Journal of Travel, Tourism and Recreation
Volume 2, Issue 1, 2020, PP 13-28
ISSN 2642-908X

Understanding Destination Image in UGC: A Lexicon Approach

Weijun Li[1], Wencai Du[2*]

[1]Ph. D Candidate, International Tourism and Management, City University of Macao, Avenida Padre Tomás Pereira Taipa, Macao, China
[2]Professor, Faculty of International Tourism and Management, City University of Macao, Avenida Padre Tomás Pereira Taipa, Macao, China

***Corresponding Author:** Professor, Faculty of International Tourism and Management, City University of Macao, Avenida Padre Tomás Pereira Taipa, Macao, China.*

ABSTRACT

This study hastaken a pioneer position to apply semantic analysis to examine destination image through user-generated blog entries. Based on semantic rules, statistical analysis and co-occurrence analysis, this paper has analyzed blog entries generated by Chinese domestic visitorson the social media platform. A total of 3,160 valid reviews have been analyzed to compute semantic scores, explored the relationship between destination component and semantic scores and, investigated the features of positive and negative blog entries respectively. The findings have shownthat several dimensions of destination image are presented in the social media with each carrying different weight. It has also been proved that there is also a strong relationship between destination image and semantic scores. Prominent different clusters have been discovered in positive and negative comments, indicating different aspects of the destination are recognized by tourists. This study has embraced the challenges of converting the large volume of unstructured data into structure data for destination marketing and management. This work has also progressed in combining both semantic rules and statistical analysis.

Keywords: *sentiment analysis, user-generated-content, destination image; big data; lexicon approach.*

INTRODUCTION

Destination image has always been an intensive research area in tourism as having a knowledge of destination image can benefit both supply-side, such as destination position and promotion, and demand-side, related to tourist decision-making and satisfaction(Balogluand Mc Cleary, 1999; Beerliand Martín, 2004; Qu, Kim, andIm, 2011; Qu, Kim, andIm, 2011). Destination image is thought to be formed by visitors' reasoned and emotional evaluation of two closely interrelated components: cognitive assessment, related to visitors' own knowledge and beliefs about the attributes of the society, and affective evaluation associated with how individual feel about the object (Balogluand Mc Cleary, 1999). Compared with induced resources, referring to the information from marketing organizations, organic information sources, such as knowledge or experience from friends and relatives, exert a significant impact on the factors determining cognitive image of destination (Beerliand Martín, 2004). Cognitive attributes are found to have a significant effect on the affective attributes, and work as the antecedent of affective attributes (Stern and Krakover, 2010; Lin, Morais, Kerstetter, andHou, 2007). Baloglu and McCleary (1999) identified both cognitive and affective components working together can generate a compound destination image. The cognitive and affective components of destination image could result in a predisposition to revisit and recommend the destination (Prayag, and Ryan, 2011; Qu, Kim, andIm, 2011). It has also been found that the strength of affective attributes is stronger than cognitive attributes to generate the overall destination image, which calls for the strengthening the affective component in destination marketing strategy (Baloglu and McCleary, 1999; Li, Cai, Lehto, and Huang, 2010).

The prevalence of related destination information in the Internet is no longer dominated by destination management organizations (DMOs). Tourists have also been responsible for the information creation. User-generated-contents (UGC), serving as electronic word of mouth (e-WOM), is perceived to be trustworthy with no commercial interest (Murphy, Moscardo, and

Benckendorff, 2007). Because the goods in tourism and hospitality market are intangible and usually quite expensive, consumers are inclined to look for others' comments to minimize the risks related to their decision making (Sparks and Browning, 2011; Ladhari and Michaud, 2015). Hanlan and Kelly (2005) proved that WOM, which is autonomous and independent information sources, has a potent effect on destination image. It acts like a halo-image and a benchmark for other destinations. Jalilvand, Samiei, Dini, & Manzari (2012) proved that eWOM can affect destination image, attitudes towards destination and travel intention. Morgan, Pritchard, and Piggott's (2003) identified that visitors with dissatisfied experience communicate derogatory information and negative WOM has compelling influence on the destination image. Valence of the EWOM has dramatic influence on consumers' trust, brand awareness, and the purchase intentions (Mauri andMinazzi,2013; Sparks and Browning, 2011; Ladhari and Michaud, 2015; Vermeulen and Seegers, 2009). Positive reviews can improve tourists' perceptions of hotels (Philander and Zhong, 2016). Dennis, Merrilees, Jayawardhena, and Wright (2009) confirmed positive attitudes of the suppliers can positively affect the demanders' intention to purchase.

Methodologies used to measure destination image are dived into qualitative and quantitative ones. Quantitative studies, such as structural equation modelling (SEM), are employed by many scholars to identify the relationship among the attributes and the strength of each (Hosany, and Gilbert, 2009; Hosanyand Prayag, 2013; Zhang, Fu, Cai, and Lu, 2014). While it is useful to examine mechanism among the variables, this method suffers several shortcomings. The design of the questionnaires is based on the attributes of the destination, which could be unreliable (i.e., certain significant attributes may be absent) and unimportant to individuals (Beerli and Martín, 2004). What's more, it cannot capture the integral and psychological impressions related to a destination (Echtner and Richie, 2003). Unstructured approaches are considered more conducive to measure the holistic components of destination image to reveal a broad and complete view, because a destination provides a wide range of products and services (Echtner and Richie, 2003; Greene, Jennifer, Valerie, and Caracelli, 2003). More recent studies show an increasingly prominent trend of applying qualitative methods and taking advantage of the UGC to analyze destination image (Banyai, and Glover, 2011). Content analysis (Banyai, and Maria, 2010; Carson, 2008; Choi er al., 2007; Law et al., 2010; Leung et al., 2010; Pan, MacLaurin, and Crotts, 2007; Wenger, 2008; Wong and Qi, 2017), textual analysis (Pan et al., 2007) and semantic network analysis (Liu, Huang, Bao, and Chen, 2019; Mali, Yafang, and Zhia, 2013) are frequently used to mine and examine the perceived destination image. Considering the merits and demerits of applying qualitative or quantitative alone, the combination of both methods is proposed.

Developed from artificial intelligence and natural language processing, sentiment analysis can detect the valence (negative, positive or neutral) and assess the strength of the sentiment (Pang and Lee, 2008; Thelwall, Buckley, and Paltoglou, 2011). Scholars in tourism and hospitality domain increasingly take advantages of both machine learning approaches (Duan, Cao, Yu, & Levy, 2013; Gu, Yoo, Jiang, Lee, Piao, Yin, & Jeon, 2018; Windasari & Eridani, 2017; Ye, Zhang, & Law, 2009) and dictionary-based approaches (Hao, Xu, & Zhang, 2019; Liu, Huang, Bao, & Chen, 2019; Mukhtar, Khan, & Chiragh, 2018) as part or main research methods. Duanet al., (2013) designed a classifier and assigned the sentiment polarity for each sentence to measure the service quality of hotel service. Ye et al., (2009) compared the performance of three supervised machine learning algorithms, namely N-grams, Naïve Bayes and SVM. They proved that the accuracy rates of three algorithms can reach more than 80% of correct classification and the SVM model and character-based N-gram model outperformed the Naïve Bayes model. Xiang et al., (2017) applied Latent Dirichlet Allocation, an unsupervised machine learning model, to discover the main topics related to consumers' experience and evaluation of hotel product. As for studies which used dictionary-based approaches, they all firstly collected tourism vocabulary and calculated the semantic values based on the linguistics rules (Hao et al., 2019; Liu et al, 2019). Although it has not been tested in tourism domain and in Chinese or English, the lexicon -based approach has outperformed supervised machine learning approach for Urdu sentiment analysis not only in terms of accuracy, precision, recall and F-measure but also in terms of the time and efforts saved (Mukhtar et al., 2018).

