CATEGORY EFFECT OF KEYWORDS ON HEALTHCARE INDUSTRY BY TEXT MINING IN WEB

[1]JAE-WON HONG, [*2]JAE-YOUNG MOON

[1]Department of Global Trade, Gyeongnam National University, 52725, Korea

[2]Division of Business, Dongseo University, 47011, Korea

E-mail: [1]jwhong@gntech.ac.kr, [2]jaymoon@gdsu.dongseo.ac.kr

ABSTRACT

This study is intended to investigate the effects of keywords by category on the diffusion of consumer's interest in healthcare industry focus on tourism using data mining in web environment. And the results are as follows: first, in the case of health care-related keywords, search activities for the categories of finance, biz, book, and social categories are found to affect the diffusion of medical tourism. Second, in the case of the physical examination keywords, search activities for the categories, such as Shopping, Travel, Social, community, are found to have some effect on the diffusion of medical tourism. Third, the role played by key words that affect the diffusion of medical tourism is appeared to differ depending on their categories. That is, for the part of health care, the categories of premeditation attribute, such as, finance, book, biz, travel, found to have impact on the diffusion of medical tourism, while in the case of physical examination, categories of impulsion, such as shopping, sports, food, are found to affect the diffusion of medical tourism. In other words, it is implied that the attributes of keywords and categories should be considered in communicating for the diffusion of medical tourism in web environment.

Keywords: *Keyword Analysis, Text Mining, Healthcare, Tourism, Web Environment*

1.   INTRODUCTION

Tourism means, "Travel for pleasure; also the theory and practice of touring, the business of attracting, accommodating, and entertaining tourists, and the business of operating tours." [1]. Korea takes the 8th place in the world in terms of the size of consumption in Korea by tourists to Korea and the number of visitors is expected to continue to increase over time [2].

Table 1: International Tourism Expenditure in 2015

Rank	Country	UNWTO Region	International tourist arrivals (2015)
1	China	Asia	$292.2 billion
2	U.S.	North America	$112.9 billion
3	Germany	Europe	$77.5 billion
4	UK	Europe	$63.3 billion
5	France	Europe	$38.4 billion
6	Russia	Europe	$34.9 billion
7	Canada	North America	$29.4 billion
8	South Korea	Asia	$25.0 billion
9	Italy	Europe	$24.4 billion
10	Australia	Oceania	$23.5 billion

The World Tourism Organization reports the following destinations as the top ten biggest spenders on international tourism for the year 2015

As such, Korea has secured its place as a country of tourism in the World attracting a lot of tourists worldwide. Recently, a new type of tourism emerged out of these various types of tourism that combines the functions of medical services, recuperation, leisure and cultural events, and many countries are developing medical tourism as an industry with heavy investment under a new

paradigm of tourism [3]. Korea should have the same level of medical services as developed counties while keeping medical costs low, to be able to offer competitive medical tourism services. Especially, in the case of medical tourism, the length of stay is longer compared to other types of tourism, making it an ideal for earning foreign currency; so many countries have strong supporting policy for medical tourism at their government level.

As suggested in the second Korea report of McKinsey, the areas that can provide new job opportunities and industry development the most are health care, social welfare service, financial service and tourism [4]. In the case of Korea, growth potential for tourism is high because the quality of medical services and infra for tourism is good while medical costs are competitive, which are good conditions for medical tourism as indicated already, backed by continuing interest of patients in Korean wave and increasing visitors from Asian countries [5].

Korean government strongly supports medical tourism as a high value-added industry by combining existing tourism industry and medical technology, and the total revenue from tourism amounted to 0.2 billon US dollars in 2013, and the number of visitors was 211,218 people from 191 countries [6]. Most of the visitors on medical tourism were for plastic surgery, but recently, more and more people are visiting Korea for physical examination services and medical treatments.

Even though Korea has globally recognized level of 'medical skill', product developments, investments and marketing are not as robust as expected due to lack of awareness in 'medical tourism' and various regulations.

In the study, therefore, data of search activities for material keywords provided by Google are collected by week and category for past 5 years period worldwide, and keywords mainly used by those foreign visitors to Korea are looked at in an effort to help increase medial tourists with a more robust marketing.

2.   RELATED WORK

According to Han and Kim [7], the concept of purposeful medical tour existed in BC 3 when aristocrats in ancient Rome used to visit resorts, such as Cyprus and Alexandria for recuperation, and this concept continued to survive into 19 century by Britons who were then Celt living in southern part of Great Britain and used to visit spar resorts in Germany in sear of clean water, and Korea also has records in veritable records of the Chosen court that kings of Chosen Dynasty used to visit hot springs, such as OnYang Hot spring in ChungChung province in Korea [8].

