

doi: [10.5821/conference-9788419184849.49](https://doi.org/10.5821/conference-9788419184849.49)

KEYWORD VERIFICATION OF VIDEO EXPRESSION NECESSARY FOR REGIONAL PROMOTIONAL VIDEOS TEXT MINING USING AWARD-WINNING FILMS FROM THE JAPAN WORLD'S TOURISM FILM FESTIVAL

Kaori MATSUBARA ^a

^a *Fukui University of Technology Graduate School, Japan, info@noahsark-design.com*

ABSTRACT

Due to the Corona situation, teleworking practices and relocations have increased, as people realize that they can do their jobs from any location. People have been relocating and buying land in remote and rural regions, away from the risks associated with the ongoing pandemic and inner-city living. The author of this paper has made several promotion films, documenting the initiatives of local people in the mountainous rural Ashimi district of Fukui Prefecture, and has been researching to verify and describe the results of these video expressions by comparing them to similar projects from different regions both in Japan and around the world. This paper researched the works of Japan World's Tourism Film Festival, which the author was also involved in managing, and used text mining to find keywords required to improve and objectively verify results.

Keywords: Text Mining, Tourism Film, Visual Expression Method.

1 INTRODUCTION

There are many marginal settlements in Japan today, including in the Ashimi area of Fukui Prefecture, which is the focus of this study. One such settlement is the Ashizumigama Charcoal Kiln, which is a local business attempting to help accommodate people relocating to the area, utilize local resources, and promote regional branding. The business engages in the sale of charcoal products, firewood, and shiitake mushrooms, placing a strong emphasis on protecting the local natural environment and forests in a sustainable way while at the same time increasing sales. While involved in the activities of this district, the author has been documenting the way of life that makes the most of local resources and has been conducting research into how best to archive their efforts through video images to find the best methods of expression for regional promotional videos.

In a previous study, the author aimed to research the different methods of video expression that would be useful for individuals or organizations involved in regional branding and promotion. As a result, it was interesting to note that even though the length of the two videos, the method of video expression, and the setting of the advertisement itself were the same, there were significant differences in viewer ratings and number of views, and the author wondered if these differences were related to other factors.

The fact that the title of one work included the name of the product, while the title of the other work was drawn from the words of the person interviewed in the video was considered by the author to be a factor that prompted a large difference in viewer ratings and interest in the video. However, it was difficult to quantify the qualitative elements for the disparity which led the author to search for methods of verifying which factors would give the viewers a favorable impression and increase viewer ratings.

Therefore, in this study, we examined 40 films nominated from 201 films submitted to the Japan World's Tourism Film Festival (JWTF) held in March 2022, along with 47 films nominated from 1542 films submitted to the international competition to determine what sort of keywords were used in the production of regional promotion films that were worthy of nomination. To verify what keywords were used to create the regional promotion videos that were worthy of nomination, a text mining analysis was used to examine them.

2 METHODS

The method of this study is an analysis using a user-local text mining tool (User Local, 2022).

First, regarding text mining, "Text Mining Techniques and Trends" (Ogi, 2015) begins, "What is text mining? Although it is commonly stated that there is no clear definition, it essentially means mining (information excavation) texts (sentences) and quantitatively analyzing texts with qualitative characteristics. This technology is called that because "data" in data mining corresponds to "text". At its core, it is realized by a combination of natural language processing, statistical analysis, and data mining technologies." Since around 2001, there have been several text mining studies in Japan.

Furthermore, IT Trends (IT Trend, 2022) indicates that the data collected has been studied since the 1960s, when computers were adopted as the backbone of business support systems. Data mining originated in the 1960s when computers were accepted as mission-critical systems to support corporate operations, and data analysis was used to create "hypotheses" for solving business problems (The History and Evolution of Text Mining, 2022). Text mining, on the other hand, has fundamentally altered conventional wisdom by eliciting "insights" before hypotheses. Machines are now able to grasp and translate natural language because of advancements in AI and deep learning technology. Siri and other AI assistants, search engines, and chatbots that use natural language processing are all examples.

According to the developer (Higuchi, 2017), KH Coder is a well-known text mining software in Japan. Higuchi states, "KH Coder is a free content analysis (quantitative text analysis) and text

mining program. It was designed to perform quantitative analysis on a variety of Japanese text-type data, including free-text descriptions in questionnaires, interview transcripts, newspaper articles, and so on. It can show lists of words in text files, concordances, and conduct other searches and statistical analyses.”

