Hashtag Trend Analysis based Travel Destination Recommendation Technique

Hyunsoo Hong, Sungwook Yoon, Hyenki Kim

Dept. of Multimedia Engineering
Andong National University,
Andong-City, Republic of Korea
e-mail: bazooka92@nate.com, uvgotmail@nate.com, hkkim@anu.ac.kr (Corresponding Author)

Abstract—The study explores a plan for the evaluation of travel destination preferences classified by emotional language by analyzing hashtags existing in image sets through the collection of Social Network Service, and by comparing with the vocabulary showing general travel satisfaction. The vocabulary to be evaluated was divided into regional, tourism target vocabulary, common expression language, and emotional expression language. This study demonstrates tourists' preference for travel destinations according to whether collected hashtags match previously retrieved emotional hashtags when tourists search for specific tourist sites. Through this, it will be possible to analyze the changing travel trends of a specific travel destination and accordingly, it will be possible to establish a travel promotion policy.

Keywords-hash-tag; trend; emotional; corpus anaysis.

I. INTRODUCTION

Through tags, most of the social network service (SNS) platforms tend to emphasize key words that express their writing and central key words. The role of keywords, categories, and metadata is expanded to express self-directed information discovery on digital documents, which is expressed as an additional expression with a hashtag. This resulted in the users of the existing metadata transferring from experts to general users. Through the hashtag processing, the folksonomy-based voluntary classification of ontologies allows various processes for corpus discovery of big data to be arbitrarily organized and used as search and analysis data [1]. The frequent use of hashtags in social network services is common to many platforms, and it is possible to intuitively perform statistical analysis, such as association analysis or frequency analysis between hashtags. Using this information, the promoter can use the hashtag to actively promote tourism information [2].

In general, hashtags are usually expressed by summarizing articles written by SNS users into noun-type keywords, and are responsible for categorizing keywords searched for information sharing.

The association analysis and the real-time pattern analysis of the promotion keywords are considered to be the means to make or strengthen tourism trends. This method of promoting tourist attractions is also used as a means of promoting tourism information to tourists. Municipal tourism information from municipalities with tourist attractions generally promote events, festivals, promotion of specific areas, Fam tour, and local companies, and it is intended to

spread through the use of not only official channels but also platforms, such as Facebook, Instagram, Twitter, Kacao Story, Band, and Blogs. Typically, these platforms are made through social connections with users and organizations, so they spread even faster through related tags.

The acquisition of information on tourist sites tends to be affected by the names of the sites and the keywords of the empirical contents posted on social platforms. The emotional and abstract experience language in the travel review, along with the name of the locality posted on social media will possibly affect the next trip destination of followers and friends exposed to the media [3].

However, since social media generally involves a rapid change in trends, it reflects the trend of recent time changes in real time. In general, the analysis of these trends reflects the tendency to change through folksonomically categorized tags. This reflected travel trend influences search preferences and strengthens or damages the emotional image along with the place's name. This research aims to provide effective promotional strategies for tourist information on targeted social network service tourist sites. Some methods are proposed, which would help to collect retrieved hashtags and select destinations in accordance with travel trends changing in real time. The study conducted research on the effective use of multiple hashtags as keywords for travel through classification and analysis.

The structure of this paper is as follows. Session 2 describes related studies, Session 3 describes tourism travel destination preference analysis plan and Session 4 describes the conclusion of the paper.

II. RELATED STUDIES

Articles posted on social networks with the subject of travel are always being uploaded from the point of view of tourism consumers, and the followers of digital connections recognize this information to be trends. Users who are interested in traveling can learn information about the desired travel destination through hashtag searches and can enhance the travel trend. A technique [4] was also proposed where, the more the travel destination name is gathered on Twitter, the more it is considered to be the travel destinations that people are likely to visit. In addition, there was a study [5] that analyzed the real-time SNS trends to recommend travel destinations and a study [6] on techniques of analysis of hashtags that appear simultaneously in the SNS environment. This study suggests the level of

recommendation according to the survey done on emotional vocabulary and the frequency analysis of existing hashtags. Unlike previous research, it is to grasp the preference through the emotional analysis through collective intelligence.