However, sentiment analysis is subjected to some drawbacks. Sentiment analysis is domain sensitive (Pang and Lee, 2008). Words with specific meaning in one domain does not carry the same meaning in another area. For example, English word 'complex' means 'complicated' or 'a group of buildings' in different context. In spite of these inherent demerits, computer-aided sentiment analysis is appealing to scholars and researchers. Wang, Gu, & Wang (2013) examined prior studies, and stated that the technique yields a rather high accuracy rate with 70% to 80% in training-test data matching tasks. Considering the objective of sentiment analysis is to acquire the overall pattern from the large volume of UGC, rather than perfect classification of all data points, this approach is acceptable and feasible.

Based on the above discussion about the importance of destination image and UGC, as well as the application of sentiment analysis, this study tries to achieve the following objectives:

- To calculate the sentiment score of blogs and detect the sentiment polarity in the blog entries on Ctrip.com through the construction of specific dictionaries;
- To explore the nature and the underlying structure of the perceived destination image in blogs;
- To explore the relationship between the perceived destination image in blog and the according sentiment score of blogs;
- To investigate the positive and negative images of the perceived destination image in blog entries on Ctrip.com;
- To providing managerial implications and recommend ways the industry can strengthen the positive destination image and improve negative destination image.

METHODOLOGY

In order to fully capture the representation of destination image in blogs, this research applied dictionary-based semantic analysis to further explore the blogs generated by tourists on the platforms. Ctrip.com was selected since it is the most popular and largest online travel agency in China with monthly active domestic users exceeding 200 million (Tencent tech, 2018). Specific research procedures are illustrated in Figure 1. The whole data collection and pre-processing were conducted in Python. Statistical analysis was performed in SPSS.

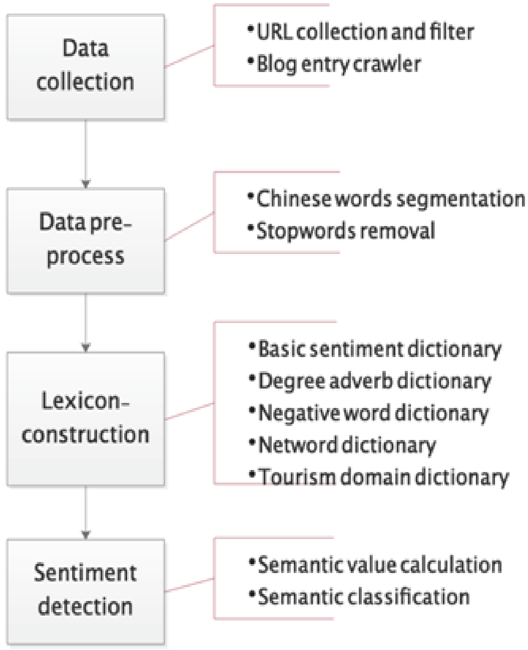

Figure1. *Steps of data collection and analysis*

Nanjing City

Nanjing, Known as a famous historical and cultural city, is located in the Yangtze River Delta of China. With 2500-year history of city building and 450-year history of capital building, Nanjing has rich cultural accumulation and unique cultural landscape. Announced by the city council, it was also among the first batch of historical and cultural cities in China. Recommended by 'Money' in 2019, a famous American magazine, Nanjing is listed in 20 world's best tourist destinations. It is particularly renowned for its ancient and modern history, literature, historical figures and local pastries. With about 134 million domestic and foreign arrivals, Nanjing has 1 World Cultural Heritage sites, 2 tourist attractions on China's world cultural heritage preliminary list and 49 national key cultural relics protection units. From the perspective of the authority, it is promoted as 'Capital of Universal Love', 'The capital of green' and 'Famous cultural city'. Therefore, considering the significance and representativeness of this city, this study uses the city of Nanjing as an example to analyze the UGC to understand online destination image.

Data Collection and Pre-Processing

A web crawler was designed to crawl blog entries in Ctrip.com. The contents and titles of 5032 blog entries were obtained from Ctrip. com. Excluding blogs which are about other destination, 3160 blogs were required in October, 2019. The blog entries were posted during the period between 2012 to 2019.

As for the pre-processing of the data, the first step was to divide the sentences into meaningful segments. There are several tools to cut Chinese sentences, such as Jieba, Snownlp, THULAC and ICTCLAS. This study applied PKUSEG to segment the sentences, which can provide segmentation model for the domain of tourism (Luo, Xu, Zhang, Ren, Sun, 2019) and which supports user-defined lexicon. The performance of PKUSEG has been proven to exceed that of JIEBA and THULAC, reaching more than 90% (Luoet al., 2019). When all the sentences were segmented under tourism model and with user-defined lexicon, words and characters which do not contribute to the specific meaning were all removed, such as '的' and '在'. In the end, 91,284 unrepeated tokens and 1,218,462 tokens in total were obtained for further analysis.

Lexicon Construction and Semantic Value Calculation

Considering the context of social media and tourism, as well as with reference to work did by Zhang, Wei, Wang, & Liao (2018), this study combined several widely recognized dictionaries and built a self-generated dictionary for this study. This studies first employed Chinese Emotional Vocabulary Ontology Database of Dalian University of Technology (CMVOD) as basic sentiment. This dictionary includes 27,467 words: positive word, negative word, neutral word, and the polarity intensity (PI)(Xv, Lin, & Pan, 2008). The PI of neutral words is assigned 0. The PI of positive words is set into five levels of 1, 3, 5, 7, 9 and the PI of negative words is categorized as -1, -3, -5, -7 and -9. As it has been previously argued that sentiment analysis is domain sensitive, the meaning of one word in a certain area does not carry the same sentiment in the tourism domain. For example, '火热' is a positive adjective in CMVOD, while it can express the negative feeling of being very hot about the weather in the domain of destination. Therefore, the author gleaned tourism vocabulary via the examination of 10% of the crawled blogs and collected 260 positive words and 202 negative words covering the area of cultural attractions, natural attractions, food, transportation, accommodations, shopping, and climate. The complementation of the words overlapped roughly 50% of the words in the CMVOD.

The semantic value of the sentences is also determined by the number of negators and degree words, the sequence of both and relational conjunction words. For example, the expression '不是 (not) /很 (very) /高兴 (happy)' is in the sequence of 'negator + degree word + sentiment word', while the expression '很 (very)/不 (not) /高兴 (happy)' is in the order of 'degree word +negator + sentiment word'. The former one has a diminished effect on the sentiment value while the latter one has an enhanced effect on the sentiment value. Therefore, Ois defined as the sentiment value of a single clause; W stands for the PI of sentiment word; D represents the magnitude of degree word. The formula for calculating the expression of the first diminished express is described as O = (-1)*D*PI*0.5 while for the enhanced expression is defined as O = (-1)*D*PI*2. Another example: 'The night market is too crowded, but the food there tastes very delicious'. The sentence is divided by the relational conjunction words into two parts, and the focus is in the latter parts. Therefore, the semantic value of the whole sentence is the sum of the two parts: 2.5 * (-3) + 2 * 2 * 5 = 12.5.

Table1.

Algorithm: sentiment analysis for a blog entry
Input: a blog entry
Output: sentiment value of a blog entry
1. A blog entry is divided into n sentences.
2. For ($i = 1; i ++; i <= n$)
3. If no relational conjunction in the sentence:
If no negator in the sentence:
$O = D * PI$
Else if the degree word is ahead of the negator:
$O = (-1) * D * PI * 2$
Else if the negator is ahead of the degree word:
$O = (-1) * D * PI * 0.5$
Else if *only n*negators in the sentence:
$O = (-1)^n * PI$
4. Else if
split the whole into main clause and subordinate clause
repeat step 3. for both clauses $O_{main}/O_{subordinate}$
$O = O_{main} + O_{subordinate}$

The sentiment value algorithm is presented in Table 1, and the PI of emotional word as well as the weight of degree words are presented in Table 2.If the semantic value (defined as V) of is greater than 0, then semantic tendency of the sentence would be considered as positive and vice versa. If V equals to 0, then semantic tendency of the sentence would be considered as neutral. Among the sentences belonging to the same blog entry, sentences with semantic value equaling to 0 will be excluded, as the study only focuses on the positive and negative aspects of destination image. Therefore, sentences with semantic value greater than 0 and less than 0 are grouped into according positive blog entries and negative blog entries.

