Earthquake's influence on inbound tourism: voices from the travel blogs

K. H. Murakami[1], H. Kawamura[2] & K. Suzuki[2]
[1]*School of Arts and Sciences, Shibaura Institute of Technology, Japan*
[2]*Graduate School of Information Science and Technology, Hokkaido University, Japan*

Abstract

Since the 2011 off the Pacific coast of Tohoku Earthquake damaged the major area of eastern Japan on March 11, 2011, the number of foreign visitors has drastically decreased. Japanese government tried to take urgent countermeasures against the loss. They spread the true information that Japan is safe by using media and Word-of-Mouth (WOM) such as influential bloggers. In this study, we investigate what WOM can be found regarding the earthquake through the blogs posted after the disaster. The blogs including earthquake-related words were collected and analyzed. Extracting the target data from the large number of data allows us to qualitatively examine the data. The text mining approach we used in this study includes morphological analysis and topic words analysis with term frequency-inverse document frequency weight. After that, the topic words and the words with high document frequency were used to find collocation of those topic words. As a result, the topic words "safe" and "normal" were found in the March 2011 blogs while they contained many other topic words about earthquake, tsunami, and nuclear crises. Moreover, gradually the number of earthquake-related words disappeared in the blogs and travel-related words appeared as time went by. In addition, we used concordance software to see the context where the topic words are used and found that the blogs written by the foreign visitors could convey true information about the disaster.
Keywords: 2011 Tohoku Earthquake, inbound tourism, travel blogs, text mining, topic words.

WIT Transactions on Ecology and The Environment, Vol 161, © 2012 WIT Press
www.witpress.com, ISSN 1743-3541 (on-line)
doi:10.2495/ST120041

1 Introduction

The 2011 off the Pacific coast of Tohoku Earthquake (the 2011 Tohoku Earthquake) occurred on March 11, 2011. The massive tsunamis caused by the M9.0 earthquake hit Fukushima I nuclear power plant and this led to nuclear meltdown later [1]. Because of this disaster and radiation accidents, the number of foreign visitors to Japan decreased dramatically in 2011. Fig. 1 displays the foreign visitor arrivals from January 2010 to February 2012 [2]. As can be seen in this figure, the arrivals in March, April, May, and June in 2011 dropped considerably. However, the countermeasures taken by the Japanese government to revive inbound tourism achieved a successful outcome and the number of foreign visitors has been gradually returning. Even though the affected areas were mainly Tohoku area and Kanto area, the decrease of the number of foreign visitors was seen all over Japan. Therefore, Japanese or local government invited foreign media and travel agencies to each prefecture to show the place was safe or Japan was safe, providing the real-time information about the radiation. Also, they invited influential foreign bloggers to the disaster area wanting them to write the progress of reconstruction in their blogs [3–5].

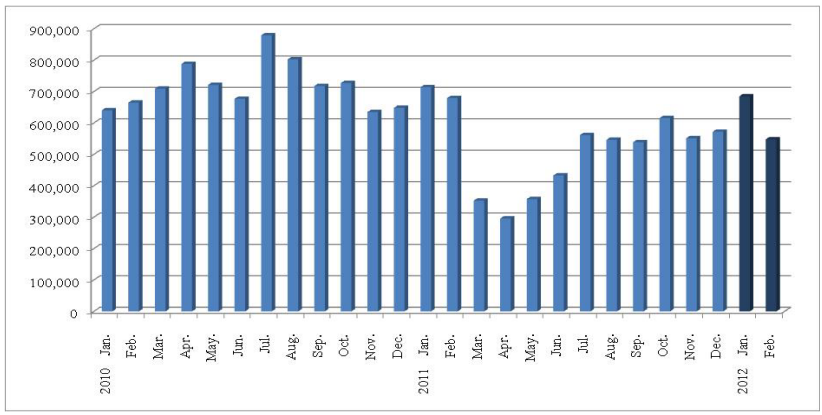

Figure 1: Trend of foreign visitor arrivals to Japan (source: Japan National Tourism Organization). *The numbers in January and February in 2012 are preliminary.