A. Folksonomy based Hashtag

Folksonomy is a new classification system based on 'tags' instead of 'directory' which is the traditional classification standard, and it means 'classification by people' (Folk + order + nomos). According to Wikipedia, "it is a newly coined word that represents the way members organize information together using freely chosen keywords" [7]. In addition to dividing into politics, economy and culture, for example, it is possible to gather information at a glance based on keywords, such as general elections and flower festivals. The difference between folksonomy and taxonomy, which is a classification system like existing book management, is that the members voluntarily organize unit information by giving meaning to individual information, and since the information is listed while interacting with the main members, it has the advantage of transmitting more meaningful information more accurately, and therefore it is in the same context as the hashtags on SNS [8]. Therefore, folksonomy is not as systematic compared to taxonomy. This is because users of the Internet are not making and learning a folksonomy system, but making it more convenient and useful. Thus, the cost of categorizing content can be reduced because there is no need to learn a complete classification scheme. Folksonomy is expandable and can change quickly depending on how users are categorizing through the Internet. Since folksonomy is the result of a collection of people's interaction, it is suitable for searches. The folksonomy-based hashtag also needs to be evaluated for changes in its trends and categories.

B. Status of Use of Promotion tags to Promote Municipal Government

In April 2017, the study conducted a questionnaire survey on the selection of promotional phrases related to Yeong-ju city to about 1,000 public officials who were all the subjects of public promotion training for Yeong-ju city.

Figure 1 shows it processed by word cloud according to frequency.

Figure 1.   Word cloud based on frequency of promotional keywords

In the top 40 keywords for promotion, 90% of language bundles, 5% of common language, and 5% of emotional corpus showed uneven usage combinations. The name of the place or specific sightseeing spots were mainly surveyed. Figure 2 below shows the results of the frequency analysis.

The results show that according to the classified keywords used by tourism promoters, there is a tendency to promote geographical names. (see Figure 2).

BuseokTemple	Sosuseowon_Confucian_Academy	apples	Sobaeksan_Mountain
105	104	72	65
the	ginseng	museom	Seonbichon
64	56	33	31
Punggi	Seonbi	museo	village
28	25	24	23
Insam	Floating_Stone	rayon	Korean_beef
21	20	20	16
town	apple	beef	BuseokTempl
16	15	15	15
of	decimal	in	Yeongju
15	13	13	11
Healing	si	bridge	forest
10	10	9	9
grapes	healing	log	Minority
9	9	9	9
Yangban	Confucian	Ginseng	Museum
9	8	8	8

Figure 2.   Promotion keywords based on survey

The hashtags used on SNS are typically used as multiple lists, and it is common to be listed mainly through the placement of the area name - the general corpus - the common corpus. This study aims to gauge the level of travel satisfaction by combining names of certain areas, common expression language containing the contents of tourist experiences, and corpora of folksonomical emotional expressions.

III.   TOURISM TRAVEL DESTINATION PREFERENCE ANALYSIS PLAN

A. Hashtag Classification Technique

This study proposes an analysis method that calculates a tourist travel destination preference index by using SNS data created by tourists on travel destinations and analyzing the frequency of specific tourism travel destination references and emotional analysis. The contents of the hashtag used as a keyword for promotion were classified into three categories in this study.

A list of words in the order of geographic, common, and conceptual words is shown in Figure 3.
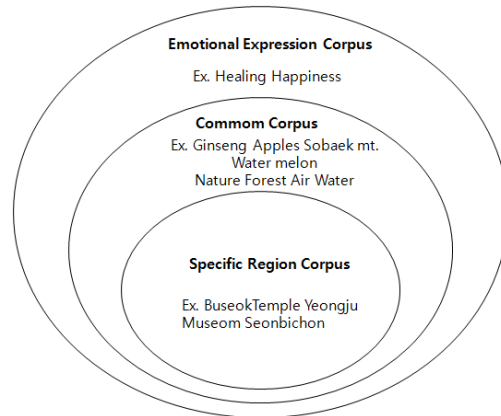

Figure 3.   Travel purpose and travel destination classification

They were divided into {specific region corpus, common expression corpus, emotional expression corpus}. In this classification, the study created a database with a real-time hashtag trend list of emotion expression corpus (Table 1).