Table2.

Category of words and the according magnitude of degree. Data source: collected and compiled by the author.				
category	Grade/Property	Magnitude of degree	Example of words	Number of words in the category
Degree word	a little, slightly	0.5	'a little', 'a bit', 'relative'	25
	comparatively	1.5	'all the more', 'more and more', 'also'	17
	very	2	'a lot', 'really', 'too'	42
	super	2.5	'completely', 'fully'	16
	extremely/most	3	'extremely', 'most'	52
Negator	not/never	-1		30
relational conjunction	concessive	0.5	'even though', 'although', 'despite'	10
	adversative	2	'however,', 'yet', 'but'	10

Statistical Analysis and Semantic Network Analysis

When dealing with big data, statistical analysis should consider the effect sizes and the explained variance rather than the conventional p value. Based on the proposition and the goal of the study, this work focused on exploring destination image-related words with the highest explanatory effect on the semantic value. The author identified that segments with high frequencies were very skewed in certain blog entries, with 78.32 % of words occurred in less than 45 % of all blog entries. Considering the assumptions based on the covariance among the word frequencies, words with low frequencies and blog entries without any high-frequency words were excluded. The optimization of the model was achieved through setting word frequency threshold to maximize the explanatory effect on the semantic value. Thus, the number blog entries and the number of words with high frequencies were reduced to 1532 and 81 respectively.

As these variables (high-frequency words) are correlated to each other, factor analysis was applied to answer the second research question: exploring the potential structure of destination image displayed blog entries. By using the factor scores as independent variables and the semantic value as the dependent variable, linear regression analysis was used to explore the relationship between the destination image presented online and semantic value.

In order to understand tourist semantic tendency, the co-occurrence of any two words in each blog entries were input and analyzed by a network analysis program called Gephi. This study conducted semantic network analysis of positive reviews and negatives reviews separately. In this study, the most mentioned words represent the nodes, the thickness of the edges between any two nodes stands for the relationship of the two nodes. In the output network, the nodes indicate the most mentioned aspects of destinations, the size of the node increases with the importance of the node, and the nodes which are in the same color means they belong to the same clusters; the thickness of the edges between any two nodes shows the degree of the closeness.

Results and Discussion

Descriptive Analysis

As discussed in the literature, the examination of destination image should consider both the cognitive and affective components. This study classified the destination of Nanjing based on the nine dimensions proposed by Beerliand Martín (2004). The categorization is illustrated in Table 3.

Table3.

Categorization of the high frequency words			
category	**sub-category**		**Number**
culture, history and arts	museum	museum, memorial hall, Taiping Kingdom History Museum, Nanjing Museum	4
	gastronomy	taste, snacks, delicious, delicious food, steamed dumpling, duck blood	6
	religion	Jiming Temple, Lingu temple, temples, Buddha peak Palace	4
	history	Sun Yat-sen Mausoleum, history, Ming Tomb, President Office, the walls, Republic of China, Rain flower pavilion, six dynasties, ancient capital, Gaochun county, tombs, Memorial Gateway, Imperial Palace	13
	culture and arts	Confucius Temple, Qinhuai scenic area, architecture, Gate of	9

		China, Qixia mountain, square, East Chinese Gate, culture, lion bridge	
natural resources		scenic spot, Xuanwu lake, South of Yangze river, fauna, scenery, environment, nature,	7
atmosphere of the place		city, beautiful, gorgeous, modern	4
natural environment		weather	1
general infrastructure	transportation	metro, vehicle, train station, train, bus, transportation, airport, express way, station, taxi, bridge, boat, car	13
tourist infrastructure	resorts and theme parks	park, hot spring reserve, Tang mountain,	3
	accommodation	room, hotel,	2
	dinning	restaurant, food street	2
tourist leisure and recreation		night street, sightseeing, shopping, night life, Hunan road, taking photo, LIBRAIRIE AVANT-GARDE,	7
others		like, recommend, feeling, friend, famous, experience	6

High frequency words in UGC fall into the night categories: culture, history and arts, atmosphere of the place, natural environment and natural resources, general infrastructure, tourist infrastructure, as well as tourist leisure and recreation, and there is a lack of the dimension of political and economic factors and social environment. Obviously, the dimensions of culture, history and arts are frequently mentioned by tourists in the blog entries. Attractions of modern history, such as Sun Yat-sen Mausoleum (7478), President office (6448), and Imperial Palace (4421), are of great importance because of the associations with Sun Yat-sen (forerunner of modern democratic revolution and former interim president of the Republic of China) and Yuan Shikai (First president of the Republic of China). Tourists also showed great interest in ancient history attractions. For example, the Ming Tomb (4296), Rain Flower pavilion (3149), and Imperial Palace (3637) were popular choices among tourists. These attractions were constructed during the Ming dynasties. A relationship can be found that blog entries which talked about these attractions are more likely to refer Nanjing as 'Jinlin', 'ancient capitals', a city with 'six dynasties. The Memorial Gateway (4026) and the Imperial Palace (3847) are sub-attractions in the scenic spots of the Ming Tomb. Tourists also had a preference for museums, such as Nanjing Museum, which displays the events throughout the history of Nanjing. Due to the high frequency words related to culture and history, the cognitive destination image strengthened, as close engagement with local culture can enhance the cognitive image of the destination (Kim, 2017). Mura (2015) proposed that tourists' experience of local cultural components would enrich their knowledge of and emotional bonds with the destination.

Tourists were also interested in visiting temples, with Jiming Temples having the high frequency (3354), followed by the Lingu Temple (2331) and Buddha Peak Palace(1896). This is due to the fact that since Buddhism was introduced into China in Nan Dynasty, Nanjing has been the center of Chinese Buddhist culture for a long time, making itself a well-known destination for the worship of Buddha. Tourist's visit to temples in the destination can be explained by their search for spiritual supporter they would like to transcend themselves in their beliefs (Digance, 2003).

Gastronomy represents a unique food culture of Nanjing. Tourist enjoyed local and traditional snacks (1283), including small steamed meatfilled buns (828) and Duck blood and bean-starchy vermicelli soup (836). No foreign food can be found in the UGC. Although no research has investigated the effect of local food on the cognitive and affective destination image, Choeand Kim, (2018) proved that tourists' attitudes to local food have positive impact on their intention to recommend and their intention to revisit the destination. In addition, tourists prefer to try these snacks in the vendor stands or small dining establishments in night street (409) or pedestrian street (582).

Eastern Suburb Scenic Area (4208) and Xuanwu Lake (3654) were scenic spots most mentioned by tourists. Eastern Suburb Scenic Area covers a huge area. Although Sun Yat-sen Mausoleum and the Ming Tomb is located in the Scenic Area, the whole area is also renowned for the spectacular view with the mountains, water, and forests being integrated. Similarly, Xuanwu Lake also have abundant history, serving as the largest Royal Garden Lake in China, the only remaining Royal Garden in the south of the

Yangtze River. Its significance and its magnificent beauty make it popular among tourists.

In the blog entries, tourists talked about the way they came to the destination and how they move around within the destination. In the UGC, as for the access to the destination, trains (1972) are more preferred than the aircraft (850)and express way (812). In terms of how tourists move around the city, metro is more often discussed and used than the buses and cabs.

When it comes to the places where tourists can shop, the most frequently mentioned place is Xinjiekou area (1203). With a bronze statue of Sun Yat Sen in the center, it is a famous commercial center in China with a history of one hundred years. Right now, this area is a huge complex with large shopping mall, many dining establishments and multifunctional facilities. Furthermore, the bookstore 'LIBRAIRIE AVANT -GARDE' (789) is also very popular among tourists. This is a theme bookstore, combining cultural salon, coffee, art gallery, film, music, creativity, life and fashion. Tourists especially enjoy taking photos there.