Medical tourism can be defined as a new type of tourism in which medical services and tourism services are combined to attract more tourists by offering high level of medical services and various products for tourists [3]. In the past, most of medical tourism was served by wealthy countries as people in poor countries traveled to wealthy countries and the country that experienced the most visitors was United States of America. In 70s-80s, medical tourism had strong negative image because a lot of people visited USA to give birth to children or to avoid military services, but today, medical tourism has developed into a new type of tourism and made popular in various facets.

Korea has become a popular destination for medial tourism in the world. And various data show evidences that Korean medical skills have reached the same level as developed. But high proportion of medical tourists to Korea are still for plastic surgery and physical examination services, it is, therefore, necessary to have policy level support and market strategy to provide better diversified medical tourism services.

According to Connell [9], the factors that contribute to the growth in medial tourism are high level of medical spending by tourists and high level of medical services in the destination countries, favorable air fares and foreign exchange rates, and high level of medical costs in the country of destination.

Goodrich [10], medical tourism is a combination of health services and tourist products, as a purposeful attempt to attract tourists. Hall [11] said that medical tourism is one type of tourism but its main point is health-related. Law [12] defined medical tourism as activities of tourism and medical services that occur by or for individuals away from their residences.

As such, looking at medical tourism based on existing studies, we need to make more serious efforts to attract more high risk patients beyond the areas of medical treatments for eyes, skin, ear-nose-throat and plastic surgery, by combining high level of medical services and low level of medical costs and tourist products.

3.   RESEARCH PROGRESS

3.1 Data Collection by Data Mining

The data used in this study is data on weekly amounts of keywords search provided by Google data mining engine. In this study therefore, 100 indicates a week of the largest amount of

search activities for given period of time, and other weeks are expressed in their relative values based on the scale of 100 thereof, showing the keywords distribution by week clear. And for certain keyword, it shows search activities by category, such as health, games, science, finance, news, books, business, governmental organizations, shopping, sports, food, travel, communities, social groups, hobby and leisure, and it also provide the data by country and reason, thereby presenting the frequency distribution of keywords on global scale.

For this study, we collected search activities by category and week for 5 years from November 2011 to October 2016.

3.2 Choosing Keywords

In this study, we search related major keywords based on search activities related to medical tourism using data mining method from Google, and investigate the effect of the search patterns of keywords by the consumer of medical tourism. The keyword analysis of the document unit level is shown in Figure 1 [13-14].

Table 2: Search Volume of keyword 'Medical Tourism' by Country

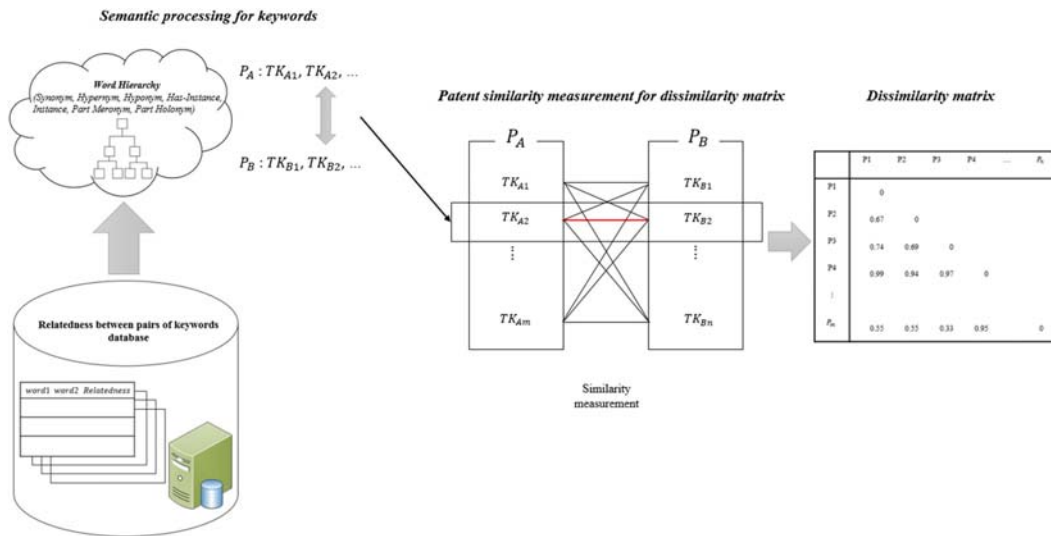| Rank	Country		City	
	Name	Index	Name	Index
1	Singapore	100	Gurgaon	100
2	India	78	Singapore	36
3	Malaysia	58	New Delhi	33
4	Philippine	56	Bengaluru	30
5	Australia	25	Mumbai	30
6	Canada	24	Pune	18
7	United States	19	Seoul	17
8	United Kingdom	10	Toronto	11

Figure 1: Procedure for keyword dissimilarity matrix

To this end, the keyword of 'medical tourism' is selected as a major keyword based on the amount and patterns of keywords search by the consumer of medical tourism, and the two terms 'physical examination' and 'health care' are selected as key words that affect medical tourism.