The word-of-mouth (WOM) can become either harmful rumors or countermeasures. It is important to know how to use the WOM as countermeasures. One of the WOM sources is blogs in tourism market. Blogs are used for tourism research to study destination images [6–8]. In this study, we investigate what WOM can be found regarding the earthquake through the bogs posted after the disaster by using the text mining approach.

2 Methodology

2.1 Data overview

In this study, we examine the travel blogs posted after the 2011 Tohoku Earthquake in order to see what WOM related to the disaster can be found in the blogs.

The data was retrieved from the travel blog website, TravelPod (www.travelpod.com). This is one of the largest travel blog websites in the world. The number of postings related to Japan travel is 15,267 as of March 6th, 2012. The total number of blogs we used for this study is 886. Only English blogs were collected for the analysis. The blogs posted from March 12, 2011 to February, 2012 were collected. Since there are a variety of contents those blogs have, only the blogs which have earthquake-related words were extracted to examine the contents of the blogs. The number of earthquake-related blogs is 156. Fig. 2 shows the number of English blogs and earthquake-related blogs of each month. In March, the number of earthquake-related blogs is the highest. April, July, August, November, and January are the following. In total, 17.6% of all blogs (156 blogs out of 886 blogs) contain earthquake-related words.

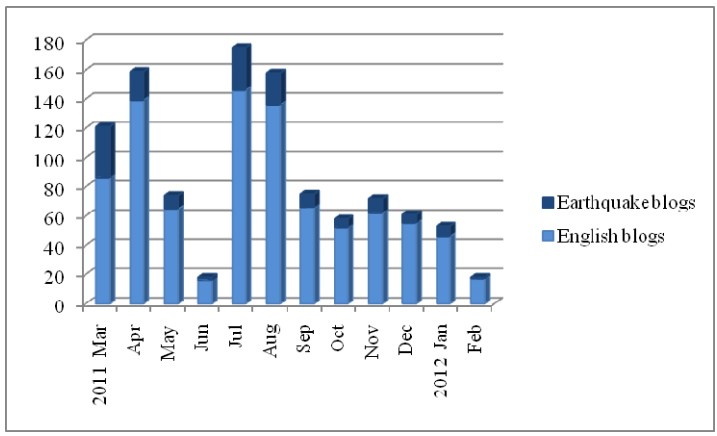

Figure 2: The number of English blogs and earthquake-related blogs.

We made the criteria to choose the earthquake-related words. First, the words which represent the disaster such as “earthquake”, “tsunami”, and nuclear crises or nuclear power plant (“nuclear”) were chosen. Therefore, those three words can be seen as the most related words. Second, synonyms or relevant words of “earthquake”, “disaster”, and “radiation” were selected by suing a thesaurus dictionary. Third, keywords whose number of search volume increased rapidly in 2011 in Google [9] were picked up. Finally, for the names of places, the reports by the Fire and Disaster Management Agency [10] were used. The names of the town/city where the number of deaths was more than 100 and the nuclear evacuation zones in Fukushima prefecture were chosen. From these criteria

54 words were selected. Table 1 shows the representative words of earthquake-related words. Therefore, "earthquake" includes its synonyms and "miyagi" includes the town/city names that were affected. The number in parentheses is the number of blogs in which either of 54 words appeared.

Table 1: Earthquake-related words (representative words).

earthquake (151)	tsunami (40)	nuclear (37)	disaster (29)
radiation (26)	aftershock (14)	evacuate (5)	blackouts (4)
miyagi (16)	fukushima (15)	iwate (4)	

2.2 Text mining

The earthquake-related blogs were sorted by the month when each blog was posted in order to see how many earthquake-related words appear in each month. The text mining approach was used for the earthquake-related blogs. For morphological analysis, open source software, TinyTextMiner (TTM) [11] was used. By TTM, the word frequency and the document frequency of all data were generated. After that, the term frequency-inverse document frequency (TF-IDF) weight was applied to find the topic words. TF-IDF weight is a weight used as a statistical measure to evaluate how important a word is in a corpus. Eqn. (1), (2), and (3) show the mathematical details as follows:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

$$idf_i = 1 + \log \frac{|D|}{\left|\left\{ j : t_i \in d_j \right\}\right|} \tag{2}$$

$$tfidf(i,j) = tf_{i,j} \times idf_i \tag{3}$$

where $n_{i,j}$ is the number of occurrences of term t_i in document d_j, $|D|$ is the total number of documents, and $|\{j: t_i \in d_j\}|$ is the number of documents that includes term t_i.