In terms of the accommodation discussed by tourists in UGC, tourists who post blogs online are more likely to live in hotels rather than hostels, as words with high frequency do not reveal much information about hostels. In addition, tourists especially pay more attention to the room in the hotel (1583). Discussion about the service of the hotel, and the friendliness of the staff can be rarely found.

Overall, the frequency distribution is highly concentrated in the dimension of culture, history and arts, with 81 words accounting for 78.32% of the total frequency of all words. Tourists' cognitive image reflected in blog entries center around the history, culture and arts, nature and gastronomy. Firstly, tourists were interested in the historical sites both ancient attractions, for example the Ming Tomb and the Walls, and modern locations, such as Sun Yat-sen Mausoleum and the President Office. Secondly, tourists explore local food by paying visit to the night street. In summary, most of the words with high frequency are related to the destination attributes and objective, while a small proportion of the words used by tourists represent their evaluation or feelings about the destination. It is also interesting to notice that the travel party ('friends' in this case). These words reflect the diverse dimensions of the destinations and their subjective feelings about the destination at the aggregate level.

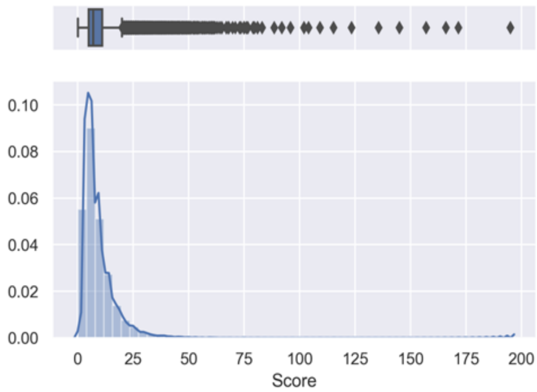

Figure2. *The distribution of semantic score of all the reviews*

Figure 2 shows the distribution of the semantic score of the whole blog entries. From the box plot, it is obvious that the scores of 75% of blog entries are below 12.546 and there is no blog entry whose score is negative. This may imply that tourist overall semantic tendency towards the destination is positive. It is also noticeable that the scores of some blog entries are extremely high. In order to understand this seemingly outliers, the author further examined the according blog entries and discovered that blogs whose scores were extremely high were also have many sentences. Although these scores are extremely high, they are meaningful and acceptable. Therefore, this study used the average semantic score of each blog entries for further analysis.

Statistical Analysis

Table4.

Factor loadings with words (shows only with loadings > 30)							
Words (N = 38)	Factors loadings						
Affective feelings (7)	F1	F2	F3	F4	F5	F6	F7
city	.809						
Environment	.584						
Scenery	.677						
Walking	.783						
International	.674						
Clean	.549						
prosperous	.532						

Ancient Historical attraction (11)							
Jiming temple		.775					
Lingutempe		.767					
Culture		.702					
Six dynasties		.66					
Wall		.684					
Xuan wu lake		.593					
Ming Xiaoling Mausoleum		.589					
Confucian Temple		.598					
History		.598					
Qinhuai river		.602					
Culture		.753					
Modern historical attraction (5)							
Republic of China			.736				
Museum			.682				
Taiping Kingdom History Museum			.681				
Zhonghua Gate			.801				
Presidential palace			.744				
Accommodation (3)							
Hotel				.561			
Room				.751			
service				-.767			
F & B (4)							
Duck blood soup with vermicelli					.692		
Steamed Bun Stuffed with Juicy Pork					.760		
Snacks					.590		
restaurant					-.426		
Transportation (4)							
Metro						.761	
Train station						.598	
Bus						.599	
Taxi						-.413	
Leisure and Recreation (4)							
Tangshan area							.623
Hot spring							.512
Holiday							.456
Shopping							-.426
Eigenvalue	12.53	16.00	2.89	2.76	2.52	2.46	2.06
Cumulative variance	8.32%	17.64%	24.155	30..43%	377.36%	42.47%	48.81%

Factor analysis was conducted in this study to explore the potential semantic structure of these words. Factor analysis is also very useful to reduce the number of words and classify the words into more meaningful groups. As the factor loadings obtained in this study were relatively low, this study consulted to the methods employed by Xiang et al.(2015). Therefore, this study set the cutting off loadings at (+/-).40 to acquire as many words as possible and threshold of eigen value at 2 to avoid the difficulty of interpreting the 'small' factors. Table 4 shows that there are seven factors composing of 35 words out of the 81 words and explaining 48.81% of all variance. Each factor was named based on destination image dimension these words belonged to. The first factor, consisting of 7 words, was named as the feelings as 'international', 'prosperous', 'clean' occurred with 'environment' and 'scenery'.

It is interesting to notice that the second factor and third factor are all associated with historical attractions. In order to differentiate the two factors, the second factor is named as 'ancient historical attraction' and the third factor is called 'modern historical attractions'. The forth factor is accommodation, including positive factors of 'hotel', and 'room' and a negative factor of 'service'. This suggests that the service tourists encountered may NOT take place in the hotel or about the room. In the fifth factor, which is named as F & B, it is interesting to note that the loadings of 'restaurant' is negative, suggesting that when tourists talked about 'snacks', 'Steamed Bun Stuffed with Juicy Pork' and 'Duck blood soup with vermicelli', the places where they consume such snacks are not likely in the 'restaurant'. The sixth factor represents the means of transportation tourists used in the destination, including positive factor loadings of ‘metro’, ‘train station’ as well as ‘bus’ and negative factor loadings of ‘taxi’. This may imply that tourist who have mentioned ‘metro’, ‘train station’ and ‘bus’ are unlikely to talk about ‘taxi’. The seventh factor ‘Leisure and Recreation’, includes four words, with shopping having negative factor loadings and suggesting tourists who mentioned ‘Tang Mountain Area’, ‘Hot Spring’ and ‘Holiday’ are not likely to do ‘shopping’.

On the whole, these factors identified the prominent components of destination images in UGC as words with high frequencies in UGC showed satisfied loadings in each factor. With a close examination of the frequency table and the factor loadings, it can be identified that most high-frequency words are also kept in the factor list, such as ‘Sun Yat-sen Mausoleum’, ‘Ming Tomb’, ‘snacks’, ‘architecture’, ‘city’ and ‘park’. While some words, like ‘night street market’ and ‘Nanjing Museum’, which are closely related to the formation of destination image, were expected in be found in the factor list, these words were not significant in the factor analysis.

Table5.

Results of linear regression analysis

Model	Unstandardized coefficients		Standardized coefficients		t	Sig.
	B	Std.error	Beta			
Constant	22.536	2.690			39.576	.000
Affective feelings	-.492	2.690	.250		14.578	.000
Ancient Historical attraction	.347	2.690	.003		.183	.000
Modern historical attraction	.803	2.690	-.138		-8.135	.000
Accommodation	.532	2.690	-.056		5.489	.000
F & B	.336	2.690	.068		-3.309	.000
Transportation	.437	2.690	.021		4.001	.213
Leisure and Recreation	.083	2.690	-.005		-.299	.000

Dependent variable: average semantic score of blog entries;
Adjusted R square: .682.

This study used seven dimensions of destination image as independent variables and the average semantic score of each blog as the dependent variable to explore the relationship between these factors and the sentiment tendency of the destination. Table 5 displays the ANOVA results. The whole seven factors except ‘Transportation’ were all significant at the p = .01 level. The factor, which has the largest factor loading is ‘affective feelings’ of .25. It implies that this factor is closely related with semantic value. Although most words with high frequency are in the dimensions of ancient history, the co-efficiency of this dimension is the least. Combining the factor loadings of each dimensions, the results could be more thought-provoking. In the factor of F & B, the semantic value of the blog was related to the discussion of the words such as ‘Duck blood soup with vermicelli’, ‘Steamed Bun Stuffed with Juicy Pork’, and ‘Snacks’. However, the negative sign of the word ‘restaurant’ indicates that blog entries with have high semantic values are not likely to mention ‘restaurant’, which may imply that tourists did not have these snacks in the formal restaurant. It is interesting to notify that dimensions of ‘modern history’, ‘accommodation’, ‘leisure and recreation’ are also negative. In terms of the factor of modern history, considering the factor loadings of this factor are all positive. If the semantic value is low of the blog entries, this factor is more likely to be discussed by tourists. As the factor of accommodation, the co-efficient of this factor was negative, while the factor loading of ‘service’ was negative. This suggests that in blog entries which semantic value was low, the word ‘service’ was NOT likely to be mentioned in the context of those words. Finally, in the factor of ‘leisure and recreation’, this factor has a negative relationship with the semantic value. Tourist who experienced hot spring or had a holiday in the Tang Mountain would have negative semantic score, while the positive sign of shopping indicated that a low semantic value is not likely to be linked to the discussion of the shopping.

