

# Analysis of Social Network Service Data to Estimate Tourist Interests in Green Tour Activities

**HyungChul Rah**

Graduate School BigData Convergence  
Chungbuk National University, Cheongju 28644, South Korea

**Sungho Park, Miok Kim, Youngbeen Cho**

Agricultural Bigdata Division  
Rural Development Administration, Jeonju-si, 54875, South Korea

**Kwan-Hee Yoo**

Department of Computer Science  
Chungbuk National University, Cheongju 28644, South Korea

## ABSTRACT

*Social network service (SNS) data related to green tourism were used to estimate preferred tour sites and users' interests. Keywords related with green tour activities were employed to search the SNS data. SNS data were collected from Korean blogs such as Naver and Daum from June 1<sup>st</sup> to August 31<sup>st</sup> between 2015 and 2017 using text-mining solution. During the study period, seven hundred and five posts were analyzed. Associated words that frequently co-occurred with keywords were classified into different categories depending on the nature of associated words. Associated words included swimming pools and camping sites (location); experience and swimming pools (attribute); and water play and culture (culture/leisure). Our data suggest that SNS users with experience of green tourism in Korea exhibited interest in green tourism with swimming pools, camping sites, experience, water play and/or culture rather than particular popular sites. Based on the findings, it is recommended that preferred facilities such as swimming pools should be provided at green tourism sites to meet the users' needs and to facilitate green tourism.*

**Key words:** Green Tour Activities, Social Network Service, SNS, Text-mining.

## 1. INTRODUCTION

Many tourists have used various social network services (SNS) to search tour information in advance, and to share their experiences with friends and families while traveling or after traveling by posting various information at social media [1]. It is not only exchange of new information on their tours but also sharing of happiness and fun from the tours and activities. SNS users are able to share their entertainment and fun through SNS channels, which allows one-to-multiple interaction. This means that sharing of interesting information about particular places can provide potential tourists with desire to visit as well as information about the places [1]. These a priori activities may influence potential tourists on decision-making [2]. Previously, it has been suggested that social media can play an influential

role as a communication channel to inform the values of green tour activities because tourists who experienced rural amenities want to share their experiences using various social media [3].

Considering that the purpose of tourism studies is to explore interests and desires of tourists, eventually in order to establish marketing promotion plans and to promote tourism industry, it is critical to understand main interests of tourists, information that tourists are willing to share, and tourist reviews or comments that are most influential to potential tourists in the future [4]. It will also be important to estimate tour sites or facilities that tourists prefer or want during their stay in order to promote tourism industry, especially green tour activities that are to be improved in Korea [1], [5].

In this paper, we analyzed Korean SNS data in order to identify 1) related words that most frequently co-occurred with green tour activities and their frequencies in social media, 2) preferred characteristics and locations by the tourists who experienced green tourism, 3) issues to be addressed in order to promote green tour activities by analyzing social network data.

---

\* Corresponding author, Email: [khyoo@chungbuk.ac.kr](mailto:khyoo@chungbuk.ac.kr)  
Manuscript received Apr. 06, 2018; revised Jul. 30, 2018;  
accepted Aug. 13, 2018

## 2. RELATED STUDY

Up to date, no studies have reported the use of SNS data in the field of green tourism. However, a few studies have been reported that utilized SNS data in other fields of tourism in order to understand current interests and needs of tourists in Korea. Lee et al [6] used SOCIALmetrics™ of Daumsoft, one of the text mining solutions for Korean SNS data, in order to analyze SNS data in tourism by using related word analysis and opinion mining analysis. Jung et al [7] used the same solution in order to examine the idea of tourists on railway tour by using Twitter and blogs. Kim et al [8] used Textom™ of the IMC, another text mining solution for Korean SNS data, in order to identify related words for temple stay in tourism. The authors drew temple stay-related words in the aspects of places and attributes, and suggested current tour trends and user preference of temple stay. Cho et al [9] collected tourism data related with Chungbuk province by crawling Korean news and blogs data and applied these data to promote tourism to Chungbuk province and to provide suggestions for policy making. Recently, Lee et al [10] analyzed Korean SNS data in order to investigate awareness of ski resort visitors by using one of the text mining solutions, Textom™. The authors estimated the interests of ski resort users in Korea based on the degree of relationship. Based on the literature review of related studies, similar method from other fields of tourism has been applied to green tourism in this study.

## 3. ANALYSIS Methods OF ASSOCIATED GATHERING DATA

The purpose of the study is to collect SNS data related with green tourism in Korea, and to analyze collected SNS data in order to estimate the interests of the users of green tour activities, and to provide recommendations to promote green tourism industry. Keywords related with green tour activities were used as search keywords among social media (blogs) to collect related posts. Social media data were collected by using a text mining solution for Korean SNS data, called SOCIALmetrics™ (Daumsoft, Seoul, Korea). The SOCIALmetrics™ solution contains crawlers for social media data, process text using natural language processing (NLP), text mining technologies, and application programmer's interface (API) for users' query (Fig. 1).

