The Study of Using VGI to Analyze the Tourist Satisfaction About Taichung Jazz Festival

Shunjen Lai[1], Chengting Wu[2], Tienyin Chou[2], Meiling Yeh[2]

[1]Program in Civil and Hydraulic Engineering, Feng Chia University, Taichung, Taiwan

[2]GIS Research Center, Feng Chia University, Taichung City, Taiwan

Email address:
alexsjlaig@gmail.com (Shunjen Lai)

To cite this article:
ShunJen Lai, Chengting Wu, Tienyin Chou, Meiling Yeh. The Study of Using VGI to Analyze the Tourist Satisfaction About Taichung Jazz Festival. *Advances in Sciences and Humanities*. Vol. 4, No. 2, 2018, pp. 16-24. doi: 10.11648/j.ash.20180402.11

Abstract: Since 2003, the Taichung Jazz Festival has become one of the major annual events regularly held in Taichung City. The number of tourists and the tourism business opportunities brought by this festival has been increasing year by year, even reaching more than 1 million participating tourists/times ever since 2015. In terms of traditional assessment methods for great events, we used to obtain analytic information such as visitor satisfaction or the number of people through questionnaires. However, different levels of issues concerned by tourists cannot be easily understood through standardized questionnaires. Due to the popularization of online platforms and smart phones, people tend to voluntarily provide some information when they are participating in an activity. Such coordinated information is namely "Volunteered geographic information" (VGI), ex. "check-in" created by anyone. People can show their positive and negative messages by expressing their words about certain places (food, landscape, etc.), which can can make up for the shortcomings of traditional questionnaires. In this study, through the API provided by Facebook and by writing a web crawler program, we downloaded a total of 46,260 comments/messages written by people during the period of the Jazz Festival. Then, by means of Chinese word segmentation and through keywords, statistical analyses were conducted on two indicators shown by these tourists regarding the Jazz Festival: 1. Satisfaction about this event: To analyze people's positive and negative evaluations of the handling of this event, as well as their feelings; 2. Suggestions for event improvements: To analyze all aspects of concrete problems and suggestions for improvements proposed by people for this event. In this study, through collecting VGI data and constructing unstructured information analysis methods, explorations were made, concerning people's intuitive feeling about Jazz Festival from a mass perspective. In addition, comparisons and analyses against traditional questionnaires were conducted. Therefore, the findings of this study can serve as a reference for future leisure activity surveys combined with VGI data analyses.

Keywords: VGI, GIS, Big Data, Taichung Jazz Festival

1. Introduction

Tourism Industry is a very important part of urban economy. It belongs to Quaternary Industry. Because no pollution will be produced during the process of consumption and experience, it also belongs to Green Industry. Each city allocates its budgets according to its own financial conditions and its planning about the tourism industry. It can invest in public facilities, develop and construct sightseeing spots, or organize themes and itineraries according to the characteristics of the city. It uses various kinds of urban marketing methods to enhance the visibility of the city and promote tourist crowds, in order to create incomes and benefits for its urban economy. Each city has its own unique characteristics. The features of a city include humanities, history, culture, historic spots, ecology, cuisine, leisure spots, specialty products, agricultural products, shopping, natural landscapes, industrial activities, cultural activities, urban landscapes, etc., We can say that the scope is quite extensive. However, in order to flourish the tourism industry, we need well-organized planning, resource allocation, construction of various hardware and software,

convenient transportation networks, and readily-accessible public information, so as to create bright spots, emit charms and expand urban marketing, etc., all of which require the coordination and cooperation between various governmental departments. What's more, full achievements will not be reached without the assistance from non-governmental resources.

In this study, explorations of the Jazz Festival were made. The Jazz Festival originated from the conception created by Former Mayor Hu Zhi-Qiang and the Cultural Affairs Bureau of Taichung City Government, in hopes of making Taichung become the Edinburgh Festival in Asia. We found that the proportion of people loving jazz music is higher than that of those loving other music from people's feedbacks about the music festival "Encore!". Perhaps it is because jazz music can span eastern and western cultures best and it can meet the needs for people of all ages while corresponding to the city atmosphere. Therefore, jazz music events are included in Taichung Shining Art Festival and then be expanded as a separate music event. In the concert, famous jazz groups at home and abroad are invited to give performances. The venues for early events were scattered in Taichung Park, FengLe Sculpture Park, Jinguo Pathway and so on. As from 2005, this event is held on a small fixed stage within Jinguo Pathway and on the main stage in Civil Square. Such places are open space, and no tickets are required. The budgets of the music festival events are mainly from Taichung City Government and the participating enterprises. Five-starred hotels in Taichung City will also set up stalls on the scene to provide food and beverage. People can sit on the grass of Civil Square in front of the main stage. Later, picnic cultures even develop. As crowds of people gathered, there are even many people already occupying the space of their seats long before these evening shows begin.