When studying text mining for the first time, the author tried utilizing this KH Coder, but after preparing and preprocessing the text data to be verified, the chart diagram of the “co-occurrence network,” which reveals the links between texts, which KH Coder is good at, was not retrieved properly. This most likely happened due to the author being trained in how KH Coder worked. However, some preparation was required to perform morphological analysis based on keywords and automatically re-analyze the text mining based on specific elements of voice and text, which the author believed required a significant amount of preprocessing effort.

As a result, the author chose to use the aforementioned user-local text mining tool for the research and analysis of this study because it was straightforward to grasp.



Figure 1. User Local AI Text Mining

2.1 Source data for analysis

The original data used by User Local was an Excel document containing the titles, synopses, production backgrounds, targets, and goals of 40 films nominated from 201 films submitted to the Japan section of the JWTF in March 2022, and 47 films nominated from 1542 films submitted to the international section.

Excel sheets comprising the titles, synopsis, production background, targets, goals, and other information for 47 regional promotion videos selected from 1542 videos submitted to the International Competition were used.

2.2 Pretreatment

The text data for the text mining study included 908 words or 18,728 characters for the Japanese sector and 7,897 words or 48,899 characters for the foreign area, as shown in Table The original Japanese data was exported to CSV data from Excel, with extraneous symbols (underbars, special parentheses, etc.) and coined words that could be separated eliminated, or concatenated as a single sentence preprocessed. Because the foreign component was written in English, specific coined words were checked, but no additional preparation was required, so the Excel data was converted to CSV.

Table 1. Text Mining Number of Source Data

	International	Japanese
Words	7,897	908
Characters	48,899	18,728

2.3 Analysis Method

Based on the previous study “Factor Analysis of Film Evaluation Using Text Mining and Network Analysis,” the author hypothesizes that the evaluation of films nominated for this year's Tourism Video Festival will be influenced by the factors that affect the evaluation of films nominated for regional promotional films (Ogawa, 2022). The author believes that by focusing on nouns and adjectives in the provided materials, they may extract the characteristics that influence the nominated films' evaluation. By doing so, the author will clarify the evaluation criteria for the works chosen by the film festival jury and consider their video expression approaches.

3 RESULTS

Table 2 shows the content visualized by text mining in the user's local area. This investigation will focus on word clouds, word frequencies, and co-occurring keywords.

Table 2. Types of analysis visualized

Type	Details
Word Cloud	Several words with high scores are selected and illustrated with a size corresponding to their value. The color of the words differs according to their part of speech, with blue representing nouns, red verbs, green adjectives, and gray moving words.
Word Frequency	The table shows the frequency of words that appear in a document. The size of the “score” shown for each word indicates how characteristic the word is in a given document. Usually, the more often a word appears in a document, the higher the score, but words such as “say” or “think” that appear frequently in all documents will have a lower score.
Co-occurring keywords	The figure shows a line connecting words that have similar patterns of occurrence in a sentence. The larger the number of occurrences of a word, and the stronger the degree of co-occurrence, the thicker the line is drawn.

3.1 Results of analysis by Word Cloud

The first step is to examine the word cloud's results. The word cloud is set up so that distinct results are displayed in order of score and frequency of occurrence.

The score (user local, 2022) is a number that represents the word's “importance.”, “Words that appear frequently in general papers are given less weight since they are not important” according to text mining, while “words that do not appear frequently in general documents, but solely in the document under study, are given more weight.” Words that appear infrequently in general papers but frequently in the document under study, on the other hand, are given additional weight. The TF-IDF approach is a statistical process that is used to extract these distinguishing words. The “score” is a value that takes into account not only the number of occurrences but also the degree of importance when using this method. A high-scoring word is one that characterizes the text.