TABLE I.   CORPUS-SPECIFIC PERSONALITY REGULATION

	Explanation	Example
Specific Region Corpus	Name of tourist destination, geographical name, tourist destination	BusekTemple Jeju
Commom Expression Corpus	Non-border language, such as natural environment	Mountain, Sea, Apples
Emotional Expression Corpus	Abstract word representing travel purpose	Healing Happiness Memories

B. Analysis Plan

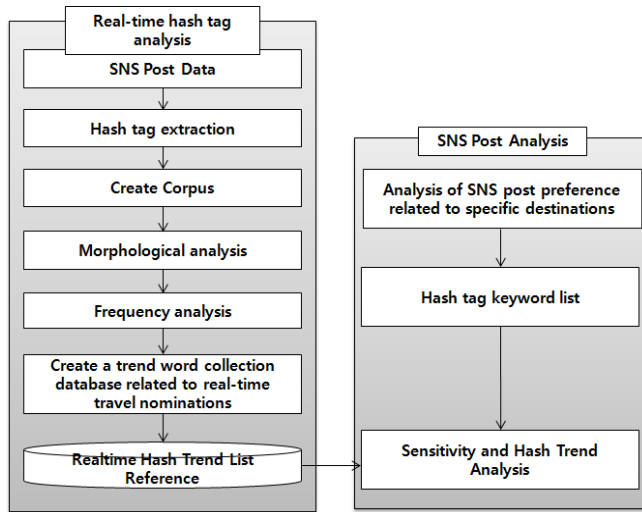The flow chart of the proposed method is shown in Figure 4.

Figure 4.   Flowchart of Research

First, collect hashtags from posted writings on SNS. Then, analyze morphemes through corpus analysis. Finally, collect related hashtags used in the same sentence. Analyze articles containing representative keywords related to various travel location names through hashtag collection of posts on social network services. The study uses Korean morpheme analyzer [9] to classify articles of social platforms, such as Instagram, Facebook, and Twitter into noun units and extract candidate keywords. Then, the term frequencies in which the keywords appear are listed in order. In this case, words not relevant to the research purpose are filtered by Stop Word. In addition, to identify the real-time trend area, words related to the region and emotional words are extracted. Finally,

compile only the keyword related to the area and list the ranking. The items with highest frequency have priority.

A name dictionary for tourism travel destination is constructed in advance, and based on this, SNS including a tourist travel destination name is extracted, and through the term frequency algorithm, the frequency of related keyword references within the SNS text of each travel destination is obtained. In addition, emotional words for sightseeing travel destinations are extracted as keywords, and a trend dictionary database of emotional vocabulary collected from posts about travel is created.

$$P(t) = F_t \left(\sum_{n=0}^{\infty} HT_n + V_n \right) \qquad (1)$$

Equation (1) is the preference index ($P(t)$). t is a tourist destination, Ft is the total number of t references, Ft is the reference frequency of t, n is the number of posts including t, HT is the matching ratio of the hashtag abstract language words including travel-related emotional expressions, and Vn represents the emotional probability of t.

C. Emotional word Evaluation

The emotional direction can be encoded by evaluating positive, neutral, and negative emotional words evaluated in the hashtag of the image set appearing when searching for a specific tourist destination [10]. In the emotion analysis, positivity indicates the positive direction and negative indicates the decrease in the preference in the negative direction. Emotional probability is calculated by extracting the SNS including the tourist destination and then using the term frequency algorithm to find the frequency of reference within the SNS text for each tourism destination. Emotional hash tags refer to tags or keywords that can define emotional probability through collective intelligence. In addition, the emotions expressed in the text are analyzed to reflect positive or negative opinions about the tourist destination. In this process, we use open Hangul API, which is an open platform that can be used in research to identify positive, neutral, and negative emotion words in sufficient number through collective intelligence [11]. The final tourist destination preference index is calculated by using the frequency of comments and emotional analysis results for each tourist destination. By evaluating the combination of hashtags created, it is possible to create a preference index for a specific region and adjust the direction of the related promotion policy.