Accordingly, the purpose of this study is to examine the internal and external results and benefits of the City Government's efforts to promote urban tourism activation. By combining with geographic information, smart mobile vehicles and the data mining models of big data, this study intended to understand people's intuitive feeling about the tourism activities in Taichung City (external benefits), which can be provided as the decision-making basis for the future development of tourism industry and the allocation of resource input. The implementation methods and procedures in this study are combined with the data of timeline. In terms of the variations of check-in information in the Jazz Festivals of different years held by Taichung City, the distribution patterns of check-in created by people were demonstrated by means of space, through the Overlay Analysis made by GIS software. Then, explorations were made regarding the positive and negative messages shown by the visitors, while constructive suggestions were extracted from such messages.

2. Method

This study explores the evaluation benefits of Public Participation GIS applied to large-scale tourism activities. Therefore, in terms of literature review, inductions would be made in the aspect of "the wisdom of crowds", for exploring the process of community data mining. In addition, by means of the visual platforms communicating coordinates, locations and photos through spatial information, and by combining with unstructured data processing methods, it was also meant for analyzing common or hidden messages to dig the value of "the wisdom of crowds", so as to make the judgment criteria more accurate.

2.1. Development of Big Data

Half a century ago, the rapid development of computer information science has begun to develop rapidly, resulting in swift accumulation of information volume. The digital data has been clearly seen a tremendous growth. The areas of greatest interest to experts are data mining and knowledge discovery [1]. In particular, information has been developed and accumulated in a rapid and explosive manner in this century, resulting in the formation of big data, massive data and Mega data. Big data are generally divided into structured data, semi-structured data, unstructured data; and the characteristics of big data are: 1. Volume of data; 2. Velocity of data; 3. Variety of data; 4. Veracity. Called "4V", these are the characteristics of big data.

Text mining has become a main trend, and text mining has also been combined with other research areas, for example, computational linguistics, Information Retrieval (IR) and data mining [2]. Big data are accumulated at unimaginable rates every day. For example, more than 10 million new photos may be uploaded by Facebook users per hour, while 3 billion likes or comments may be posted. There are also more than 400 million comments posted on Twitter each day. In 2013, Intel company announced a statistical record: in every minute, Google performs 2 million searches; Facebook has newly-added 350G of data volume, with 1.8 million people post likes; 72 hours of videos are uploaded to YouTube; 70 domain names are registered; 104,000 photos are shared on Snapchat; 278,000 "Tweets" are issued on Twitter. Huge amount of data can be generated on social networking sites (SNSs) within a minute, plus other data sources. Such a speed of accumulation is hard to be imagined. In such a case, such tremendous or complex data sets are hard to be processed via traditional data processing applications. However, with the newly-developed Data Mining software tools, a large number or different formats of data clusters can be accumulated or combined, to analyze or extract data relevance. Through such data relevance, researchers or data analysts can determine the status of road services, the pace of flu spread, the satisfaction or disadvantages of governmental services, the patterns of crime, and even the forecasts of the trends of public domains and commerce, in a speed faster than before.

2.2. The Wisdom of Crowds, and Social Network Analysis (SNA)

We can put forward the concept of "the wisdom of

crowds", whose central argument is that a diverse collection of individual autonomous decisions might make certain types of decision-making, forecasting and statistical sampling even better than those accomplished by experts. The nature of such information can be closer to reality even more. The congregate information from individual populace of social network information tallies with this feature. In addition, thanks to the rapid popularization of internet and location-based mobile devices, the development of these techniques has provided the ferment for the issues regarding the wisdom of crowds [3].

Facebook, Twitter, LinkedIn, Instagram and Micro-blog, those are the new type of social media, it develops very rapidly. It has the characteristics of high-speed communication, instant information updates, and strong interactivity. It has a huge impact on Internet users and non-user's social lives [4]. In early days, geographic information needs to be constructed manually. However, the development of today's geographic information has moved from the age of poor data to the age of massive data [5]. Most of the providers of this massive geographic information take photos, give check-ins and upload messages to the internet via smart devices. The persons, matters, timing, locations and objects revealed by them may contain coordinate points or can more or less reveal the geographic information of their whereabouts locations. Such spontaneous mass information can be referred to as "Volunteered Geographic Information" (VGI) [6]. In the commonly-used online network community platforms such as Facebook, Twitter, Flickr or Plurk and other related communities, there are features available for people to take photos, give check-ins and upload to the communities via smart devices and then share with others. At present, in terms of the study regarding the disasters in a city, the things shared by the populace all carry the coordinate points of spatial information or reveal the geographic information of their whereabouts through their semantic meaning. Driven by social communities, such information will be continuously transmitted to the platforms of social network sites (SNSs), and then be seen or used by other members of the platform. If geographic messages are presented through such information, members can know the locations released. Through the positions of such locations and space, the events occurring in the space can be analyzed further [7].