The top parts of speech of the words exhibited in the order of score and frequency of incidence in the international section (Figure 2) were predominantly verbs in the order of score and nouns in the order of frequency of recurrence. Nouns are the most commonly used parts of speech in titles, synopses, and production objectives, among other things. The words “unique” and “different” stood out on the Word Cloud, and they were prominently shown in both score order

Table 3. The top 50 most frequently appearing terms and their parts of speech

International								Japanese							
No.	part	word	count	No.	part	word	count	No.	part	word	count	No.	part	word	count
1	noun	people	33	26	verb	take	16	1	noun	魅力	48	26	noun	a	12
2	noun	film	30	27	noun	island	15	2	noun	映像	45	27	noun	表現	12
3	noun	video	29	28	noun	one	15	3	noun	観光	31	28	verb	いただく	12
4	noun	campaign	28	29	verb	see	15	4	noun	自然	30	29	verb	感じる	12
5	noun	tourism	27	30	noun	sauna	14	5	noun	動画	29	30	verb	できる	12
6	noun	world	27	31	noun	cannes	14	6	verb	いく	26	31	Adjective	美しい	12
7	noun	culture	25	32	noun	adventure	14	7	noun	発信	24	32	noun	is	11
8	verb	make	25	33	noun	region	14	8	noun	制作	22	33	noun	方々	11
9	noun	city	24	34	verb	want	14	9	noun	島	20	34	noun	心	11
10	noun	place	23	35	Adjective	natural	14	10	noun	地域	20	35	noun	トアロード	10
11	noun	destination	21	36	Adjective	new	14	11	verb	伝える	18	36	noun	中央区	10
12	noun	experience	20	37	noun	australian	13	12	verb	もらう	17	37	noun	人々	10
13	noun	life	20	38	noun	australia	13	13	noun	旅	15	38	noun	歴史	10
14	noun	nature	19	39	noun	area	13	14	verb	訪れる	15	39	noun	多く	10
15	noun	event	19	40	Adjective	main	13	15	noun	富山	14	40	noun	世界	10
16	noun	time	19	41	Adjective	local	13	16	noun	暮らし	14	41	noun	東京	10
17	verb	find	19	42	Adjective	different	13	17	noun	街	14	42	noun	my	9
18	noun	holiday	18	43	noun	territory	12	18	verb	知る	14	43	noun	日常	9
19	verb	show	18	44	noun	u	12	19	noun	pr	13	44	noun	体験	9
20	noun	year	17	45	verb	look	12	20	noun	風景	13	45	noun	大阪	9
21	Adjective	unique	17	46	verb	go	12	21	noun	豊か	13	46	verb	歩く	9
22	noun	audience	16	47	noun	tourist	11	22	noun	i	13	47	verb	行う	9
23	noun	day	16	48	verb	promote	11	23	noun	文化	13	48	verb	持つ	9
24	verb	visit	16	49	verb	focus	11	24	noun	森	13	49	noun	奄美大島	8
25	verb	experience	16	50	verb	create	11	25	noun	まち	12	50	noun	people	8

Furthermore, there was only one adjective in the Japan portion of the top 50, the word “beautiful” therefore the author looked at the ratio of nouns to adjectives in the worldwide and Japan categories again, and the findings are displayed in Table 4. In the Japan segment, adjectives occurred only 2.1 percent of the time, indicating that nouns and verbs appear to be the primary components of Japanese tourism videos. The effect on the spectator is a topic for further investigation, but the findings of this study clearly illustrate that the approach to tourism video creation in Japan and other nations is vastly different.

Table 4. Differences in adjective occurrence rates

	International section	Japan section
<i>Noun</i>	1,075	1,469
<i>Verb</i>	343	267
<i>Adjective</i>	350	31
% Adjectives	32.6%	2.1%

3.3 Results of analysis by Co-occurring keywords

Figures 4 for the international section and 5 for the Japanese segment illustrate the results of the co-occurring keywords.

First, the international section's co-occurring keywords were organized into seven major groups. The most common group included the nouns “people”, “place”, “film” and “time” which led to the verbs “make”, “find”, “show” and “get” while adjectives such as “natural”, “local”, “real” and “beautiful” were displayed as related keywords. By connecting the keywords in this group, they can be regarded as keywords that provide suggestions for video titles and expressions.

