Analysis of Preference of Tourist Destination using Twitter: Case Study on Theme Park in Seoul, South Korea

Inyoung Chae*, Jiyoung Kim*, Youngmin Lee*, Kiyun Yu*

* Seoul National University; ciy5164@snu.ac.kr; soodaq@snu.ac.kr ; dal-danka@snu.ac.kr; kiyun@snu.ac.kr

Extended Abstract

Recently, the number of travel-related texts posted by SNS users has increased and studies have been conducted to analyze them for the purpose of recommending meaningful tourist destinations for tourists.

However, existing studies do not consider cases where the contents include complaints about a destination because they consider only the frequency of destination mentions. Therefore, the results of the studies may not always recommend destinations that tourists actually prefer. Therefore, in this study, we quantified the preference for tourist destinations by considering positive and negative opinions about the destination mentioned in the SNS.

The subjects of this study were 118 theme parks that are tourist destinations in Seoul, the capital of South Korea. The subjects were selected from the re sults obtained after searching for the terms 'theme park in Seoul' using the POI search function in Naver map, a representative map of South Korea. A mong the 138 search results obtained, we removed several POIs that were n ot theme parks. Afterward, 118 theme park POIs were used for the study. S NS data used in this study comprised of Twitter posts collected by using the open API based on REST. We collected data from 22 July 2015 to 26 Februa ry 2016. After de-duplication, 273,515 posts were used in this study. Therefo re, in this study, we calculated the frequency of theme park mentions in the text of the aforementioned Twitter data and the preference for the theme pa rks in Seoul using sentimental analysis.

In order to quantify the preference for the theme parks, we first created a co rpus from collected Twitter data and removed stop words. Second, we extra cted Twitter texts containing theme park POIs among 273,515 Twitter data. A total of 1,674 Twitter texts were extracted, and the number of theme park

Published in "Proceedings of the 13th International Conference on Location-Based Services", edited by Georg Gartner and Haosheng Huang, LBS 2016, 14-16 November 2016, Vienna, Austria.

POIs mentioned more than once in the text was 38. Third, morphological a nalysis for natural language processing was conducted.

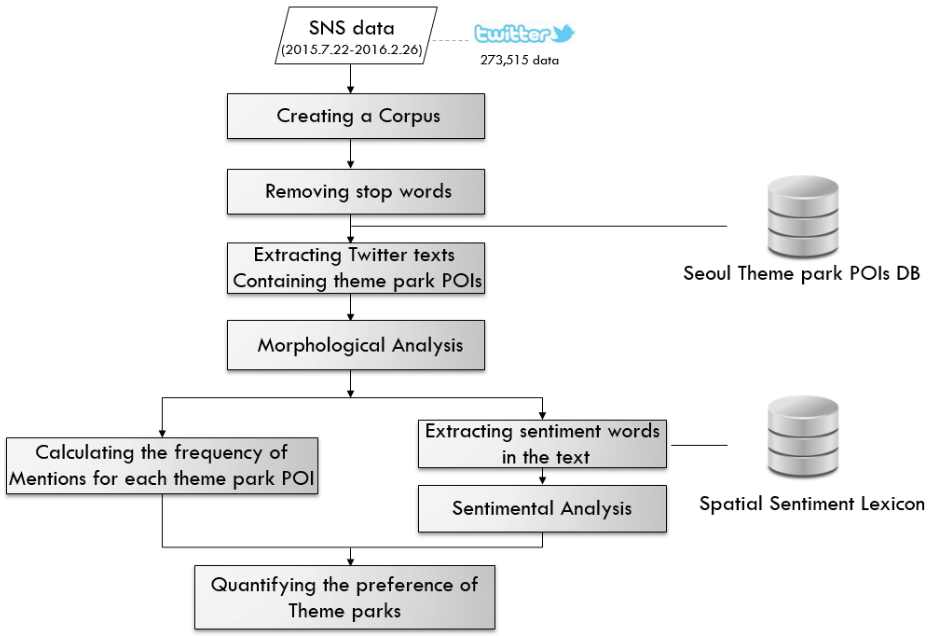

Figure 1. Research workflow

Next, we calculated the frequency of mentions for each theme park POI. We also analyzed the sentiment expressed in the extracted text to identify posit ive and negative opinions about the theme park POIs. Here, in order to anal yze the sentiment about the theme park, sentiment word describing places s uch as 'quiet or crowded' should be extracted. Thus, by utilizing existing spa tial sentiment lexicon such as in *Table 1*, we classified the sentiment present ed in the Twitter text as positive, neutral, or negative.