However, the information in online communities is usually enormous. It is mainly composed of unstructured text materials. Information providers are not volunteers. Information demanders obtain the information applied in a space from a huge amount of public information, through already available information. Since information providers are not crowdsourcing providers or volunteers, the biggest difference between such data characteristics and traditional data is that the data sources are diversified, with numerous types. Most of them are unstructured data, and are updated very fast, resulting in greatly-increasing data volumes. The greatest importance about big data is to explore available information, discover the models, find out the correlation between the data, and then evaluate the situations for predicting the future. In terms of the promotion of tourism and recreation, when people arrive at a tourist spot for experiencing relevant travels or leisure activities, such information will be posted on the internet and then become the information with their comments on these activities. Such information constantly increases every day and forms a huge database. However, the information provided by the populace belongs to non-structural information. The issues, such as how to effectively make non-structural text descriptions become structured and specialized and how to explore the relevance and applicability of their spatial distribution, have become important topics worth exploring.

2.3. Speech Recognition of Community Data

2.3.1. Semantic Mining

Under the tendency in which unstructured data are accumulated rapidly on the internet, text mining can help explore a variety of unforeseen, innovative, and important information or knowledge [8]. However, the analysis of these internet data needs the application the techniques of semantic analysis or opinion mining - text mining is mostly semi-structured or unstructured data; therefore, pre-processing should be conducted in advance, and its filtering mechanism is very important. As for the retrieval of spatial information, Rudolf in 2012 even proposed a method of analyzing the semantic meaning in Twitter for determining a user's geographical position in real time. In this experiment, the correctness of the results was determined by 93 Internet users on Amazon Mechanical Turk, which was provided as the information source of the wisdom of crowds [9].

2.3.2. Data Pre-Processing

Preprocessing Procedure: (1) In Syntactic Analysis, lexicon is required for conducting Tagging processing. By removing pleonasm, (2) The extracted Terms are filtered and screened, to decide which syntactic terms should be preserved; (3) The frequency of words is analyzed through statistical methods or algorithms such as TF-IDF [10].

2.3.3. Segmentation Processing of Words and Sentences

In segmentation processing of words and sentences, the Academia Sinica uses various articles collected from 1981 to 2007 to establish a balanced corpus. It also researched and developed a CKIP Chinese Word Segmentation System to undertake the processing and mark the parts of speech. All the words can be retrieved in the corpus and removed duplicate words, to obtain the lexicon of Maximum Matching Algorithm. Through Maximum Matching Algorithm, the words with the longest word length in the lexicon are compared, for segmentation of words and sentences. Then, combined with the Chain rule in N-gram Language Model, two discrimination methods have been applied, namely positive maximum matching algorithm and reverse maximum matching algorithm, so as to get a much more possibly accurate results of segmentation of words and sentences [11].

2.3.4. Framework of Mining

According to the semi-automated opinion mining process

as well as other relevant literature, the common processing procedures for emotion analysis or de-mining system are generally as follows: (1) Data collection: Collect data, through the web crawler or official API; (2) Data Preprocessing: Through segmentation of words and sentences, process the complete semantic meaning of the article and then exclude the words that are not significant but frequently shown in semantic analysis or emotion analysis, such as pronouns, prepositions, adverbs, stop words and repetitive data, so as to transform unstructured data into semi-structured or structured data that the system can understand, which can enhance the analytical accuracy of subsequent stages of opinion mining. (3) Opinion Mining: Through Text Mining techniques, calculate eigenvalues to retrieve the thematic classification of key words automatically, for analyzing opinion tendencies.

2.4. Research and Analysis Methods

The VGI mode of operation is applied in this project, to provide external information related to the Jazz Festival instantly from the perspective of the populace. In addition, through multiple non-structural information inductions including the data such as check-in keyword screening, semantic analysis, word classification and community spatial distribution, judge whether each message is from the location of Jazz Festival and whether its description relates. In addition, the interrogation and analyses of the results of the existing Jazz Festival (2012-2016) within the Taichung City Government were compiled. Spatial statistical analyses and conversions combined with internal and external information were conducted, to comprehensively examine the effectiveness of Jazz Festival promotion and the rationality of resource utilization.

2.4.1. Data Collection and Integration of the Wisdom of Crowds

In terms of data collection, it can be divided into structural data and non-structural data. The structural data is the official survey data conducted by Taichung City for the Jazz Festival, including the information about the number of people, satisfaction surveys, questionnaires, etc., for performing data inventories to understand the information related to the satisfaction about the existing events. Non-structural data is social media data, mainly based on Facebook social networking sites, to retrieve relevant data through web crawler programs or community API tools.