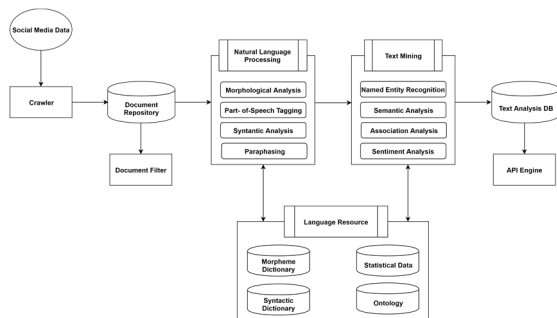


Fig. 1. Structure of SOCIALmetrics™ Social Big Data Mining Platform for crawling SNS data, natural language processing and text mining (modified from [11]).

The scope of data collection and analysis was limited to blogs in Korean such Naver blog (section.blog.naver.com) and Daum blog (blog.daum.net) which covers more than 90% of blog markets in Korea [12], [13]. In order to search SNS data related with green tourism, search keywords that are related with green tour activities, summer vacation and user review were used (Table 1). The NLP module divided input text into sentences, and segmented the word forms in each sentence into a string of morphemes (Fig. 1). The segmented morphemes were grouped into syntactic units via syntactic analysis. Once syntactic units were constructed, expressions denoting named entities such as people, locations, and organizations are recognized. Association analysis, then, was performed to identify tuples of search keyword with associated words. Associated words were grouped by category for map and trends of the words. Associated words most frequently co-occurred with search keywords were listed with the frequencies. Sentiment polarities for search keywords are determined through sentiment analysis. User reviews and comments for their experience in green tour activities were evaluated by using sentiment analysis.

Data were collected for the blog posts that were generated between 2015 and 2017 during the holiday season (from June 1, 2015 to August 31, 2015; from June 1, 2016 to August 31, 2016; from June 1, 2017 to August 31, 20167). During the study period, seven hundred five posts were collected for analysis.

Table 1. List of keywords used to search green tour activities from social media data

Category	Keyword
Green tour activities	green tourism, green village, rural tourism, rural traditional theme village, farm stay, tourist farm, farm tour, rural homestay
Summer vacation	summer vacation, vacation, family trip
User review	review, vacation reviews

## 4. ANALYSIS RESULTS

Associated words that most frequently co-occurred with green tour activities in social media communication of the users of green tour activities may reflect their interests. The associated words with the highest frequency of co-occurrence during the study period include 'experience', 'tour', 'rural area', 'summer', and 'variety' (Table 2). In order to identify places and attributes preferred by users of green tour activities, associated words were grouped into several categories such as location, attribute, and culture/leisure. In location category, associated words such as 'rural area', 'swimming pool', 'camping site', 'village', and 'valley' were the top five associated words that most frequently co-occurred with green tour activities during the study period (Table 3). In attribute category, associated words such as 'experience', 'summer', 'swimming pool', 'village', and 'nature' were most frequently (Table 4). In culture/leisure category, most frequent associated

words include ‘tour’, ‘summer vacation’, ‘water play’, ‘culture’ and ‘travel’ (Table 5).

Table 2. Top five associated words that most frequently co-occurred with green tour activities in social media during the study period (between 2015 and 2017)

2015		2016		2017		Overall	
Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.
Experience	139	Tour	80	Tour	60	Experience	271
Rural area	138	Experience	77	Experience	55	Tour	240
Summer	103	Rural area	46	Rural area	37	Rural area	221
Tour	100	Enjoy	33	Summer	31	Summer	134
Variety	91	Variety	33	Water play	22	Variety	124

Table 3. Top five associated words that most frequently co-occurred with green tour activities in location category during the study period (between 2015 and 2017)

2015		2016		2017		Overall	
Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.
Rural area	64	Rural area	46	Rural area	37	Rural area	147
Camping site	32	Village	33	Swimming pool	18	Swimming pool	61
Village	25	Valley	25	Water fall	13	Camping site	60
Valley	23	Swimming pool	22	Camping site	12	Village	58
Swimming pool	21	Camping site	16	Pension	11	Valley	48

Table 4. Top five associated words that most frequently co-occurred with green tour activities in attribute category during the study period (between 2015 and 2017)

2015		2016		2017		Overall	
Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.
Experience	85	Experience	77	Experience	55	Experience	217
Summer	44	Village	33	Summer	31	Summer	75
Nature	36	Program	27	Swimming pool	18	Swimming pool	61
Village	25	Facility	24	Nature	18	Village	58
Swimming pool	21	Swimming pool	22	Program	12	Nature	54