Of the data collected, search activities for medical tourism are observed the most in Singapore, and the next countries are India, Malaysia and Philippines, Australia, Canada, United States, United Kingdom, respectively. Of the cities, the biggest amount of search activities was observed in Gurgao of India and the next cities are Singapore and New deli, Bengaluru, Mumbai, Pune, Seoul, Toronto, respectively.

Table 3: Definition of Variables

Variable		Definition
Medical tourism		Amount of keyword searching a week: Category of "medical tourism"
Healthcare / Physical examination	Health	Amount of keyword searching a week: Category of "health"
	Finance	Amount of keyword searching a week: Category of "finance"
	News	Amount of keyword searching a week: Category of "news"
	Book	Amount of keyword searching a week: Category of "book & literature"
	Beauty	Amount of keyword searching a week: Category of "beauty & fitness"
	Biz	Amount of keyword searching a week: Category of "business & industry"
	Government	Amount of keyword searching a week: Category of "government"
	Shopping	Amount of keyword searching a week: Category of "shopping"
	Sports	Amount of keyword searching a week: Category of "sports"
	Food	Amount of keyword searching a week: Category of "food & beverage"
	Travel	Amount of keyword searching a week: Category of "travel"
	Community	Amount of keyword searching a week: Category of "online community"
	Social	Amount of keyword searching a week: Category of "group & social"
	Hobby	Amount of keyword searching a week: Category of "hobby & leisure"

3.3 Research Model

In this study, the dependent variable is the amount of search activities for the keyword 'medial tourism', and the independent variables are the amount of search activities for sub-categories of 'healthcare' and 'physical examination'. That is, there are 14 sub-categories, which are Health, Finance, News, Book, Beauty, Biz, Government, Shopping, Sports, Food, Travel, Community, Social, and Hobby. But categories, such as games, science, real estate, pet, are excluded because they are considered to have little relevance medical tourism.

4. RESULTS

4.1 Feature of Variable

During the period of time, mean search activity by theme for 'medical tourism' is found to be 59.540(STD=8.363). The highest score in the category of 'healthcare' is for 'beauty (M=80.920, STD=7.393), and the next highest scores are for biz (M=79.238, STD=6.763), Hobby (M=78.686, STD=7.071), respectively. In the category of Physical Examination, Health (M=79.057, STD=8.883) is the only one that recorded a high level of score, unlike with the 'healthcare' category, but the other categories are observed small, with a high level of standard deviation for search activities by theme.

Table 4: Characteristic of Statistical Variable

Variable		Frequency	Mean	Standard deviation	Minimum value	Maximum value
Medical Tourism		261	59.540	8.363	39	100
Healthcare	Health	261	64.241	6.666	36	100
	Finance	261	63.146	9.471	40	100
	News	261	43.429	7.419	21	100
	Book	261	63.333	11.912	31	100
	Beauty	261	80.920	7.393	46	100
	Biz	261	79.238	6.763	46	100
	Government	261	38.762	6.924	19	100
	Shopping	261	72.502	6.463	47	100
	Sports	261	70.977	9.076	34	100
	Food	261	69.460	7.667	44	100
	Travel	261	73.900	7.423	43	100

	Community	261	70.805	10.076	32	100
	Social	261	47.625	7.542	24	100
	Hobby	261	78.686	7.071	49	100
Physical Examination	Health	261	79.057	8.883	47	100
	Finance	261	37.410	12.910	0	100
	News	261	37.927	14.133	0	100
	Book	261	27.989	11.926	15	100
	Beauty	261	29.667	12.852	0	100
	Biz	261	44.418	15.670	19	100
	Government	261	52.866	17.530	18	100
	Shopping	261	23.533	11.640	12	100
	Sports	261	26.571	12.803	0	100
	Food	261	34.433	17.187	0	100
	Travel	261	34.843	14.619	0	100
	Community	261	28.287	23.324	0	100
	Social	261	39.548	13.466	15	100
	Hobby	261	34.525	14.343	18	100

