users successful bloggers by providing deeper and wider support in content creation. To have a competitive advantage over others WSPs, full range of support —as much as possible— will be desired. In this perspective, our results have broadened the discussion in theories of weblog success factors to WSPs' role in weblog content creation.

When separately examined, the quantity of posts and the quantity of comments were both positively correlated to the number of visitors [Safran, Kappe 08]. But, considered together, the quantity of posts was inversely related to weblog success while the number of comments was significantly and positively contributed to weblog success. From this result, another suggestion can be made: in order to have more successful weblogs, a WSP should promote its users to improve the quality rather than the quantity of weblog content. For higher quality of posts, offering incentives to users could be a solution. For example, a WSP could give free background music to its users satisfying the certain quality level in weblog contents. Or, it can introduce cyber money grant to those who received most recommendations or those who recommended good posts to others most, for instance.

With all these findings, this study has some limitations. First, the success measure used in this study restrictively represents the 'readership' side of weblog success. As [Du, Wagner 06] mentioned, weblogs are powerful tools which enable people to freely express and exchange their personal opinions, knowledge and information. In this sense, weblogs serve their owners and users as social media: compared to the classical media such as newspapers or televisions, activities in the weblogs are two-way processes. Successful weblogs have wide readership as popular classical media do. In addition to this powerful readership, activities in the weblogs involve interaction between the authors and readers: collaborative processes are unavoidable during production and diffusion of weblog contents (see [Quiggin 06]). In spite of the interactive features of weblogs, the weblog success measured by the number of weekly unique visitors could not reflect the 'social interaction' side of weblog success. Second, one cannot get needed information on the relationship between weblog maturity and weblog success: as mentioned in Section 3.3, the introduction of weekly unique visits only reduces the possible impact of weblog maturity on weblog success. In addition, the visits to weblogs are due to various purposes: To access successful weblog contents; reinforcement of personal relations. To measure the sole impact of weblog content on weblog success, other measures can be selected. For example, the number of inbound links or total time spent on a weblog by all visitors could represent weblog success. Yet, these measures can be biased to calculate genuine, unbiased weblog success: the total time spent on a weblog would be biased as one could just open several web pages together and does not actually concentrate on certain weblog content. In this sense, we believe that the number of unique visitors per week is an appropriate measure. As all hits by one user are counted as one visit, unique-visitor method excludes any repeated visits of identical blog reader.

This study explored weblog success using relatively short-term data sets and considered readership side of weblog success. For further advancement of weblog success studies, this research opens several areas to be explored. First, we may analyze weblog success regarding the social interaction point of view as well as readership perspective: the strength of the social ties in weblog activities, etc. Second, we may investigate how maturity of weblogs affects the success of weblogs in the long term: WSP's support may be less effective to bring weblog success in the long term investigation; maturity itself, or reputations built on during the life of a weblog may affect weblog success. Third, we may expand our model to activities related to the participants' genuine features (e.g., the characteristics of blog owners and/or blog readers) as well as activities related to weblog contents.

Acknowledgements

This research was supported by the KCC (Korea Communications Commission), Korea, under the CPRC (Communication Policy Research Center) support program supervised by the IITA (Institute for Information Technology Management) (IITA-2009-C1091-0901-0003).

References

[Blood, 04] Blood, R.: "How Blogging Software Reshapes the Online Community"; Communications of the ACM, 47, 12 (2004), 53-55.

[Daum Directory Search 08] Daum Directory Search: (2008), http://directory.search.daum.net/site_list.daum?dirseq=169938 (Data collected from April 6, 2008 to May 4, 2008). (In Korean)

[Du, Wagner 06] Du, H. S., Wagner, C.: "Weblog Success: Exploring the Role of Technology"; International Journal of Human-Computer Studies, 64, 9 (2006), 789-798.

[Garcia, Hackbarth 08] Garcia, A. E., Hackbarth, K. D.: "Approximation to a Behavioral Model for Estimating Traffic Aggregation Scenarios"; Journal of Universal Computer Science, 14, 5 (2008), 731-744.