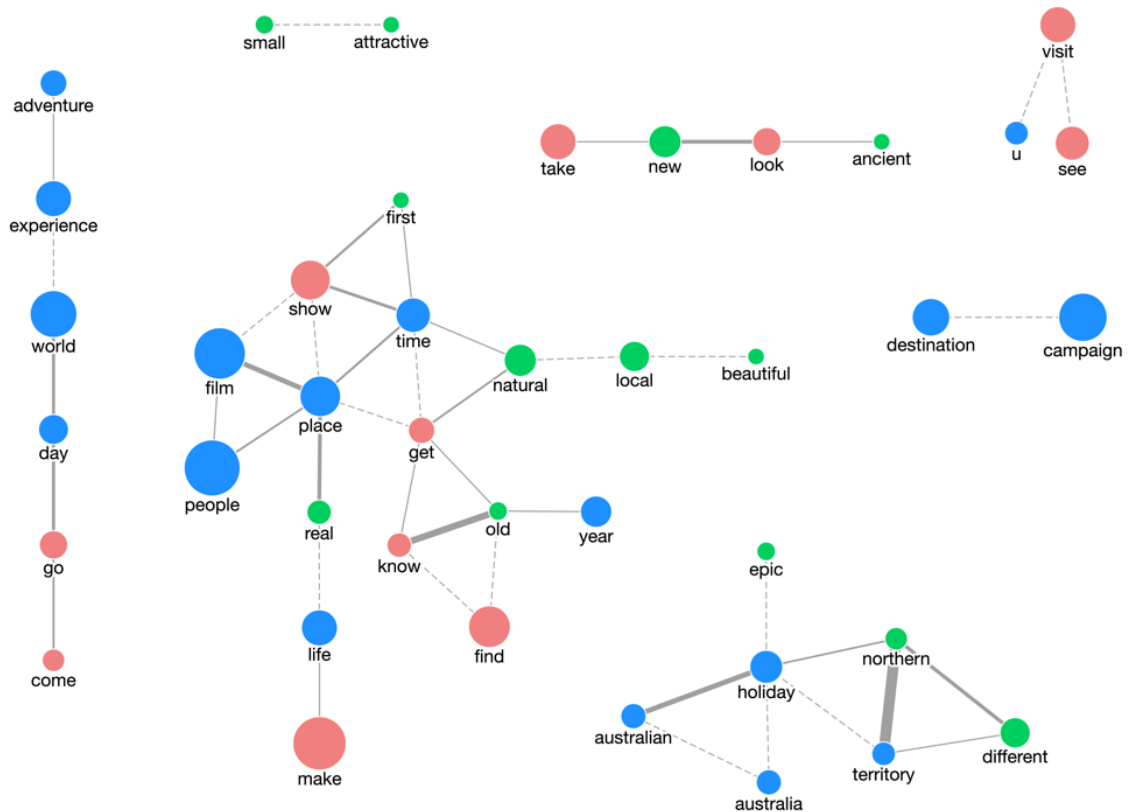


Figure 4. Co-occurring keywords International Results

For example, “If you come to this location, you will meet someone” (Matsubara, 2020), the title of a regional promotion film produced in a marginalized hamlet that is also the author's research base, earned the Grand Prize in the 2nd SDGs Creative Award. “People” are the “place” in “this place” and the “someone” in “you can meet someone”.

Similarly, the words used in the title of (Matsubara, 2022), another interview video produced in the previous study, were conceived without using specific place names or activities, with the goal of making viewers imagine what happens, what they see, and what they feel when they arrive at that location.

“Northern” and “territory” are strongly co-occurring words, as indicated by the thick gray line linking the co-occurring keywords. This is because the phrase “northern territory” appeared multiple times in the text detailing the summary and goals of one video production, and it was a keyword that should have been pre-processed as a single word. As a result, the author believes that this group included the related nouns “Australian” and “Australia” as well as the name of the region. Because the other groups were merely linearly related, it's safe to believe that the

keywords with the highest frequency of recurrence were linked to the phrases that accompanied them.