In this project, the mining will be conducted, by means of Facebook, the social networking site with the highest current usage rate. Then, keyword filtering was conducted according to the ID data sets collected, while easy coding was conducted, including field information such as label,

message content, coordinates, date and time, etc., Later, after the candidate information has been acquired, the information with real significance was screened and then filtered out, which will be converted into structured data for follow-up analyses.

2.4.2. Data Quality Assessment and Screening

Data Mining is meant for extracting the potentially-useful information and knowledge hidden in large, incomplete, heterogeneous, vague, stochastic, and practical application data (non-structural data). As for the collection mechanism in this project, the screening was conducted through two methods: Supervised and Unsupervised. Supervised Method is for conducting the first stage of screening against Facebook information database, through the self-selected keyword lexicon (training sample), to retrieve the keywords related to tourism and community, while the training samples were expanded through the related words to obtain the first-screened database. In the second stage, through the Chinese Word Segmentation System (Academia Sinica), inappropriate or less representative keywords were screened out by obtaining the occurrence frequency of each keyword after being segmented, so that the messages in the database can be more accurate and streamlined.

As for the database constructed in this project, the attributes of the database can be converted into suitable database formats, through different searches and document exchange applications and screening mechanisms. In the future, it can provide applications for the databases of different formats or the needs for other tasks.

2.4.3. Big Data Decision and Benefit Evaluation

Use the aforementioned internal and external data to undertake the integration of event output and spatial analysis, while connecting the timeline, in order to understand the amount of change in the check-in messages of different years. Through Cluster Analysis, Dasymetric Dot Distribution, Hotspot Analysis, Cost-effectiveness Analysis (CEA) and so on, the outcomes and special benefits from the implementation of the music festival have been evaluated, combined with big data decision analyses.

2.5. Research Framework

In Figure 1, the research of this project includes three main parts: (1) Data collection and analysis: the process of gathering and establishment of the Wisdom of Crowds (external materials), (2) Evaluation and screening of data quality: how to questionnaire information about Jazz Festival (internal materials), (3) decision and analysis situation: scenarios of big data decision analysis.
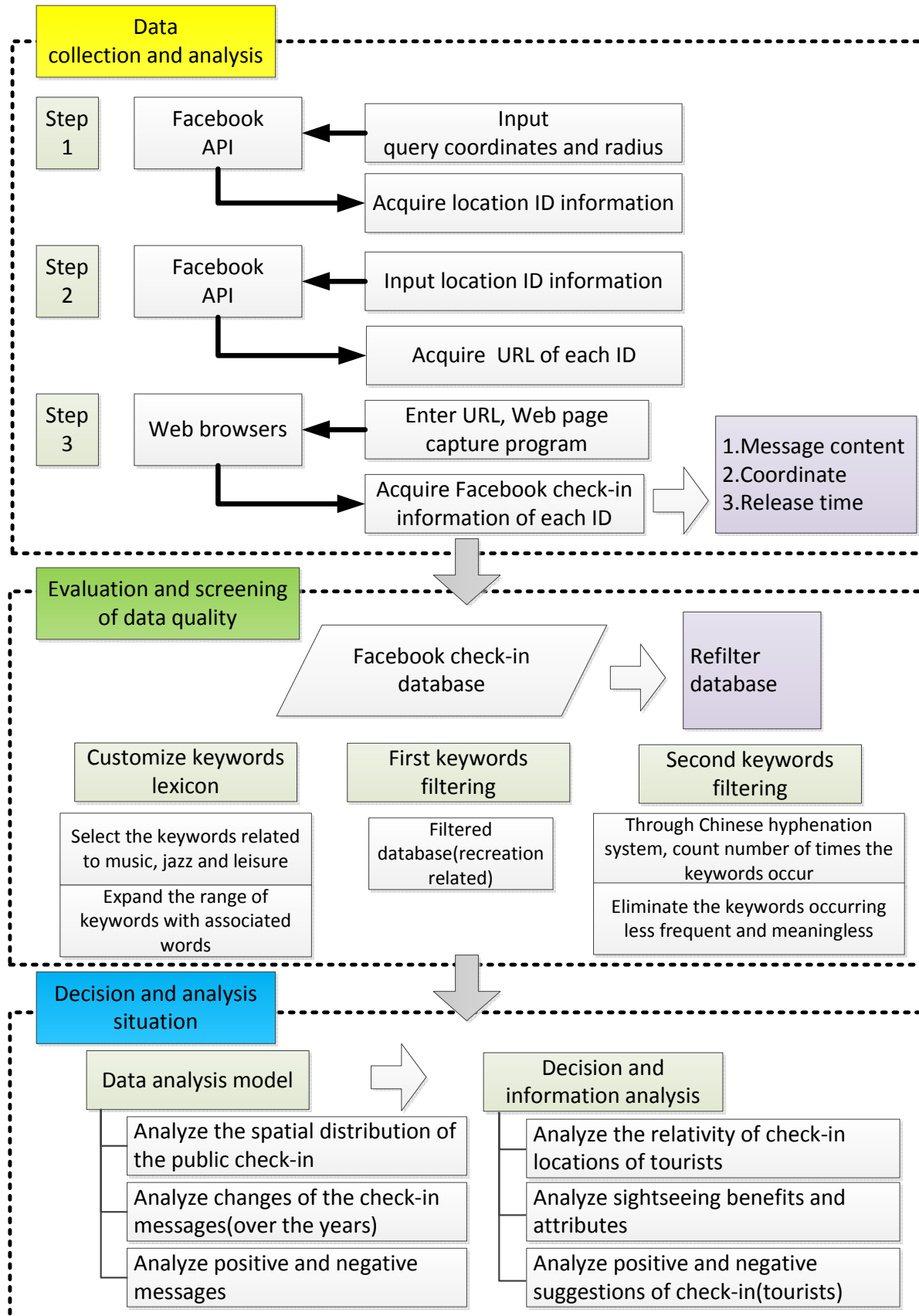
Figure 1. *Overall process of research framework.*