Table 5: Correlation between Medical tourism and Healthcare

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0.27													
3	0.01	0.74												
4	0.24	0.92	0.72											
5	0.47	0.49	0.24	0.44										
6	0.14	0.61	0.43	0.46	0.31									
7	0.40	0.72	0.42	0.53	0.52	0.60								
8	0.25	0.91	0.68	0.99	0.45	0.45	0.50							
9	0.33	0.70	0.50	0.61	0.56	0.38	0.66	0.60						
10	0.27	0.67	0.41	0.54	0.33	0.45	0.62	0.53	0.50					
11	0.07	0.72	0.69	0.71	0.28	0.46	0.45	0.70	0.47	0.47				
12	0.19	0.59	0.32	0.44	0.25	0.56	0.65	0.43	0.40	0.53	0.35			
13	0.17	0.57	0.40	0.49	0.26	0.39	0.50	0.48	0.42	0.40	0.36	0.34		
14	0.28	0.93	0.73	0.99	0.46	0.47	0.53	0.98	0.64	0.55	0.72	0.43	0.50	
15	0.27	0.64	0.50	0.58	0.42	0.40	0.54	0.58	0.65	0.53	0.49	0.43	0.40	0.60

1=Medical Tourism, 2-15=Healthcare [2=Health, 3=Finance, 4=News, 5=Book, 6=Beauty, 7=Biz, 8=Government, 9=Shopping, 10=Sports, 11=Food, 12=Travel, 13=Community, 14=Social, 15=Hobby]

Table 6: Correlation between Medical tourism and Physical Examination

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
2	0.25													
3	0.12	-0.07												
4	0.15	-0.04	0.16											
5	0.11	0.07	0.18	0.18										
6	0.06	-0.05	0.16	0.11	0.11									
7	0.06	0.09	0.08	0.03	0.03	0.10								
8	0.10	0.17	-0.09	0.06	0.08	0.00	-0.05							
9	0.30	0.04	0.16	0.19	0.18	0.13	0.06	0.00						

10	0.16	-0.04	0.19	0.27	0.14	0.09	0.06	0.10	0.13					
11	0.09	-0.02	0.15	0.12	0.03	0.04	0.11	-0.07	0.15	0.09				
12	0.23	-0.05	0.13	0.08	0.09	0.14	-0.01	0.06	0.29	0.16	0.11			
13	0.12	0.27	0.17	0.06	0.16	-0.05	0.10	0.11	0.07	0.01	-0.04	-0.11		
14	0.16	0.23	0.09	0.04	0.09	0.10	0.10	-0.01	0.06	0.00	0.08	0.07	0.02	
15	0.16	0.05	0.05	0.15	0.24	0.07	0.04	-0.03	0.31	0.13	0.07	0.09	0.12	0.12

1=Medical Tourism, 2-15=Physical Examination [2=Health, 3=Finance, 4=News, 5=Book, 6=Beauty, 7=Biz, 8=Government, 9=Shopping, 10=Sports, 11=Food, 12=Travel, 13=Community, 14=Social, 15=Hobby]

The results of correlation analysis for search activities for Medical tourism and the search activities for the healthcare category indicate that there are significant relations, except finance and food, and the correlations between independent variables appears relatively high (See Table 5).

According to the results of correlations analysis between the search activities for 'medical tourism' and the search activities for 'physical examination' by category, correlations are found to be comparatively low between dependant variables and independent variables, and between inter-independent variables (Table 6) as well.

4.2 Results of Empirical Test

The following are the results of regression analysis by stepwise method of the effects of the sub-categories of healthcare-related keywords. And this methodology is to avoid Multicollinearity occurring due to high correlations between these variables.

Table 7. Effects of Healthcare on Medical tourism

Variables	B	β	t-value	p-value
Finance	-0.265	-0.300	-3.7***	0.000
Book	0.199	0.284	4.5***	0.000
Beauty	-0.128	-0.113	-1.7	0.091
Biz	0.408	0.330	4.5***	0.000
Food	-0.162	-0.148	-1.9	0.064
Social	0.383	0.346	3.8***	0.000

*p<.05, **p<.01, ***p<.001, F-Value=21.04, Adj-R2=0.316

In the category of 'healthcare', the following 6 variables are found to have significant effects, with the categories of biz, book, and social being large effects. And this implies that the diffusion of medical tourism can be augmented by vitalizing healthcare-related businesses, publicity through books and strategic approach to social relations. It is also determined that the diffusion decreases when consumers take into account the financial aspect of healthcare.