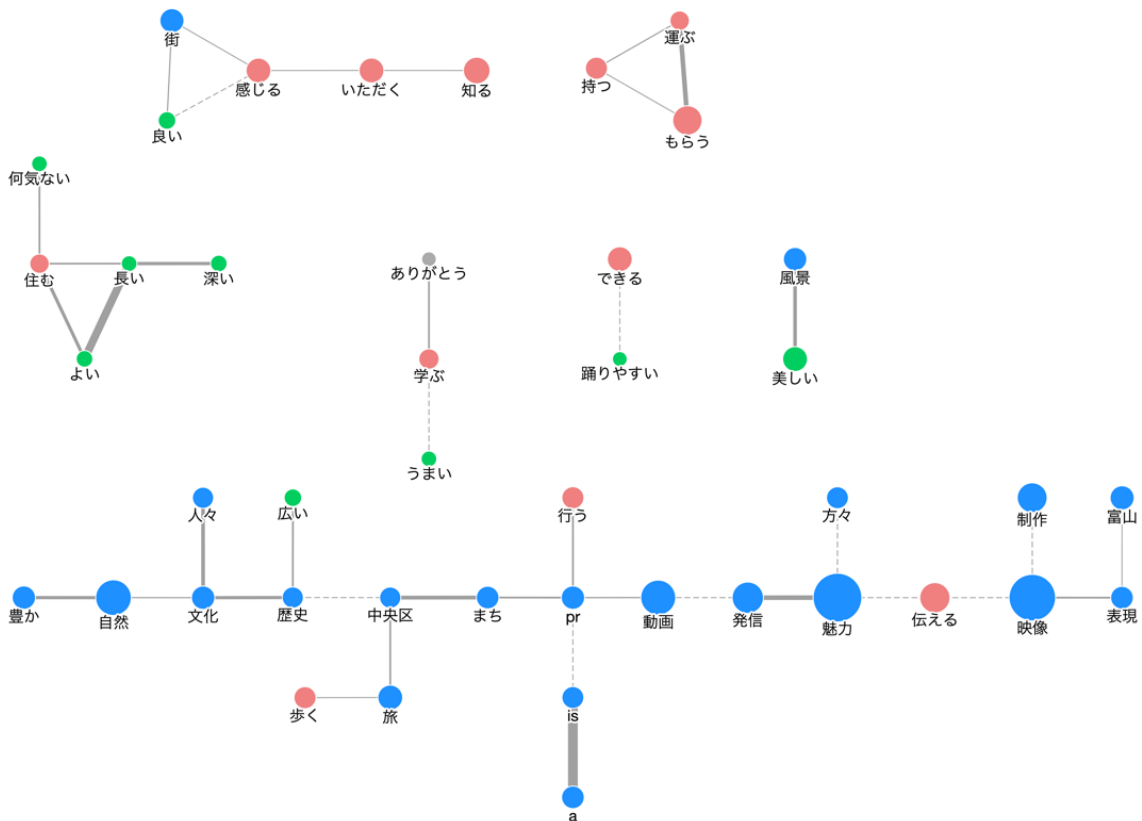


Figure 5. Co-occurring keywords Japan Results

In the Japan section, like in the international section, the results of the co-occurring terms were organized into seven primary groupings. However, in contrast to the international section, most of the keywords in the Japan section were connected by linear relationships, with nouns like “attraction” and “transmission” connected to the verb “convey,” nouns like “image” and “expression” connected to the verb “convey” and nouns like “city” + “PR”, “city” + “place name,” and “culture” + “history” and “culture” + “people” connected to the verb “culture.” The terms “culture” and “people” become increasingly linked to “culture” and “people.” Some sentences derive from keywords like “feel the good things about the town” and “let people know the excellent things about the town” which are associated with the adjective “good” in the group where the verbs “feel”, “receive” and “know” are coupled with the noun “town.” It’s possible that grouping co-occurring keywords into noun-centered, verb-centered, and adjective-centered groupings is a feature of the Japanese language, but further research on this is required in the future.

4 CONCLUSIONS

In this study, using text mining analysis as a validation approach, the author investigated whether keywords can be identified in regional promotion videos submitted to the Japan World’s Tourism Film Festival.

The adjectives “unique”, “different”, “natural”, “new” and “local” shone out on the word cloud in the international section, perhaps due to the heightened attention to nature and local cities in the aftermath of the Corona Disaster. Nouns like “people”, “location”, “film”, and “time” were the most common co-occurring keywords, leading to verbs like “create”, “discover”, “show” and “get” as well as adjectives like “natural” and “local”, “real” and “beautiful.”.

The bulk of keywords in the Japan section were nouns, but location names and terms like “attraction”, “video”, “PR” and “tourist” stood out, confirming that the film is a promotional film targeted at luring tourism. Furthermore, as compared to the international section, the word cloud was distinguished by the presence of adjectives in off-center places. Many co-occurring keywords were thought to be expressed in terms of notions such as “city” + “PR” + “conveyance” with nouns such as “attraction” and “transmission” leading to verbs such as “conveyance” and nouns such as “image” and “expression” leading to verbs such as “conveyance”.