D. Evaluation of Emotion Expressions Corpus according to Trend

The study collected emotional language according to trend through big data analysis and created a database with travel related emotion expression corpus (*HT*). If a match is found in the vocabulary used, such as hashtag searched in the travel destination, it can be evaluated for the purpose of following the trend and rated as preference. Also, it can be used to promote local events, festivals, and specific information in addition to the corpus of the local name

through collection of frequent hashtag changed in real time. Figure 5 is part of an example of keyword extraction for travel purposes in the proposed study with purpose specific travel destination classification technique [12].

Figure 5.   Emotional language examples

Alone Friendship Date Everyday Healing Friends Sea Vacation Backpacking Guesthouse Daily Family Romance Food porn Communication Weekend Camping Youth Walking Top-restaurant Happiness Man Memories Drive Parenting Itchy-feet Railroad Honeymoon I-love-you Alone Train-travel Nature

It is possible to search common corpus including emotional language that is investigated in real time and maintain it as database (HT) and use it as promotion keyword to improve search efficiency.

IV.   CONCLUSION AND FUTURE RESEARCH

This study analyzed the preference of travel destinations by analyzing the hashtag existing in the image set through the collection of SNS, classifying it as emotional language, and comparing it with the vocabulary showing general travel satisfaction. The vocabulary to be evaluated was divided into regional, tourism target vocabulary, common expression language, and emotional expression language. This study demonstrates tourists' preference for travel destinations according to whether collected hashtags match previously retrieved emotional hashtags. Through this, it is possible to analyze the changing travel trend of a specific travel destination and accordingly, it will be possible to establish a travel promotion policy. The limitation of the study is that it is necessary to be able to react to trends quickly because it is necessary establish promotion policies in accordance with real–time trends of recommended emotion language of many existing travel destinations. Through practical data collection and analysis in the future, a study will be done on the conditions and phenomenon of the combination of the hashtag of SNS data and whether they express a strengthening or weakening of promotion effect.

REFERENCES

[1] H. H. Kim, "A Tag-based Music Recommendation Using UniTag Ontology," The Korean Society Of Computer And Information, Journal of the Korea Society of Computer and Information, vol. 17(11), pp. 133-140, November 2012.

[2] S. H. Lee, S. J. Park, M. H. Lee, and D. H. Hwang, "Efficiency Analysis of Multi Tag-based Web Image Search," Korea Multimedia Society, pp. 260-263, 2009.

[3] H. J. Park, "A Qualitative Exploration of Folksonomy Users' Tagging Behaviors," Korean Society for Library and Information Science, Journal of The Korean Society for Library and Information Science , vol. 45(1), pp. 189-210, February 2011.

[4] I. Y. Chae, M. K. Moon, M. S. Kim, and K. Y. Yu, "A proposal on tourist destination preference analysis based on emotion using SNS," The Korean Society for Geospatial Information Science, pp. 131-132, 2016.

[5] Y. S. Kim, S. H. Jun, M. Jeon, and E. Hwan, "Tourist destination recommendation scheme based on real-time SNS trend analysis", Korea Information Science Society, pp. 291-293, 2014.

[6] S. J. Kim and S. D. Lee, "A Co-Occuring HashTag Analysis Technique In SNS EnvironMents SNS," The Korea Contents Society, pp. 223-224, 2014.

[7] Wikipidia, Folksonomy, [Online]. Available from: https://en.wikipedia.org/wiki/Folksonomy/, 2017.07.09.

[8] J. M. Lee, "A Conceptual Access to the Folksonomy and Its Application on the Web Information Services," The Korean Biblia Society For Library And Information Science, Journal of the Korean Biblia Society for Library and Information Science, vol. 18(2), pp. 141-159, December 2007.

[9] KoNLPy, KoNLPy: Python Korean Langueage NLP, [Online]. Available from: http://konlpy-ko.readthedocs.io/ko/v0.4.3/ 2017.07.09.

[10] J. K. An and H. W. Kim, 2015, "Building a Korean Sentiment L exicon Using Collective Intelligence," Journal of Intelligence and Information Systems, vol. 21(2), pp. 49-6, 2015.

[11] Open Hangul, Open Hangul API, [Online]. Available from: http://api.openhangul.com/, 2017.05.02

[12] J. H. Moon, J. Rew, Y. Kim, and E. Hwang, "Purpose-based Travel Destination Classification Scheme Using SNS Hashtags," Korea Information Science Society, pp. 121-123, 2015.