Table 5. Top five associated words that most frequently co-occurred with green tour activities in culture/leisure category during the study period (between 2015 and 2017)

2015		2016		2017		Overall	
Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.	Assoc. words	Freq.
Tour	64	Tour	80	Tour	60	Tour	204
Summervacation	41	Water play	21	Water play	22	Summervacation	75
Travel	21	Summer	17	Summervacation	17	Water play	61

Culture	20	vacation	11	Culture	13	Culture	44
Water play	18	Culture	11	Travel	10	Travel	42

As a result of sentiment analysis, average rate of positive words was 62%, and average rate of positive words was 7% suggesting almost two thirds of the user have positive opinion on green tour activities (Table 6). Most frequently negative words include ‘hot’, ‘difficult’, ‘inconvenient’, ‘crude’, ‘tired’, ‘complicated’ and so on. Original blog posts where negative words appeared were examined in order to identify issues of the users on green tourism. It turned out most of negative words were meant to everyday life in contrast to summer vacation, such as ‘hot’ from ‘hot summer weather’ and ‘complicated’ from ‘escape from complicated city life’ except ‘inconvenient’ from ‘inconvenient due to the small number of restrooms at green tourism sites.

Table 6. Sentiment analysis on green tour activities during the study period (between 2015 and 2017)

Opinions	2015	2016	2017	Average rate
Positive words	418 (64%)	276 (58%)	257 (66%)	(62%)
Negative words	50 (8%)	37 (8%)	23 (6%)	(7%)
Others	189 (29%)	161 (34%)	112 (29%)	(30%)
Total	657	474	392	

### 5. DISCUSSION

Social media users utilize social media in order to share their interests and experiences by posting various tour-related information as well as to gather tour information of their interests by chatting and searching information on SNS [1]. This study was performed in order to estimate tourist interests in green tour activities, preferred sites for green tourism, and issues to be addressed in green tourism of Korea by collecting and analyzing the social media data related to green tourism created in Korean language. Social media data have been used in tourism field to estimate the needs of tourists in various fields of tourism except green tourism [6]-[10]. These previous studies have used social media data, especially blog data because blog sites managed by individual users or consumers provides comparison of competitors in the tourism market so that their words of mouse on their blogs is regarded reliable and may influence many peer tourists [14]. Previous studies presented ways to collect and analyze SNS data in the different fields of tourism by using text mining solutions.

What we drew from our study include the followings in regard to preferences in green tourism (Table 7). These patterns have remained consistent during the study period between year 2015 and 2017. These findings are beyond our expectation because we expected to see more keywords related with popular tour sites from tourists’ blogs, which was not the case.

Table 7. Preferred characteristics and locations by the tourists who experienced green tourism that were found from the study

Category	Findings
Overall	Green tourism users are generally families with children at preschool or school based on the blogs.
Place	The families with kids were more interested in green tourism sites with swimming pools, camping sites, pension facilities, and clean restroom for their kids concerning place category.
Culture/Leisure	Green tourism users were also interested in green tourism sites where their kids can experience cultural activities, water play and camping site concerning culture/leisure category
Attribute	their interests included green tour sites with experiences and swimming pools concerning attribute category.

Regarding issues to be addressed in order to promote green tour activities, it is recommended to provide preferred facilities at green tourism sites such as swimming pools and restroom facilities to meet the users' needs of the green tour activities and to facilitate green tourism based on the users' reviews on blogs. Lack of hygiene facilities such as modernized restrooms have been an issue for rural tourism in the aspects of tourist [15].

## 6. CONCLUSION

We presented a use case that utilizes social media data to estimate interests of tourist in green tourism for the first time. Our data suggests that social media users who experienced green tourism in Korea are interested in green tourism sites with swimming pools and camping sites (location category); (cultural) experience and swimming pools (attribute category); water play and culture (culture/leisure category) rather than particular popular sites or places. This study may shine a light on blind spots on green tourism. However, our findings represent a part of green tourism users who shared their experiences on social media. It is necessary to carry out the follow - up studies to verify the findings by surveying unrepresented green tourism users.

## ACKNOWLEDGEMENT

This research was carried out partly with the support of the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2013-0-00881) supervised by the IITP (Institute for Information & communication Technology Promotion), and partly with the support of "Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ01283801)" Rural Development Administration, Republic of Korea.