3. Results

3.1. Research Scope and Data Collection

The research scope of this study is mainly centered on the territory of Jazz Festival, including two blocks. As shown in the follow diagram, the blue block is the main area of performance site. The area is about 17,500 square meters, which is about the size of 2.5 standardized soccer stadiums. Most of the land in the Blue Block is covered with grass, for people to sit there and participate in these events easily. The

red blocks are the vicinity of these areas for commercial zones, trails or parking lots, to provide support for participants in other aspects, including food and beverage, shopping, taking a rest, toilets and other activities. The total area of such vicinity is about 107,400 square meters.

In this study, a total of 46,260 messages were collected from all the people who checked in the surrounding area via Facebook the social network platform. The check-in data within the research scope had been screened out further, and 16,646 messages were selected. Such messages were distributed in 115 positions (blue triangle positions as shown below).
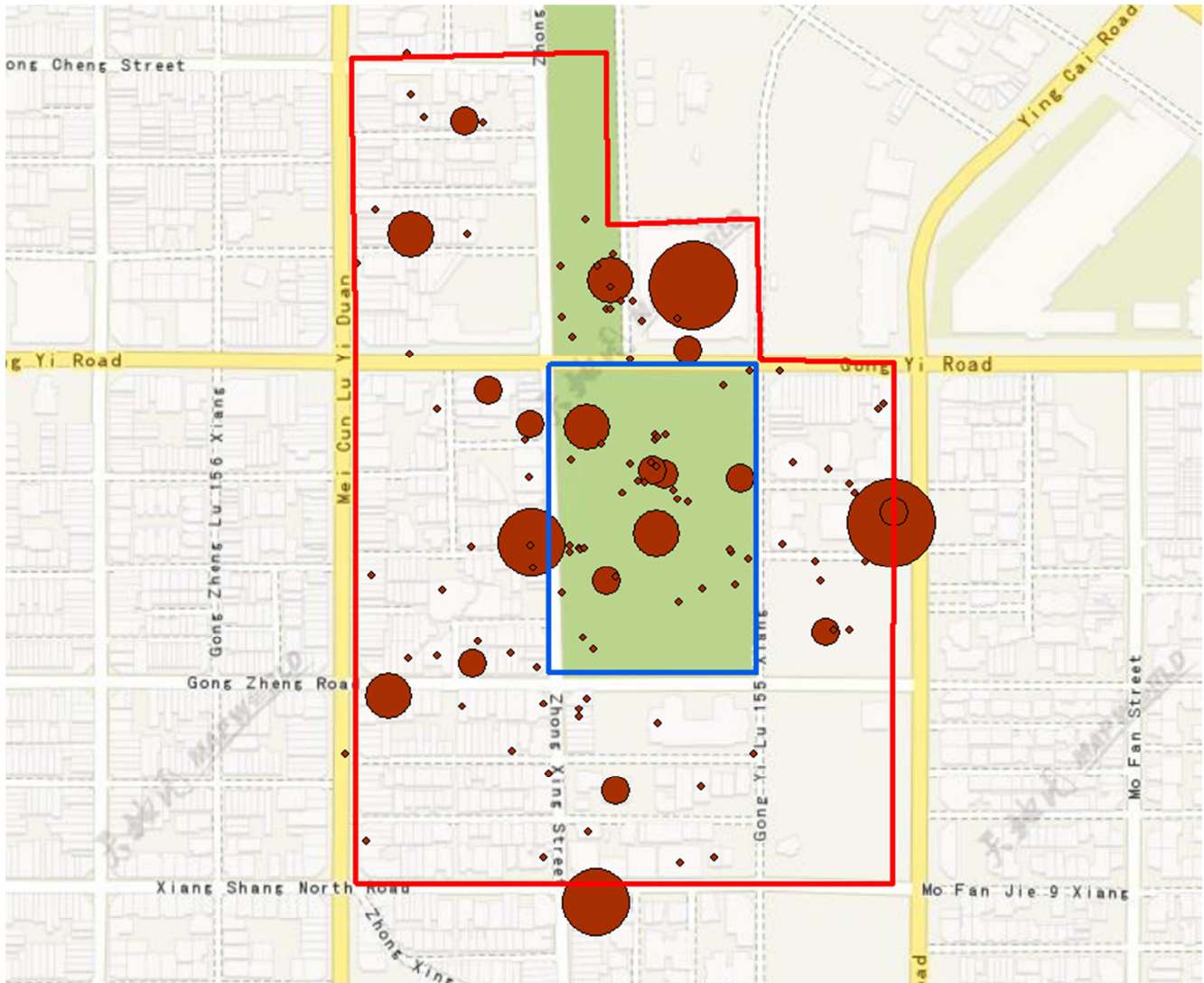