In addition, concordance software, AntConc [12], is used for collocation analysis to see the co-occurrence between topic words (search words) and other words, and concordance analysis to see how the topic words and phrases are commonly used in the target blogs. Fig. 3 illustrates a sample of collocation analysis results obtained from March 2011 blogs. In the collocation function of AntConc, there are two choices of statistical measures, which are mutual information and T-score. This study chose mutual information as the statistical measure. The equations used for this statistical measure are described in [13]. Fig. 4 shows a sample of concordance analysis results obtained from March 2011 blogs.

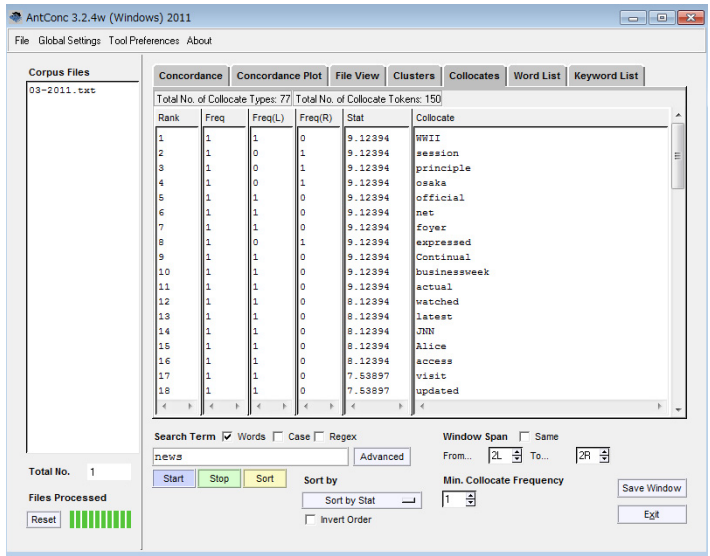

Figure 3: A sample of collocation analysis results.

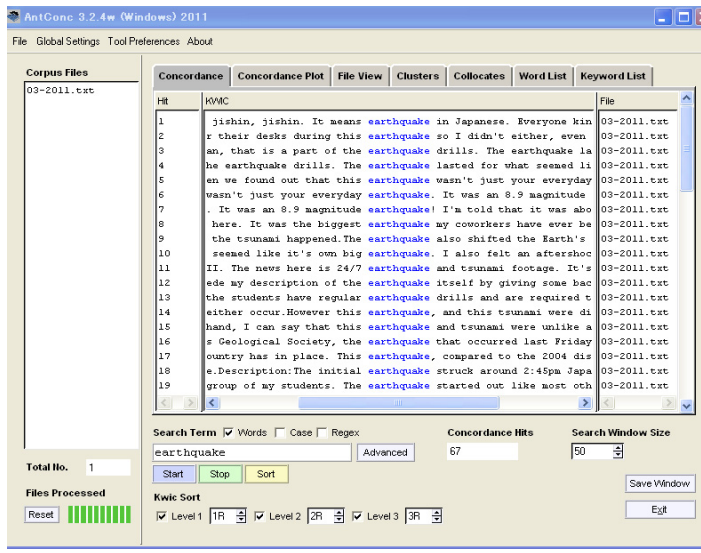

Figure 4: A sample of concordance analysis results.

2.3 Procedure of analysis

Our aim is to see what WOM can be found in the blogs posted after the 2011 Tohoku Earthquake. In order to achieve our goal, the following steps were taken.

(1) Collect the all English blogs related to Japan travel posted from March 12, 2011 to February, 2012 (886 blogs were collected). Sort the blogs by the posting date to deal with the blogs of the month as one document.

(2) Use TTM to produce the document frequency of each word. For this document frequency, nouns, adjectives, and verbs are selected. In this study, 30830 words were selected.

(3) Make a list of the earthquake-related words (54 words were chosen). Obtain the blogs containing those words (156 blogs were obtained).

(4) Use TTM for these 156 blogs to produce the term frequency and the document frequency of each word. Apply TF-IDF weight to extract topic words. In addition, not only TF-IDF weight, but also the words with high document frequency are used to obtain topic words.