Semantic Network Analysis

Last section indicated that although the overall tendency of all the blog entries was positive, there were negative sentiment. Thus, the author separated the negative sentiment and positive sentiment of each blog entry, and harnessed

Gephi to further explore the content and structure of tourist sentiment. The positive and the negative sentiment image of tourist to the destination is illustrated in Figure 3 and 4. The network analysis indicated that there were significant differences in the positive and the negative sentiment image.

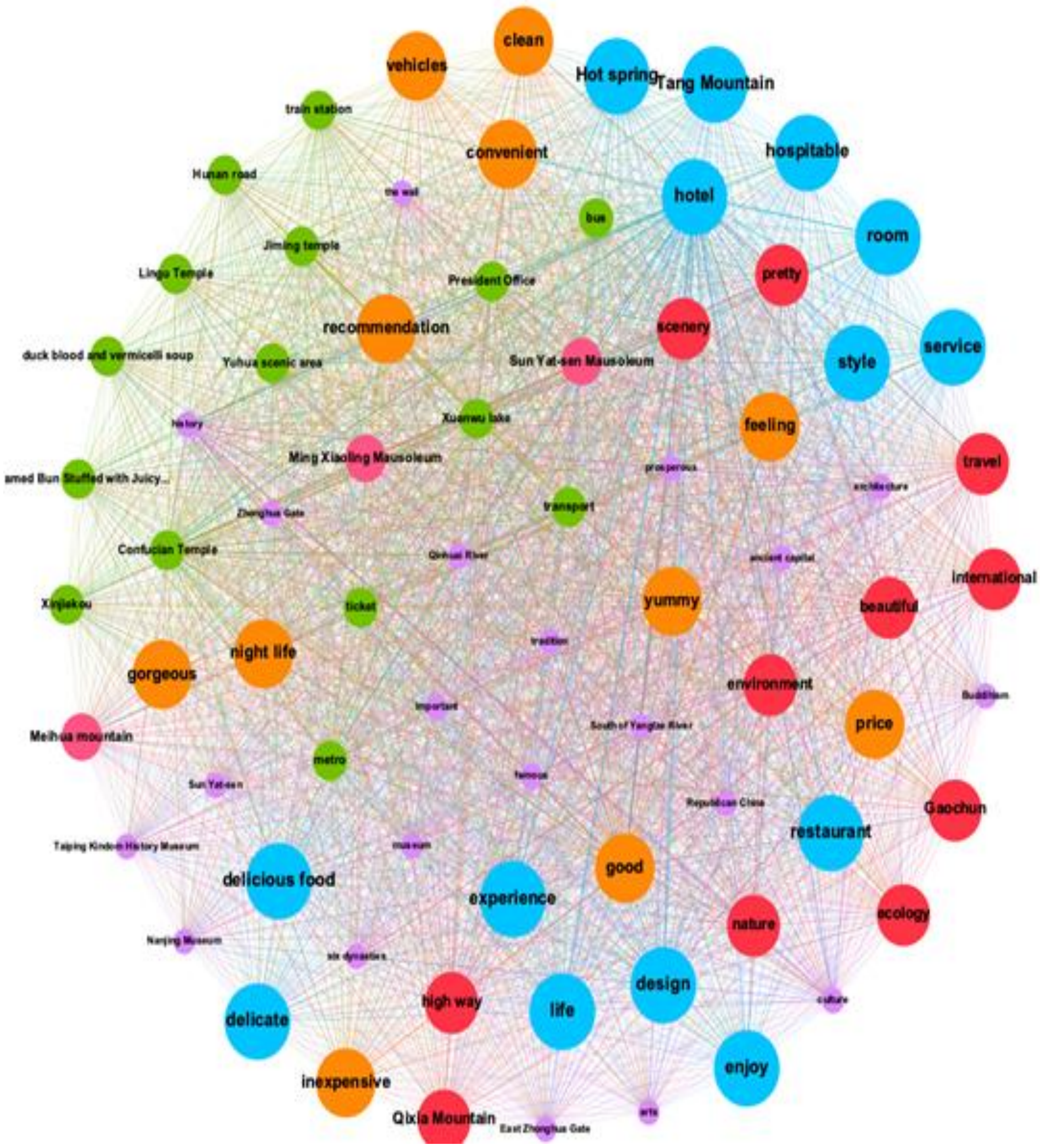

Figure3. *Destination image of positive blog entries*

The sentiment image of the of positive blog entries (Figure 3) consists of six clusters: (1) tourist travel experience of hot spring in Tang Mountain, (2) the beautiful scenery of nature and ecology in Qixia Mountain, (3) shopping and local food experience in Xinjiekou district and Hunan Road, (4) the convenience access of and inexpensive price level of local transportation, (5) important historical and cultural attractions, (6) three famous attractions in Eastern Suburb Scenic Area. The six clusters show tourists behavior preference when they were in the destination. Although the clusters reflect tourist visits to historical and cultural attraction, the sizes of the nodes were relatively small, indicating that tourists prefer to discuss the experience in the hot spring in Tang Mountain, their feelings about the nature, and satisfied with their use of local transportation system. Overall, although the destination of Nanjing is famous for its ancient and modern culture, tourist preferred activities were more related to recreation and they pay more attention to the transportation system in the destination.