Figure 2. Spatial Distribution of Visitor Check-in.

3.2. The Keywords Rank of VGI Database

As for the 16,646 messages that had been screened out, this study selected the keyword messages and quantities related to the Jazz Festival, through the keyword ranking method for unstructured data, which can be further used as the training samples for database screening. After the semantic meaning of all the words has been screened out, the study selected the verbs and nouns that appeared more than 50 times in connection with the Jazz Festival. The compilation was made

as shown in the following figure. As for the analysis of the number of keyword occurrence in the comments created by the populace, it indicated that there were a higher proportion of the words associated with jazz music and the theme of the activity. As for the expressions of their satisfaction about this event, there was a higher frequency of the words such as "wonderful" and "appreciation". Negative words appeared for less than 50 times.
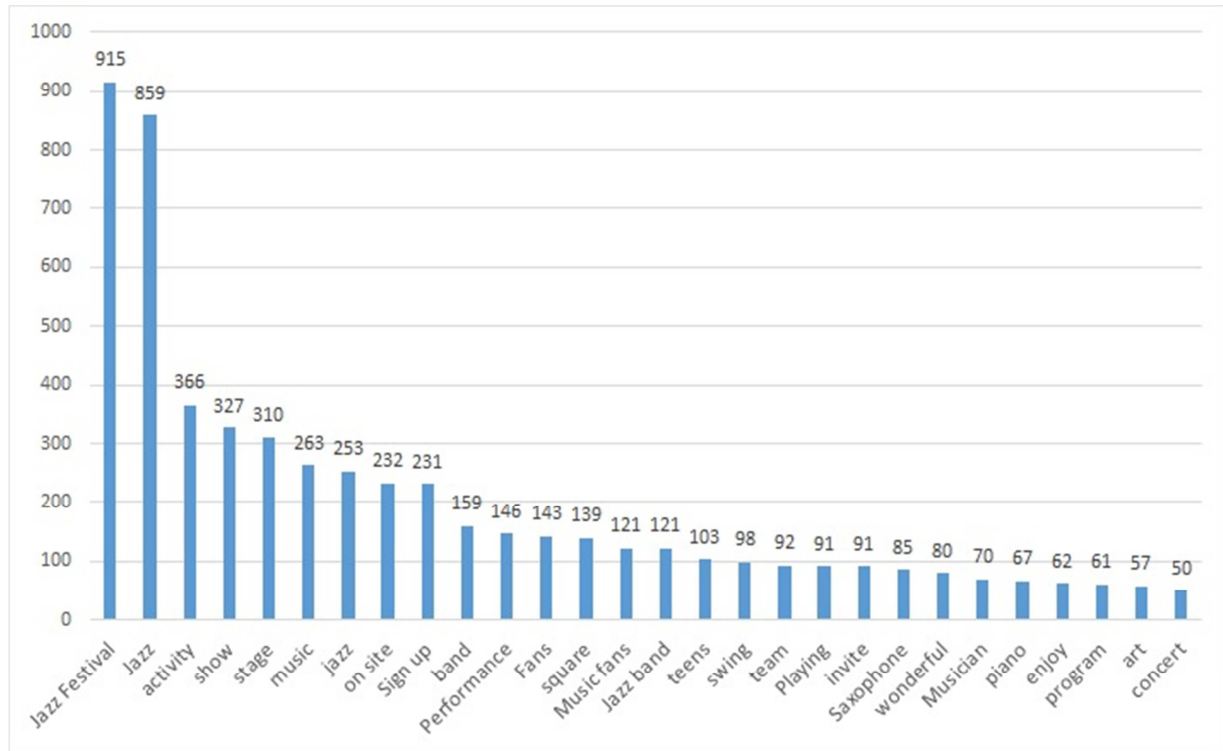
Figure 3. *The number of Keywords Rank of VGI Database.*

4. Discussion

Table 1 shows the official survey data of the Taichung City Government, including information on the number of days handled, the total number of participants, overall satisfaction, etc. However, the official data is too brief to reflect the opinions of the participants on the shortcomings of the activities. Through the results of this study, as shown in Figure 4, it can be shown that the massive VGI mass data indeed reflected the detailed and specific opinions of the masses on the issue of activity handling.