(5) Use AntConc to analyze collocation of the topic words and to see how the topic words are used in the corpus.

3 Findings

The 886 English blogs and the 156 earthquake-related blogs were examined in this study. Fig. 5 illustrates the transition of the number of earthquake-related blogs in each month. Most of the time, several earthquake-related words among the 54 words were contained in one earthquake-related blog. The words used in this figure are representative words listed in Table 1. Even though the number of blogs collected for each month is different, this result can be considered as the

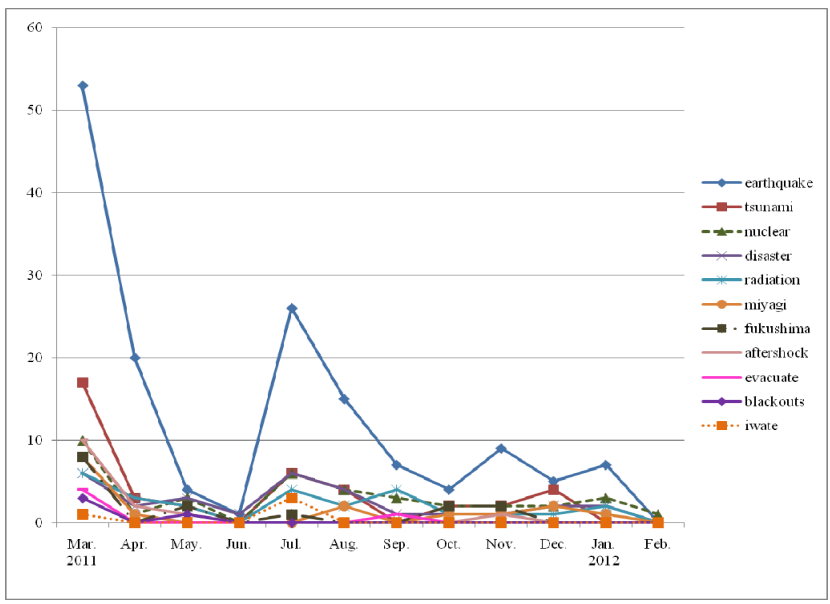

Figure 5: The number of blogs contained earthquake-related words.

amount of earthquake WOM. In March 2011 blogs, most of the earthquake-related words occurred with high frequency. However, the number of blogs for each word except "earthquake" didn't occur much after April blogs. This can be assumed that many blogs were written about earthquake in March when the 2011 Tohoku Earthquake hit Japan, but the blogs with earthquake topic did not last long and the earthquake-related words decreased from the following month.

Table 2: Top 10 topic words with TF-IDF in blogs of each month.

2011 Mar	TF-IDF	Apr	TF-IDF	Jul	TF-IDF
earthquake	0.0097	people	0.0075	day	0.0053
people	0.0075	see	0.0070	time	0.0052
day	0.0074	time	0.0066	see	0.0052
hours	0.0067	way	0.0055	tokyo	0.0052
tsunami	0.0066	believe	0.0044	people	0.0048
news	0.0061	stomach	0.0043	festival	0.0042
everyone	0.0058	good	0.0042	few	0.0042
aftershock	0.0058	earthquake	0.0041	many	0.0040
time	0.0051	bus	0.0041	temple	0.0038
know	0.0051	day	0.0038	first	0.0035
Aug	TF-IDF	Nov	TF-IDF	2012 Jan	TF-IDF
time	0.0065	day	0.0074	bomb	0.0075
day	0.0061	coffee	0.0071	people	0.0072
few	0.0057	hiroshima	0.0062	hiroshima	0.0069
people	0.0047	next	0.0050	first	0.0056
way	0.0047	time	0.0048	peace	0.0054
hostel	0.0045	park	0.0047	tower	0.0046
see	0.0043	lake	0.0046	time	0.0045
Tokyo	0.0043	aerobics	0.0046	Tokyo	0.0045
go	0.0043	trip	0.0041	few	0.0044
police	0.0041	water	0.0041	station	0.0044

3.1 TF-IDF and document frequency for topic words

In order to see this assumption is correct, we analyzed the earthquake-related blogs. Nouns, adjectives, and verbs were extracted by TTM and TF-IDF weight was applied to each word. The top 10 words were selected according to TF-IDF value. The words "Japan", "Japanese", person's names were excluded from the selection since the blogs used for this analysis are all related to Japan travel. Also

contain earthquake-related words even though all the data for this analysis is earth related blogs. In July, August, November, and January 2012 blogs, travel-related words such as “Tokyo”, “Hiroshima”, or “temple” are ranked within the top 10.