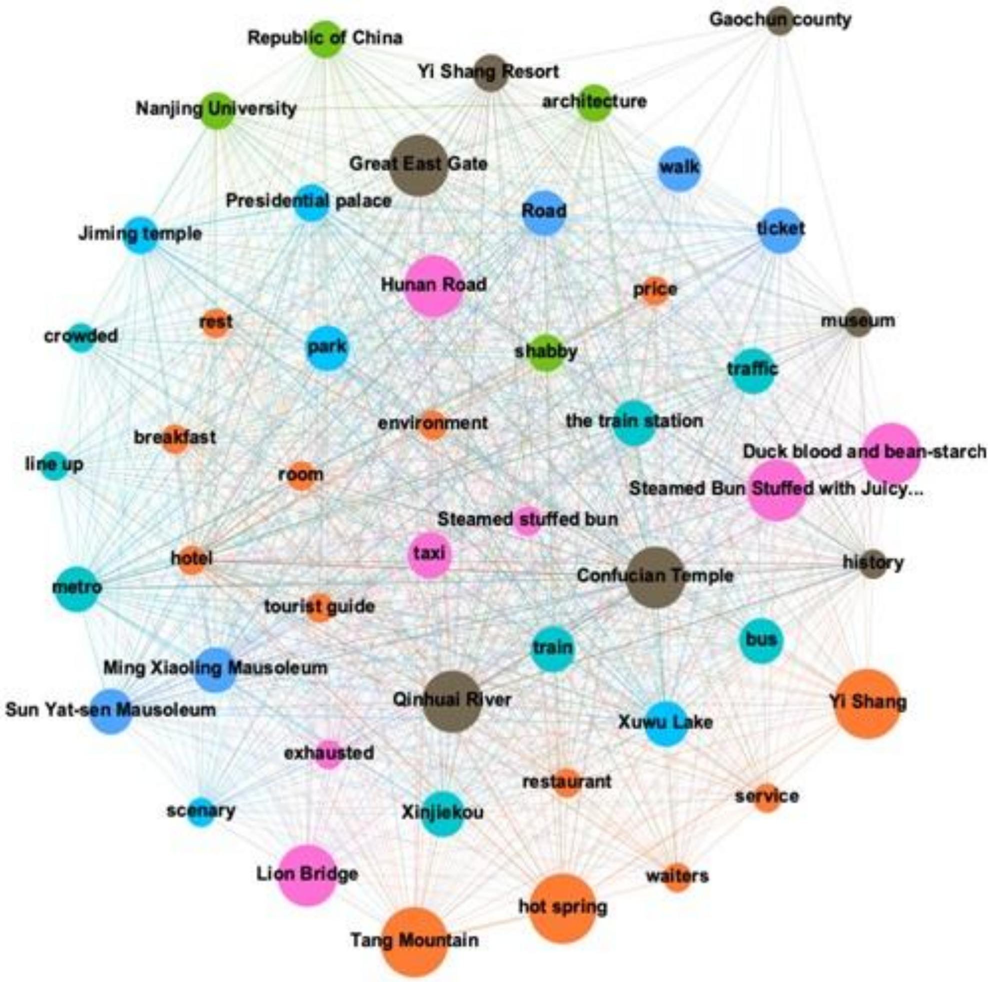

Figure4. *Destination image of negative blog entries*

As for negative reviews (Figure. 4), it is interesting to notice that while some key words appeared in positive reviews are also in negative reviews, tourists' sentiment tendency towards them vary. Snacks and night street where tourist consume these snacks are the most importance cluster indicated in the rose color. These experiences are more likely to be associated with exhausted feelings in the negative sentiment image. Although the hots spring reserve, Tang mountain, service and room both appear the positive and negative sentiment image, tourists were dissatisfied with the waiters, tourist guides, breakfast and the price level. Similarly, the clusters of transportation were in both images. Tourists in UGC complained the environment was crowded and they often lined up, which resulted in the negative sentiment value. The keywords, for example 'Museum' and 'Gaochun county' are new in the negative sentiment image. As no sentiment words are in this cluster, the author looked up the original blog entries and found that 'tourists felt very pity the museums was closed when they arrived there (No. 93)' and 'the imitated medieval castle was completely deserted and unsafe with poor facilities'. Interestingly, in the last cluster, the architecture of University of Nanjing or Republic of China is associated with shabby. Although shabby is classified as a negative expression in most contexts, 'shabby' linked with architecture in the period of Republic of China is likely to indicate the traditional and convention style of the buildings constructed in that period.

The findings in this study accorded with those inprevious findings. The overall sentiment tendency towards the destination is positive (Lu, 2017). Tourists are more positive towards of natural and cultural connotation of the destination. In the UGC, tourists often considered Nanjing important and famous due to its cultural and natural aspects. Tourists also explored the destination through local food and they usually consumed food in pedestrian street rather than in restaurants. Although tourist enjoyed the atmosphere and lifestyle in the hot spring in Tang Mountain, they were dissatisfied about the price level and other service encounters. Tourists These components and the strength of the sentiment are associated with the tourism promotion by NMACT, who greatly promoted Nanjing with recreational activities. Some of the findings only obtained in this study should also receive attention. The compliant about the crowdedness of the public transportation system requires DMOs to consider the use of public transportation also by tourists. Previous studies (W, Zh, Hu, C, and Xi, 2018) has investigated the destination image projected by NMACT and identified that NMACT emphasis the image of rural tourism. This study discovered that Gaochun county is only in the negative sentiment image. The result can be explained that DMO's promotion already make tourists pay attention to the rural activity in Nanjing, but DMO should make more efforts to improve tourists experience there to reverse their sentiment tendency.

Conclusions and Implications

This study explored the UGC through statistical analysis and sentiment analysis. Based on the semantic rules and the construction of tourism dictionary, this study firstly calculated the semantic score and detected the valence of the sentence. The methodology used in this study can tap the comprehensive positive and negative aspects of the destination.

As for the contribution of this study in practical terms, the findings illustrated tourists' preferred activities and attitudes towards the destinations. Although tourists frequently talked about modern and historical attractions in the blog entries, these two factors have opposite effect on the destination image. In addition, the strength of these two factors was weaker than tourists' overall feelings towards the destination, which inform the DMOs that they should place more emphasis on the stimulation of these feelings in the marketing strategies. What's more, especial the negative sentiment image reveals that crowdedness of the transportation and the exhausted feelings during the snacks consumption in night street should be improved. The price level in the hot spring resort in Tang Mountain should receive the attentions from DMOs as well.

In terms of the advancement in methodology, this study harnessed UGC and the linguistic rules to create the semantic value, which worked as an indicator to measure the satisfactory of the tourists' attitudes towards the destination. Although the calculation of semantic value in this study strictly followed the linguistic rules, it is unknown whether this method can outperform machine learning. What's more, this study responded to the suggestion provided by Banyai and Glover (2011). Previous studies which used UGC as the raw data would only count the word frequency as reflection. However, this study firstly used the distribution of the word frequency to filter the less qualified blog entries for the analysis of destination image. This study also combined factor analysis and regression analysis to identify the contributing factors of destination image.

However, this study still suffers a few limitations. Firstly, this study only used the blog entries in Ctrip. as the sample. Although the volume is large, further study can include as much variety as possible. Secondly, certain clusters in negative images only display the objects without any sentiment words, such as the grey cluster. These kinds of clusters arose further questions: whether the absences of semantic words is due to the granularity or whether further use of the negative blog entries can reveal comprehensive negative aspects of the cluster.

References

[1] Agapito, D., Oom do Valle, P., & da Costa Mendes, J. (2013). The cognitive-affective-conative model of destination image: A confirmatory analysis. *Journal of Travel & Tourism Marketing*, *30*(5), 471–481

[2] Baloglu, S., & McCleary, K. W. (1999). A model of destination image formation. *Annals of Tourism Research*, *26*(4), 868-897. doi:10.1016/s0160-7383(99)00030-4

[3] Banyai, & Maria. (2010). Dracula's Image in Tourism: Western Bloggers versus Tour Guides. *European Journal of Tourism Research*, *3*(1), 5-22.

[4] Banyai, M., & Glover, T. D. (2011). Evaluating Research Methods on Travel Blogs. *Journal of Travel Research*, *51*(3), 267-277. doi:10.1177/0047287511410323

[5] Beerli, A., & Martín, J. D. (2004). Factors influencing destination image. *Annals of Tourism Research*, *31*(3), 657-681. doi:10.1016/j.annals.2004.01.010

[6] Capriello, A., Mason, P. R., Davis, B., & Crotts, J. C. (2013). Farm tourism experiences in travel reviews: A cross-comparison of three alternative methods for data analysis. *Journal of Business Research*, *66*(6), 778-785. doi:10.1016/j.jbusres.2011.09.018

[7] Carson, D. (2008). The `blogosphere' as a market research tool for tourism destinations: A case study of Australia's Northern Territory. *Journal of Vacation Marketing*, *14*(2), 111-119. doi:10.1177/1356766707087518

[8] Gartner, W. C. (1994). Image formation process. *Journal of Travel & Tourism Marketing*, *2*, 191-216.

[9] Choe, J. Y., & Kim, S. (. (2018). Effects of tourists' local food consumption value on attitude, food destination image, and behavioral intention. *International Journal of Hospitality Management*, *71*, 1-10. doi:10.1016/j.ijhm.2017.11.007

[10] Choi, S., Lehto, X. Y., & Morrison, A. M. (2007). Destination image representation on the web: Content analysis of Macau travel related websites. *Tourism Management*, *28*(1), 118-129. doi:10.1016/j.tourman.2006.03.002

[11] Chon, K.-S. (1990). The role of destination image in tourism: A review and discussion. *The Tourist Review*, 45(2), 2–9.

[12] Cox, C., Burgess, S., Sellitto, C., & Buultjens, J. (2009). The role of user-generated content in tourists' travel planning behavior. *Journal of Hospitality Marketing & Management*, *18*(8), 743-764.