## REFERENCES

- [1] S. Y. Ryu and S. W. Yoo, "A study on the perceptions and interests of tourists to Gangwon-do - Focusing on social media big data analysis," *International Journal of Tourism and Hospitality Research*, vol. 31, no. 2, 2017, pp. 63-81.
- [2] H. J. Jung and G. H. Lee, "Factors Influencing Online Travel Information Value and User's Continuous Intention in Social Media," *Journal of Tourism Sciences*, vol. 36, no. 5, 2012, pp. 289-308.
- [3] Y. J. Mun and J. H. Lee, "The Effects of Continued Usage Intention of Rural Tourism Social Media," *Journal of the Korea Academia-Industrial Cooperation Society*, vol. 12, no. 11, 2011, pp. 4812-4818.
- [4] H. B. Kim and H. S. Chang, "The effects of destination images after visit on tourists' attitude and word-of-mouth communication," *Journal of Tourism Sciences*, vol. 32, no. 3, 2008, pp. 209-229.
- [5] M. O. Kim, *Estimation of Preference and Value of Rural Amenities for Green Tour Activities -An Application of Mixed Logit Model*, doctoral dissertation, Chungbuk National University, Cheongju, 2011.
- [6] Y. J. Lee and J. H. Yoon, "A Study on Utilizing SNS Big Data in the Tourism Studies : Based on an Analysis of Key Words for Tourism Information Search," *International Journal of Tourism and Hospitality Research*, vol. 28, no. 3, 2014, pp. 5-14.
- [7] M. Jung, H. Rho, and M. Park, "Reaction survey of tourist who railway utilizing SNS bit data," *Proc. 2014 Autumn conference of Korean society for railway*, 2014.
- [8] B. K. Kim and M. K. Kim, "Keywords Network Analysis of Temple Stay Using Big-Data," *Journal of Tourism Sciences*, vol. 39, no. 5, 2015, pp. 27-40.
- [9] W. S. Cho, A. Cho, K. Kwon, and K. H. Yoo, "Implementation of smart chungbuk tourism based on SNS data analysis," *Journal of the Korean Data and Information Science Society*, vol. 26, no. 2, 2015, pp. 409-418.
- [10] J. M. Lee, J. H. Lee, and M. J. Kim, "A Study on Perception of Ski Resort Using Big Data Analysis," *The Korean Journal of Physical Education*, vol. 56, no. 4, 2017, pp. 415-430.
- [11] H. Woo, Y. Cho, E. Shim, K. Lee, and G. Song, "Public trauma after the Sewol Ferry disaster: The role of social media in understanding the public mood," *International journal of environmental research and public health*, vol. 12, no. 9, 2015, pp. 10974-10983.
- [12] H. Woo, *Development of Influenza Surveillance Model Based on Internet Search Query and Social Media Data*, doctoral dissertation, Graduate School of Public Health, Seoul National University, 2015.
- [13] H. M. Kim, "KOREA Online Marketing," 2014.
- [14] H. Soh, "Examining the Effects of Message Sidedness and Rewarded Consumer Referral in the Context of Blog Product Reviews," *Journal of Practical Research in Advertising and Public Relations*, vol. 5, no. 2, 2012, pp. 112-143.
- [15] D. B. Park and M. Lee, "Consumer Preferences of Rural Tourism Accommodations," *Journal of Rural Tourism*, vol. 13, no. 2, 2006, pp. 41-62.



#### **HyungChul Rah**

He is a visiting professor of Graduate School BigData Convergence at Chungbuk National University, Korea. He received the B.S. in veterinary medicine from KonKuk University, Korea in 1996, and received M.S. in forest management from Korea University, Korea in 2000, and M.P.V.M. in preventive veterinary medicine and Ph.D. in comparative pathology from Univ. of California, Davis, USA in 2001 and 2006, respectively. His research interests include big data analysis in agriculture and healthcare.



#### **Sungho Park**

He is a senior researcher at Rural Development Administration, Korea. He received the B.A., M.S., Ph.D. in agricultural economics from Seoul National University in 1990, 1999, and 2008, respectively. His research interests include application of agricultural big data in agricultural policy.



#### **Miok Kim**

She is a postdoctoral researcher at Rural Development Administration, Korea. She received the Ph.D. in agricultural economics from Chungbuk National University in 2011. Her research interests include application of agricultural big data.



#### **Yongbeen Cho**

He is a head in agricultural bigdata division at Rural Development Administration, and is a Ph.D. student of Department of Big Data in Chungbuk National University, Korea. He received the M.A. in agricultural education from Seoul National University in 2002. His research interests include application of agricultural big data.



#### **Kwan-Hee Yoo**

He is a professor who is working for Department of Computer Science at Chungbuk National University, Korea. He received the B.S. in Computer Science from Chonbuk National University, Korea in 1985, and also received M.S., Ph.D. in computer science from KAIST (Korea Advanced Institute of Science and Technology), Korea in 1988 and 1995, respectively. His research interests include a u-Learning system, computer graphics, 3D character animation, dental/medical applications.