Additionally, we used the words with high document frequency in the 156 earthquake-related blogs as topic words. The number of words and a variety of word choices are depending on each blogs. Therefore, we chose 20 words with high document frequency from the top and we deleted the words which are found in other documents (blogs of other months). The words selected through this process are unique to the document. For example, “tsunami” is only found in March 2011 blogs among the top 20 high frequency words. Table 3 shows the topic words that are unique to each month and their document frequencies.

Table 3: Topic words that are unique to the blogs of each month and their document frequencies.

2011 Mar		Apr		Jul	
tsunami (n)	14	usual (j)	7	years (n)	16
news (n)	14	arrived (v)	7	end (n)	15
everyone (n)	11	traditional (j)	6	sure (j)	12
happened (v)	11	western (j)	6	told (v)	11
nuclear (j)	10	extra (j)	6		
normal (j)	9	dinner (n)	6		
safe (j)	9	everything (n)	6		
hit (v)	9	eating (v)	6		
trying (v)	8	best (j)	5		
running (v)	8	happy (j)	5		
		lovely (j)	5		
Aug		Nov		2012 Jan	
lot (n)	13	decided (v)	8	building (n)	10
room (n)	11	whole (j)	7	second (j)	6
museum (n)	11	bus (n)	7	something (n)	5
early (j)	9	hot (j)	6	floor (n)	5
		free (j)	5	signs (n)	5
				built (v)	5
				main (j)	4
				tiny (j)	4

In March 2011 blogs, “tsunami” and “news” are also found as topic words with this process. Additionally, “happened”, “nuclear”, “normal”, or “safe” are found. This result also explains the content of March 2011 blogs mainly is about earthquake.

Table 4: A part of collocation analysis results with mutual information value.

earthquake (Mar)		tsunami (Mar)		safe (Mar)	
drills	8.450	unprecedented	8.844	very	10.472
wordpress	7.865	unlike	8.844	relocation	10.472
unprecedented	7.865	threat	8.844	pray	10.472
society	7.865	sweeping	8.844	please	9.472
smaller	7.865	survival	8.844	organization	8.887
shindo	7.865	subsequent	8.844	kept	8.472
science	7.865	nowhere	8.844	currently	8.150
sat	7.865	minimizes	8.844	away	8.150
rumbled	7.865	issued	8.844	pretty	7.887
regular	7.865	hopefully	8.844	within	7.472
Tokyo (Jul)		festival (Jul)		years (Jul)	
welcome	8.593	matsuri	11.250	thumbing	10.250
weekends	8.593	thanked	10.250	teaching	10.250
synonyms	8.593	takayama	10.250	strategized	10.250
structures	8.593	organizers	10.250	pictured	10.250
string	8.593	natsumatsuri	10.250	nine	10.250
square	8.593	natsu	10.250	insolvent	10.250
shortly	8.593	matsumoto	10.250	founded	10.250
shinkuku	8.593	leaving	10.250	collected	10.250
parks	8.593	jazz	10.250	ago	9.791
open	8.593	holding	10.250	upwards	9.250
coffee (Nov)		Hiroshima (Nov)		free (Nov)	
tasty	8.971	castle	9.141	shuttle	11.041
strong	8.971	yaki	8.556	admissions	11.041
spill	8.971	stationed	8.556	wow	10.041
newspapers	8.971	river	8.556	scored	10.041
leaving	8.971	recommended	8.556	castle	10.041
downing	8.971	prayer	8.556	enter	9.456
continued	8.971	park	8.556	bus	7.582
black	8.971	memorial	8.556	coffee	6.649
iced	8.386	intense	8.556	time	6.582
wow	7.971	grounds	8.556	which	6.234

3.2 Collocation and concordance analyses

By using AntConc, we investigated collocation of the topic words to see what words co-occur with the topic words. Table 4 shows a part of the result for collocation analysis. Three topic words were chosen subjectively from the March, July and November blogs and top 10 co-occurrence words were listed with their mutual information value. Furthermore, we examined the context where the topic words are used by concordance analysis.