[13] Crompton, J. L. (1979). An assessment of Mexico as a vacation destination and the influence of geographical location upon that image. *Journal of Travel Research*, 17(4), 18–23

[14] Digance, J. (2003). Pilgrimage at contested sites. *Annals of Tourism Research*, *30*(1), 143-159. doi:10.1016/s0160-7383(02)00028-2

[15] Donthu, N., Kumar, S., &Pattnaik, D. (2020). Forty-five years of Journal of Business Research: A bibliometric analysis. *Journal of Business Research*, *109*, 1-14. doi:10.1016/j.jbusres.2019.10.039

[16] Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining Online User-Generated Content: Using Sentiment Analysis Technique to Study Hotel Service Quality. *2013 46th Hawaii International Conference on System Sciences*. doi:10.1109/hicss.2013.400

[17] Figueres-Esteban, M., Hughes, P., & Van Gulijk, C. (2016). Visual analytics for text-based railway incident reports. *Safety Science*, *89*, 72-76. doi:10.1016/j.ssci.2016.05.009

[18] Gallarza, M. G., Saura, I. G., & Garcia, H. C. (2002). Destination image: Towards a conceptual framework. *Annals of Tourism Research*, 29(1), 56–78.

[19] Garay, L. (2019). #Visitspain. Breaking down affective and cognitive attributes in the social media construction of the tourist destination image. *Tourism Management Perspectives*, *32*, 100560. doi:10.1016/j.tmp.2019.100560

[20] Gartner, W. C. (1993). Image formation process. *Journal of Travel & Tourism Marketing*, *2*(2/3), 191–215.

[21] Ghose, A., & Ipeirotis, P. (2009). The Econo Mining project at NYU: Studying the economic value of user-generated content on the internet. *Journal of Revenue and Pricing Management*, *8*(2-3), 241-246. doi:10.1057/rpm.2008.56

[22] Gu, Y. H., Yoo, S. J., Jiang, Z., Lee, Y. J., Piao, Z., Yin, H., & Jeon, S. (2018). Sentiment analysis and visualization of Chinese tourism blogs and reviews. *2018 International Conference on Electronics, Information, and Communication (ICEIC)*. doi:10.23919/elinfocom.2018.8330589.

[23] Greene, Jennifer, C., Valerie, J., & Caracelli. (2003). Making Paradigmatic Sense of Mixed Methods Practice. *In Handbook of Mixed Methods in Social and Behavioral Research*, edited by AbbsT ashakkori and Charles Teddlie. Thousand Oaks, CA: Sage, pp. 91-110

[24] Gunn, C. A. (1988). *Vacationscape: Designing Tourist Regions*. Van Nostrand Reinhold Company.

[25] Hanlan, J., & Kelly, S. (2005). Image formation, information sources and an iconic Australian tourist destination. *Journal of Vacation Marketing*, *11*(2), 163-177. doi:10.1177/1356766705052573

[26] Hao, X., Xu, S., & Zhang, X. (2019). Barrage participation and feedback in travel reality shows: The effects of media on destination image among Generation Y. *Journal of Destination Marketing & Management*, *12*, 27-36. doi:10.1016/j.jdmm.2019.02.004

[27] Hosany, S., & Gilbert, D. (2009). Measuring tourists' emotional experiences toward hedonic holiday destinations. *Journal of Travel Research*, *49*, 513–526.

[28] Hosany, S., &Prayag, G. (2013). Patterns of tourists' emotional responses, satisfaction, and intention to recommend. *Journal of Business Research*, *66*, 730–737.

[29] Jalilvand, M. R., Samiei, N., Dini, B., &Manzari, P. Y. (2012). Examining the structural relationships of electronic word of mouth, destination image, tourist attitude toward destination and travel intention: An integrated approach. *Journal of Destination Marketing & Management*, *1*(1-2), 134-143.doi:10.1016/j.jdmm.2012.10.001

[30] Kim, H., & Stepchenkova, S. (2015). Effect of tourist photographs on attitudes towards destination: Manifest and latent content. *Tourism Management*, *49*, 29-41. doi:10.1016/j.tourman.2015.02.004

[31] Kim, J. (2017). The Impact of Memorable Tourism Experiences on Loyalty Behaviors: The Mediating Effects of Destination Image and Satisfaction. *Journal of Travel Research*, *57*(7), 856-870. doi:10.1177/0047287517721369

[32] Kladou, S., & Mavragani, E. (2015). Assessing destination image: An online marketing approach and the case of Trip Advisor. *Journal of Destination Marketing & Management*, *4*(3), 187-193. doi:10.1016/j.jdmm.2015.04.003

[33] Költringer, C., & Dickinger, A. (2015). Analyzing destination branding and image from online

sources: A web content mining approach. *Journal of Business Research*, *68*(9), 1836-1843. doi: 10.1016/j.jbusres.2015.01.011

[34] Ladhari, R., Michaud, M., (2015). eWOM effects on hotel booking intentions, attitudes, trust, and website perceptions. *Int. J. Hospitality Manag*e. 46, 36–45.

[35] Law, R., & Cheung, S. (2010). The Perceived Destination Image of Hong Kong as Revealed in the Travel Blogs of Mainland Chinese Tourists. *International Journal of Hospitality & Tourism Administration*, *11*(4), 303-327. doi: 10.1080/15256480.2010.518521

[36] Leung, D., Law, R., & Lee, H. A. (2010). The perceived destination image of Hong Kong on Ctrip.com. *International Journal of Tourism Research*, n/a-n/a. doi:10.1002/jtr.803

[37] Leung, D., Law, R., van Hoof, H., Buhalis, D., 2013. Social media in tourism and hospitality: a literature review. *J. Travel Tourism Mark. 30 (1–2)*, 3–22.

[38] Li, M., Cai, L. A., Lehto, X. Y., & Huang, J. (2010). A missing link in understanding revisit intention-The role of motivation and image. *Journal of Travel & Tourism Marketing*,*27*(4), 335–348.

[39] Li, Y. R., Lin, Y. C., Tsai, P. H., & Wang, Y. Y. (2015). Traveller-Generated Contents for Destination Image Formation: Mainland China Travellers to Taiwan as a Case Study. *Journal of Travel & Tourism Marketing*, *32*(5), 518-533. doi:10.1080/10548408.2014.918924

[40] Lin, C., Morais, D. B., Kerstetter, D. L., &Hou, J. (2007). Examining the Role of Cognitive and Affective Image in Predicting Choice Across Natural, Developed, and Theme-Park Destinations. *Journal of Travel Research*, *46*(2), 183-194. doi:10.1177/0047287506304049

[41] Litvin, S. W., Goldsmith, R. E., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, *29*(3), 458-468. doi:10.1016/j.tourman.2007.05 .011

[42] Liu, Y., Huang, K., Bao, J., & Chen, K. (2019). Listen to the voices from home: An analysis of Chinese tourists' sentiments regarding Australian destinations. *Tourism Management*, *71*, 337-347. doi:10.1016/j.tourman.2018.10.004

[43] Lu M.Y. (2017). Research on image perception of Nanjing tourist destination based on network text analysis, Master's thesis, Nanjing normal university.

[44] Luo, R., Xu, J., Zhang, Y., Ren, X. , & Sun, X. (2019). Pkuseg: a toolkit for multi-domain Chinese word segmentation. *Journal of Information Hiding and Multimedia Signal Processing.*

[45] Mackay, K. J., &Fesenmaier, D. R. (2000). An exploration of cross-cultural destination image assessment. *Journal of Travel Research*, 38, 417–423

[46] Mali, X., Yafang, B., &Zhia, S. (2013). The Perceived Destination Image of Hangzhou City of China as Received in the Travel Blogs of Western Tourists. *Journal on Innovation and Sustainability. RISUS ISSN 2179-3565*, *4*(1), 101. doi:10.24212/2179-3565.2013v4i1p101-112

[47] Mauri, A.G.,& Minazzi, R., (2013). Web reviews influence on expectations and purchasing intentions of hotel potential customers. *Int. J. Hospitality Manage*. 34, 99–107.