[Guo et al. 08] Guo, X., Vogel, D., Zhou, Z., Zhang, X.: "Chaos Theory as a Model for Interpreting Weblog Traffic"; Proceedings of the 41st Hawaii International Conference on System Sciences (2008), 1-9.

[Hardy 93] Hardy, M. A.: "Regression with Dummy Variables"; Sage, CA (1993).

[Heckman 76] Heckman, J. J.: "A Life-cycle Model of Earnings, Learning, and Consumption"; Journal of Political Economy, 84, 4 (1976), S9–S44.

[Heckman 79] Heckman, J. J.: "Sample Selection Bias as a Specification Error"; Econometrica, 47, 1 (1979), 153-161.

[Herring et al. 05] Herring, S. C., Kouper, I., Paolillo, J. C., Scheidt, L. A., Tyworth, M., Welsch, P., Wright, E., Yu, N.: "Conversations in the Blogosphere: an Analysis from the Bottom Up"; Proceedings of the 38th Annual Hawaii International Conference on System Sciences (2005), 1-11.

[Kim 04] Joongtae Kim: "Comparison of Membership-type and Installation-type Weblogs"; (2004), http://www.dal.kr/blog/archives/000597.html. (In Korean)

[Kozlak et al. 08] Kozlak, J., Dobrowolski, G., Kisiel-Dorohinicki, M., Nawarecki, E.: "Anti-Crisis Management of City Traffic Using Agent-Based Approach"; Journal of Universal Computer Science, 14, 14 (2008), 2359-2380.

[Lee et al. 07] Lee, H., Jeong, E., Lee, J.: "Social Diffusion Paths for Digital Contents in Web 2.0 era"; Research Report, Korea Information Society Development Institute, (2007). (In Korean)

[Li, Lim 08] Li, M., Lim, S. C.: "Modelling Network Traffic Using Generalized Cauchy Process"; Physica A, 387, 11 (2008), 2584-2594.

[Li, Zhao 08] Li, M., Zhao, W.: "Detection of Variations of Local Irregularity of Traffic under DDOS Flood Attack"; Mathematical Problems in Engineering, 2008 (2008).

[Marlow 04] Marlow, C.: "Audience, Structure and Authority in the Weblog Community"; Proceedings of the International Communication Association Conference, New Orleans, LA, (2004), 1-9.

[MIC 07] Ministry of International Affairs and Communications of Japan: "White Paper on Information and Communications in Japan"; (2007), http://www.johotsusintokei.soumu.go.jp/whitepaper/eng/WP2007/2007-index.html.

[NIDA 08] National Internet Development Agency of Korea: "Survey on Internet Usage"; (2008), http://isis.nida.or.kr/eng/board/index.jsp?pageId=040100&bbsId=10&itemId=304&pageIndex=1.

[OECD 06] OECD: "Information Technology Outlook"; (2006), http://www.oecd.org/document/10/0,3343,en_2649_33757_37486858_1_1_1_1,00.html.

[OECD 08] OECD: "Information Technology Outlook"; (2008), http://www.oecd.org/document/20/0,3343,en_2649_33757_41892820_1_1_1_1,00.html.

[Paquet 03] Paquet, S.: "Personal Knowledge Publishing and its Uses in Research"; (2003), http://www.knowledgeboard.com/item/253.

[Pew 07] Pew: "Teens and Social Media"; (2007), Pew Internet and American Life Project.

[Quiggin 06] Quiggin, J.: "Blogs, Wikis and Creative Innovation"; International Journal of Cultural Studies, 9, 4 (2006), 481-496.

[Rosenbloom 04] Rosenbloom, A.: "The Blogosphere"; Communications of the ACM, 47, 12 (2004), 31-33.

[Safran, Kappe 08] Safran, C., Kappe, F.: "Success Factors in a Weblog Community"; Journal of Universal Computer Science, 14, 4 (2008), 546-556.