The results from these analyses give us an idea of what the blogs of the month mainly talked about. Especially the 2011 March blogs are about: “earthquake” and “tsunami” that “happened” were “unprecedented” or “unlike”; people who wrote the blogs wanted to tell that they were “very” “safe” and the places where they stayed were “currently” “safe” because they were “away” from the destruction; in the places where they were, the life was returning to “normal” “day” by “day”. These quoted words are either the topic words or collocation words. By looking at the context where those topic words are used, the WOM can be described with those words and their collocation words.

4 Conclusion

In this study, we analyzed the blogs containing the earthquake-related words by using text mining approach. Our aim for this study is to investigate the WOM after the 2011 Tohoku Earthquake in Japan from the travel blogs posted after the earthquake from March 12, 2011 to February 2012. As a result of finding the topic words applying TF-IDF weight, extracting unique high frequency words, and seeing the topic words in context by using concordance software, we found that the number of earthquake-related words gradually disappeared in the blogs and travel-related words appeared as time went by. Especially, in the blogs of March 2011 when the Tohoku Earthquake occurred, the main content is about the truth of the disaster, which is the real happening in Japan at that time. Sending out the “safe” message via media or influential bloggers is still very important countermeasure when the crises happened. However, it is also important to pass on the true information (WOM) by foreign visitors who travel in Japan after the crises.

Usually harmful rumors are spread after the crises and they could last quite long because of the confusion of real information. We found that the blogs written by the foreign visitors could convey true information right after the disasters. It is suggested that sending out this type of WOM through the media sources. It can avoid the delay in the reconstruction of the inbound tourism.

References

[1] Japan Meteorological Agency, Retrieved July 20, 2011 from http://www.jma.go.jp/jma/en/2011_Earthquake.html

[2] Japan Tourism Marketing, Statistics of visitors to Japan from overseas (original source from Japan National Tourism Organization), Retrieved March 10, 2012 from http://www.tourism.jp/english/statistics/inbound.php
[3] Hokkaido Shinbun, Hokkaido is safe after the earthquake (in Japanese), the tenth page (Jun. 29, 2011).
[4] Nikkei Shinbun, Bringing back our foreign tourists to Japan (in Japanese), (Aug. 4, 2011).
[5] The Japan Times, http://www.japantimes.co.jp/text/nn20120207a8.html (Feb. 7, 2012).
[6] Pan, B., MacLaurin, T. and Crotts, J., Travel blogs and the implications for destination marketing, *Journal of Travel Research*, 46(1), pp. 35-45, 2007.
[7] Wenger, A., Analysis of travel bloggers' characteristics and their communication about Austria as a tourism destination, *Journal of Vacation Marketing*, 14(2), pp.169-176, 2008.
[8] Murakami, K., Kawamura, H. and Suzuki, K., Analysis of travel blogs: foreigners' word of mouth on Japan travel, *Proceedings of the 12th Asia Pacific Industrial Engineering and Management System Conference*, pp.300-305, 2011.
[9] Google, Annual Search Ranking 2011, Retrieved March 6, 2012 from http://www.google.co.jp/press/zeitgeist2011/hottrends.html
[10] Fire and Disaster Management Agency, Retrieved February 26, 2012 from http://www.fdma.go.jp/bn/higaihou/pdf/jishin/144.pdf
[11] Matsumura, N. and Miura, A., TinyTextMiner, Retrieved February 26, 2012 from http://mtmr.jp/ttm
[12] Anthony, L., AntConc, Retrieved March 10, 2012 from http://www.antlab.sci.waseda.ac.jp/software.html
[13] Stubbs, M., Collocations and semantic profiles: On the cause of the trouble with quantitative studies. Originally published in *Functions of Language*, 2(1), pp.23-55, 1995. Online Address: http://www.uni-trier.de/fileadmin/fb2/ANG/Linguistik/Stubbs/stubbs-1995-cause-trouble.pdf. Accessed: March 10, 2012.