The study found that some people during the period of this event left their messages reflecting their problems about this event. Upon seeing such messages in real time, the organizer would immediately grasp the problems and make relevant improvements. Then, the messages related to the same problems would no longer appear. This means that it is an effective method to use VGI mass data as a tool for immediately grasping and improving the problems not noticed by on-site personnel. For example, in 2015, the audience felt uncomfortable about the orchestra's speech. The audience immediately left a message on the social platform, and then the host gave a positive response immediately.

In terms of the problems occurred in this event, the study found that VGI mass information could provide more detailed and specific expressions of these problems. However, as for the recognition and appreciation of this event, there were fewer detailed and specific messages. Most of the messages were emotional messages, which may provide less effective information when the organizer needs to analyze the merits of this event.

The study found that loud noises at the event site would affect the viewing quality for the surrounding audience. This problem persisted. From social network platforms, we could also find out that the host on the scene also made this known to everyone during this event, but the effect was limited. Further researches and explorations should be made, regarding how to understand the reasons for the loud noise and how to improve this issue.

The study found that the site's garbage did not leave with the audience after the event ended. From the messages on social network platforms, we could also find that the influence was poor although the host had made this teaching known to everyone. The site was always cleaned up by on-site volunteers. Further researches should be made, regarding how to understand the reasons why the audience did not take the trash away and how to improve the problems.

An interesting phenomenon was found in the study. In 2015, the city government renamed the Taichung Jazz Festival as the Huadu Arts Festival, which caused many repercussions in the VGI mass data. In the study, there were 14 effective responses, many of which were written in excited tones. They hoped to change the name to be the original Taichung Jazz Festival. In-depth discussions should be made, regarding whether any change will incur bad reactions and whether this represents that the brand of Taichung Jazz Festival has been deeply rooted in the hearts of people, since it has been held for many times as from 2003 to the then-current year.

Table 1. The official statistics and satisfaction Survey.

Year	The number of days for holding this event	Times of the event	The number of participants (people)	Satisfaction
100	9 days	54 times	850,000	72%
101	9 days	59 times	960,000	82%
102	9 days	53 times	1,050,000	81%
103	9 days	53 times	1,080,000	79%
104	9 days	49 times	1,320,000	81%
105	10 days	54 times	1,000,000	83%
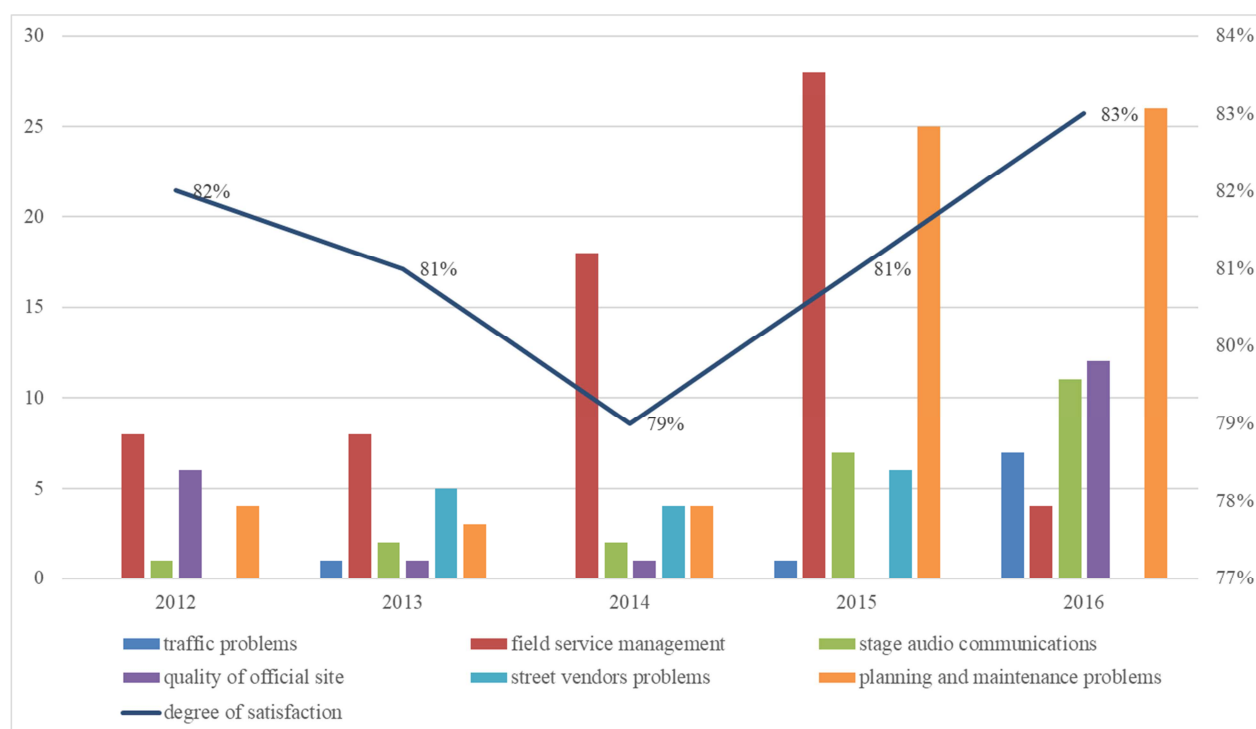
Figure 4. Comparisons on Community Check-in and Jazz Festival Questionnaire Survey.