Spatial sentiment words	POS	Sentiment polarity	Probability
Recommendation	Noun	Positive (+1)	1.000
Best	Noun	Positive (+1)	1.000
Pleasant	Adjective	Positive (+1)	1.000
Noisy	Adjective	Negative (-1)	0.976

Table 1. Example of Spatial Sentiment Lexicon

At this point, in order to consider the context in the text, we conducted senti mental analysis by dividing the sentences expressing sentiments about the me parks into three types as shown in *Table 2*.

Type	Subject	Negatives	Predicate	Example
1	Theme park POIs	not	Sentiment word about places (adjective, verb, noun)	Lotteworld, (not), good
2	Spatial feature	not	Sentiment word about places (adjective, verb, noun)	Street, (not), pretty
3	Spatial feature	not	‘many’ or ‘little’	flower, (not), many

Table 2. Types of sentences for sentimental analysis

The first type consists of sentences expressing sentiments about theme park POIs. The second type consists of sentences expressing sentiments about the spatial feature indicating the properties of the theme park such as ‘street’ and ‘atmosphere’. The last type consists of sentences that can be either positive or negative when used with the terms ‘many’ or ’little’ regarding the same spatial feature. In other words, ‘uphill’ can have a negative implication when used with ‘many’, while having a positive implication with ‘little’. *Table 3* below shows an example of Twitter text of the second type.

Twitter text	Spatial feature	Sentiment word
I have a test tomorrow, but Lotteworld Christmas atmosphere is very good!	Atmosphere	good

Table 3. Example of Twitter text of the second type

Finally, we quantified the preference by using the frequency of mentions for each theme park POI and the results of sentimental analysis. The result is shown in *Table 4*.

Rank	Theme park POI	Preference	Rank	Theme park POI	Preference
1	Lotteworld	18000.25	6	Worldcup Park	209.55
2	Olympic Park	11506.96	7	Yeouido Hangang Park	148.00
3	Children’s Grand Park	545.00	8	Namsan Park	66.00
4	Dosan Park	291.95	9	Children’s Park	38.00
5	Dongdaemun History & Culture Park	248.00	10	Noeul Park	27.00

Table 4. Part of the preference results for theme parks in Seoul

'Lotteworld,' which had the highest preference, is a typical amusement park in Seoul. Therefore, it was mentioned many times in Twitter and had a high score in the results of sentimental analysis. In addition, 'Dosan Park,' which had the fourth highest preference, is notable because it had higher preference than 'Yeouido Hangang Park,' regarded as a representative park in Seoul. In other words, many positive opinions about 'Dosan Park' appeared in Twitter. These results may provide useful information for tourists who want to visit the park on the first trip to Seoul. Furthermore, we may help tourists to plan a travel route by analyzing the preference for various types of tourist destinations.

Acknowledgement

This research, "Geospatial Big Data Management, Analysis and Service Platform Technology Development," was supported by the MOLIT (The Ministry of Land, Infrastructure and Transport), Korea, under the national spatial information research program supervised by the KAIA (Korea Agency for Infrastructure Technology Advancement) (16NSIP-B081011-03).

References

Myung, J. S., Lee, D. J. & Lee, S. G. (2008) A Korean product review analysis system using a semi-automatically constructed semantic dictionary. Journal of KIISE: Software and Applications, 35(6): 392-403

An, J., Kim, H. W. (2015) Building a Korean Sentiment Lexicon Using Collective Intelligence. Journal of Intelligence and Information Systems, 21(2): 49-67