Based on these findings, the author concluded that, in comparison to the rest of the world, current Japanese tourism videos appear to express the producers' desire to “transmit the attractiveness” and “how to make people feel and know the city's attractiveness” as well as specific place names as video expressions. The results of the international section, on the other hand, point to a more specific image of a trip to “make”, “find”, “show”, and “get” in a “unique and different”, “natural”, “new”, “local”, “real” and “beautiful” location, and it appears that the tourist images in the sightseeing video are meant to evoke this intention.

Professor Tsuyoshi Kigawa, the JWTF's representative, and Mr. Katsunosuke Honda of Honda-ya Honten, a JWTF judge, a producer of Japanese heritage, and a regional producer of Cool Japan for the Cabinet Office, spoke at a talk session (JWTF, 2022) and said, “Although the technology for photographing beautiful places and attractive parts of the region has improved, what strikes me about going there is what you can learn and notice.” He explained, “The technology for capturing gorgeous landscapes and attractive features of the region is developing, but what strikes me as interesting is what you can learn and see when you visit, which makes me want to visit. It makes you want to inquire further. It inspires you to go exploring. Rather than leaving the answers on the visuals, it leaves a lasting impression. The visuals left me feeling as if I could go there and find out for myself.” His assessment of the Japanese section was as follows. “Show-era tourism was about seeing,” Mr. Honda continued. “Tourism during the Heisei era was all about the experience. Tourism during the Reiwa era is all about discovery.” He underlined the importance of creating images that reflect changes in tourism. “I believe that by understanding and perfecting traditional culture and traditional performing arts, and sprinkling them into the visuals, we can create images that only Japan can create and that will be enjoyed by people all over the world,” he continued. He also gave us some tips on image expression that we should consider in the future when it comes to Japanese tourism images.

Keywords like “tradition”, “culture” and “performing arts” specified by Mr. Honda were positioned behind linearly related nouns in the text mining survey done at the time, and this was reflected in the visual manifestation of the Japanese regional promotion. This is why the terms “culture” and “entertainment” were nominated for prizes and won. This shows that text mining

of keywords in visual expressions will be helpful when considering visual expressions in future regional promotion videos.

ACKNOWLEDGMENTS

In writing this article, I would like to express my deepest gratitude to Tsuyoshi Kigawa, director of the Japan World's Tourism Film Festival and professor at Wakayama University's Faculty of Tourism, for supplying us with useful materials as well as his assistance and advice.

I would like to express my gratitude to Mr. Katsunosuke Honda of Honda-ya Honten for indicating the future path of regional promotion in Japan and for his numerous recommendations for this study.

REFERENCES

Higuchi, K. (2017). "Gengo Kenkyu no Bunya niokeru KH Coder Katsuyou no Kanousei". *Mathematical Linguistics*, Vol.31 No.1 pp.36-45.

IT Trend, (2022). "Now you know! The History and Evolution of Text Mining", Retrieved May 3, 2022, from https://it-trend.jp/textmining/article/history_and_evolution

JWTFF, 2022. Japan World's Tourism Film Festival. | Day1 | JWTFF 4th Japan World's Tourism Film Festival. Retrieved May 3, 2022, from <https://www.youtube.com/watch?v=7oXNGJZtd6c>

Matsubara, K. (2020). "If you come to this place, you can meet someone". Retrieved May 3, 2022, from <https://youtu.be/gOKH2-b78NI>

Matsubara, K. (2022). "There is nothing but a lot of mountains." Retrieved May 3, 2022, from <https://youtu.be/NsglCDDahDk>

Ogawa, T. (2022). "Factor Analysis of Film Evaluation Using Text Mining and Network Analysis". *Economic and Business Review*, Vol.29 No.2, March 2022.

Ogi, S. (2015). "Text Mining Techniques and Trends". *Computer Statistics*, Vol.28 No.1, pp.31-40.

User Local. (2022). "Could you elaborate on the "score", ". Retrieved May 3, 2022, from https://textmining.userlocal.jp/questions#data_q2.html