5. Conclusion

Through the semantic extraction of VGI and community big data, this study obtained the semantic data of the specific issues raised beforehand, immediately and afterwards in the process of a large-scale activity. In terms of current relevant researches, as for general VGI questionnaire surveys of tourism and activities such as Taichung Jazz Festival, there are some question options for visitors to choose from. However, due to the questions asked by the organizer and due to visitors' time, it would be more difficult to obtain in-depth and concrete opinions through such questionnaires. In terms of the information generally obtained, there was no direct relationship with the improvement of this event, except satisfaction. Through this study, we did gain a large number of tourists' positive and negative intuitive expressions, as well as in-depth and specific criticisms and suggestions. For traditional questionnaires, such information can provide more complete suggestions to make up for the deficiencies of traditional questionnaires. From the information explored in this case study; we could have concrete understanding about the details that need our attention and caution when a large-scale activity is held. The findings of this study are indeed very helpful to the city government or the event organizer; they can have in-depth understanding about the nature of a problem, also able to deal with the problems proposed by tourists in real time. All of these are the applications worthy of in-depth discussions. In the short span of the decade from VGI applications to community big data, the developments of personal mobile devices, communication speeds, and social community platforms have entered a mature stage. The volume, scope and levels of data have all reached unprecedented extents. With regard to the applications of VGI community big data, it is believed that this study is merely in its small start-up stage.

References

[1]   Yang, L., Geng, X., &Liao, H. (2016). A web sentiment analysis method on fuzzy clustering for mobile social media users. Eurasip Journal on Wireless Communications and Networking, 2016 (1). https://doi.org/10.1186/s13638-016-0626-0.

[2]   Salloum, S. A., Al-Emran, M., &Shaalan, K. (2017). Mining Text in News Channels: A Case Study from Facebook. International Journal of Information Technology and Language Studies, 1 (1), 1–9.

[3]   Surowiecki, J., &Sunstein, C. R. (2004). The wisdom of crowds: why the many are smarter than the few and how collective wisdom shapes business, economies, societies and nations. Choice Reviews Online, 42 (3), 42-1645-42–1645.

[4]   Salloum, S. A., Al-Emran, M., Monem, A. A., &Shaalan, K. (2017). A Survey of Text Mining in Social Media: Facebook and Twitter Perspectives. Advances in Science, Technology and Engineering Systems Journal, 2 (1), 127–133. https://doi.org/10.25046/aj020115.

[5]   Harvey, M., &Han, J. (2001). Geographic data mining and knowledge discovery: An overview. Geographic Data Mining and Knowledge Discovery, 9–21.

[6]   Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69 (4), 211–221.

[7]   Lin, F. C., Hsueh, N. L., Wu, C. H., Chung, L. K., & Chou, T. Y. (2015). Using the intelligence of the masses to provide disaster information and information filtering mechanisms—Taking the flood disaster as an example, Engineering, 88 (04), pp. 53-67, Taiwan, ROC.

[8]   Parker, C. J., May, A. J., &Mitchell, V. (2012). Using VGI To Enhance User Judgements Of Quality And Authority. In D. Whyatt & B. Rowlingson (Eds.), Proceedings of GIS Research UK 20th Annual Conference (Vol. Lancaster, pp. 171–178). GIS Research UK.

[9]   Gonzalez, R., Figueroa, G., &Chen, Y.-S. (2012). Tweo Locator. Proceedings of the 5th International Workshop on Location-Based Social Networks - LBSN '12, (May), 24.

[10] Salton, G., &Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. Information Processing and Management, 24 (5), 513–523.

[11] Wu, Z. H. (2014), A Framework for Web Comments-based Opinion Mining System (Master's thesis). Feng Chia University, Taiwan, ROC.