[48] Michael, N., James, R., & Michael, I. (2018). Australia's cognitive, affective and conative destination image: an Emirati tourist perspective. *Journal of Islamic Marketing*, *9*(1), 36-59. doi:10.1108/jima-06-2016-0056

[49] Morgan, N. J., Pritchard, A., & Piggott, R. (2003). Destination branding and the role of the stakeholders: The case of New Zealand. *Journal of Vacation Marketing*, *9*(3), 285-299. doi:10.1177/135676670300900307

[50] Mukhtar, N., Khan, M. A., & Chiragh, N. (2018). Lexicon-based approach outperforms Supervised Machine Learning approach for Urdu Sentiment Analysis in multiple domains. *Telematics and Informatics*, *35*(8), 2173-2183. doi:10.1016/j.tele.2018.08.003

[51] Mura, P. (2015). Perceptions of authenticity in a Malaysian homestay – A narrative analysis. *Tourism Management*, *51*, 225-233. doi:10. 1016/j.tourman.2015.05.023

[52] Murphy, L. , Moscardo, G. , &Benckendorff, P. (2007). Using brand personality to differentiate regional tourism destinations. *Journal of Travel Research*, *46*(1), 5.

[53] Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(5), 1–135. Retrieved from http://dx.doi.org/10.1561/1500 000001

[54] Pan, B., MacLaurin, T., &Crotts, J. C. (2007). *Travel Blogs and the Implications for Destination Marketing*. *Journal of Travel Research*, *46*(1), 35-45. doi:10.1177/0047287507302378

[55] Pasarate, S., & Shedge, R. (2016). Comparative study of feature extraction techniques used in sentiment analysis. *2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*. doi:10.1109/iciccs .2016.7542328

[56] Peterson, R. A., & Wilson, W. R. (1992). Measuring customer satisfaction: Fact and artifact. *Journal of the Academy of Marketing Science*, *20*(1), 61-71. doi:10.1007/bf02723476

[57] Philander, K., & Zhong, Y. (2016). Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. *International Journal of Hospitality Management, 55*, 16-24. doi: 10.1016/j.ijhm.2016.02.001

[58] Prayag, G., & Ryan, C. (2011). The relationship between the 'push' and 'pull' factors of a tourist destination: the role of nationality – an analytical qualitative research approach. *Current Issues in Tourism*, *14*(2), 121-143. doi:10.1080/13683501003623802

[59] Qu, H., Kim, L. H., &Im, H. H. (2011). A model of destination branding: Integrating the concepts of the branding and destination image. *Tourism Management*, *32*(3), 465-476. doi:10.1016/j.tourman.2010.03.014

[60] Schuckert, M., Liu, X.,& Law, R., (2015). Hospitality and tourism online reviews: recent trends and future directions. *J. Travel Tourism Mark. 32* (5), 608–621.

[61] Singh, V., Waila, P., Piryani, R., & Uddin, A. (2013). Computational Exploration of Theme-based Blog Data Using Topic Modeling, NERC and Sentiment Classifier Combine. *AASRI Procedia*, *4*, 212-222. doi:10.1016/j.aasri.2013.10.033

[62] Sparks, B.A., Browning, V., (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Manage. 32* (6), 1310–1323.

[63] Stabler, M. J. (1988). The image of destination regions: Theoretical and empirical aspects. In B. Goodall & G. J. Ashworth (Eds.), *Marketing in the tourism industry: The promotion of destination regions (pp. 133–159). London: Routledge.*

[64] Stern, E., &Krakover, S. (2010). The Formation of a Composite Urban Image. *Geographical Analysis*, *25*(2), 130-146. doi:10.1111/j.1538-4632.1993.tb00285.x

[65] Sun, M., Ryan, C., & Pan, S. (2014). Using Chinese Travel Blogs to Examine Perceived Destination Image. *Journal of Travel Research*, *54*(4), 543-555. doi:10.1177/0047287514522882

[66] Tasci, A. D., & Gartner, W. C. (2007). Destination image and its functional relationships. *Journal of Travel Research, 45*(4), 413–425.

[67] Tencent tech. (2018, December 28). Sun Jie, CEO of Ctrip, delivered a new year's speech: Ctrip has more than 200 million monthly active users. Retrieved from https://tech.qq.com/a/20181228/004823.htm

[68] Thelwall, M., Buckley, K., andPaltoglou, G., (2011). Sentiment in Twitter events. J. Am.*Soc. Inf. Sci. Technol. 62* (2), 406–418.

[69] Tussyadiah, I. P., Park, S., &Fesenmaier, D. R. (2008). Assessing the effectiveness of consumer narratives for destination marketing. *Journal of Hospitality & Tourism Research*, *35*(1), 64-78.

[70] Vermeulen, I.E., & Seegers, D., (2009). Tried and tested: the impact of online hotelreviews on consumer consideration. *Tourism Manage. 30* (1), 123–127.

[71] Vishwakarma, P., & Mukherjee, S. (2019). Forty-three years journey of Tourism Recreation Research: a bibliometric analysis. *Tourism Recreation Research*, *44*(4), 403-418. Doi: 10.1080/02508281.2019.1608066

[72] Waila, P., Singh, V. K., & Singh, M. K. (2013). Blog text analysis using topic modeling, named entity recognition and sentiment classifier combine. *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. doi:10.1109/icacci.2013.6637342

[73] Wang, J., Gu, Q., & Wang, G. (2013). Potential Power and Problems in Sentiment Mining of Social Media. *International Journal of Strategic Decision Sciences*, *4*(2), 16-26. doi: 10.4018/jsds.2013040102

[74] Wenger, A. (2008). Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination. *Journal of Vacation Marketing*, *14*(2), 169-176. doi:10.1177/1356766707087525

[75] Windasari, I. P., & Eridani, D. (2017). Sentiment analysis on travel destination in Indonesia. *2017 4th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. doi:10.1109/icitacee.2017.8257717

[76] Wong, C. U., & Qi, S. (2017). Tracking the evolution of a destination's image by text-mining online reviews - the case of Macau. *Tourism Management Perspectives*, *23*, 19-29. doi: 10.1016/j.tmp.2017.03.009

[77] W, Zh, Hu , C, and Xi. (2018). The difference between the propaganda image of the official destination and the perceived image of tourists in the network context - a case study of Nanjing city. AREAL RESEARCH AND DEVELOPMENT, 37(3), 94-100

[78] Xiang, Z. &Gretzel, U. (2010) Role of Social Media in Online Travel Information Search. *Tourism Management*, *31*, 179-188

[79] Xiang, Z., Schwartz, Z., Gerdes, J. H., &Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management*, *44*, 120-130. doi:10.1016/j.ijhm.2014.10.013

[80] Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, *58*, 51-65. doi:10.1016/j.tourman.2016.10.001

[81] Xv, L. H., Lin, H. F., & Pan, Y. (2008). The construction of Chines emotional vocabulary ontology database. *Journal of The China Society for Scientific and Technical Information*, *27*(2), 180-185.

[82] Ye, Q., Zhang, Z., & Law, R. (2009). Sentiment classification of online reviews to travel destinations by supervised machine learning

approaches. *Expert Systems with Applications*, *36*(3), 6527-6535. doi:10.1016/j.eswa.2008.07 .035

[83] Zhang, H., Fu, X., Cai, L. A., & Lu, L. (2014). Destination image and tourist loyalty: Ameta-analysis. *Tourism Management*, *40*, 213–223.

[84] Zhang, S., Wei, Z., Wang, Y., & Liao, T. (2018). Sentiment analysis of Chinese micro-blog text based on extended sentiment dictionary. *Future Generation Computer Systems*, *81*, 395-403. doi:10.1016/j.future.2017.09.048

[85] Zhou, X., Xu, C., &Kimmons, B. (2015). Detecting tourism destinations using scalable geospatial analysis based on cloud computing platform. *Computers, Environment and Urban Systems*, *54*, 144-153. doi:10.1016/j.compenvurbsys. 2015. 07.006

Citation: *Weijun Li, Wencai Du, "Understanding Destination Image in UGC: A Lexicon Approach", Journal of Travel, Tourism and Recreation, 2(1), 2020, pp 13-28.*