On the other hand, the same methodology is employed for the regression analysis of the effects of the sub-categories of physical examinations-related keywords. According to the results, 'shopping' category is found to have the biggest effect, whereas the other factors are found to have similar level of effect. And this implies that it is necessary to improve the 'shopping' category, which is a secondary factor to the 'physical examination', for the diffusion of medical tourism.

Table 8. Effects of Physical Examination on Medical tourism

Variables	B	β	t-value	p-value
Shopping	0.166	0.231	3.8***	0.000
Sports	0.071	0.109	1.9	0.063
Travel	0.084	0.147	2.4*	0.017

| Community | 0.041 | 0.115 | 2.0* | 0.049 |
| Social | 0.084 | 0.135 | 2.4* | 0.020 |

*$p<.05$, **$p<.01$, ***$p<.001$, F-Value=9.55, Adj-R2=0.141

Lastly, when considering all the sub-categories of 'healthcare' and 'physical examination', the following are the effects on the search activities for medical tourism. That is, the diffusion of medical tourism can be expanded further by stimulating social groups to help business side of healthcare, and by improving shopping, truism and secondary services as well.

Table 9. Effects of Healthcare and Physical Examination on Medical tourism

Variables		B	β	t-value	p-value
health care	Finance	-0.221	-0.250	-3.3**	0.001
	Book	0.153	0.218	3.5**	0.001
	Biz	0.388	0.314	4.1***	0.000
	Travel	-0.151	-0.134	-2.0*	0.045
	Social	0.370	0.334	4.1***	0.000
physical examination	Shopping	0.131	0.182	3.4**	0.001
	Sports	0.072	0.110	2.2*	0.030
	Food	0.051	0.105	2.1*	0.041
	Travel	0.074	0.130	2.5*	0.015

*$p<.05$, **$p<.01$, ***$p<.001$, F-Value=19.01, Adj-R2=0.384

5. DISCUSSION

In this study, the categorical effects of keywords on the diffusion of medical tourism are analyzed based on the proliferation of keywords by Google data mining method in web [15]. That is, of major keywords factors affecting the search by medical tourism keywords by consumers, it is search frequencies by category of the 'healthcare' and 'physical examination' keywords that is employed to explain the effects.

The findings are as follows: first, of the health care-related keywords, the search activities for the categories of finance, biz, book and social are found to have effect on the diffusion of medical tourism. And this implies that for the diffusion of medical tourism, it is necessary to vitalize health care-related businesses/industries, employ reference effect of social groups, and to provide necessary financial services.

Second, of physical examination-related keywords, search activities for categories, such as shopping, travel, social, community, are found to have effects on the diffusion of medical tourism. And this implies that physical examination services should be combined with shopping and travel services, and appropriate communication efforts should be made to socially promote such services throughout respective community.

Third, the roles of keywords that affect the diffusion of medical tourism are found to differ by category. In other words, in the case of health care, categories of premeditation attribute, such as finance, book, biz and travel, are fund to affect the diffusion of medical tourism, while in the case of physical examination, categories of impulsion attribute, such as shopping, spots, food, are found to affect the diffusion of medical tourism. And this implies that some communication strategy that takes into account both attributes of keywords and categories is necessary for the diffusion of medical tourism.

This study is meaningful from the view point that categorical effects of related keywords are considered for the diffusion of medical tourism, thereby trying to provide companies and policy makers with some marketing communication strategy to promote medical tourism, based on such findings.

Despite these efforts, however, there are some limitations. First, the study does not reflect social, economic and cultural characteristics of different countries. Second, keywords that are related to medical tourism are not sufficiently reflected. In future studies, therefore, country characteristics, keywords structures, and various web environments should be considered as well.

ACKNOWLEDGMENTS

REFRENCES:

[1] The World Tourism Organization, Available: https://en.wikipedia.org/wiki/World_Tourism_rankings, 2015

[2] The World Tourism Organization, The World Tourism Organization reports the following destinations as the top ten biggest spenders on international tourism for the year 2015, 2015.

[3] Korea Health Industry Development Institute, Research of Advanced Healthcare System and Healthcare Industry Promotion, 2006.

[4] McKinsey, the 2nd Korea Report, 2013[internet], Available: https://www.alog.auric.or.kr/FileDownLoad.aspx?alogid=KG2530&fileNum=7023.

[5] M. K. Kim, S. M. Yoon, and S. N. Yoon, "The Exploratory Study on Development of Medical Tourism Product in Korea," *The Tourism Sciences Society of Korea*, Vol. 32, No. 6, 2008, pp. 431-447.

[6] Korea Health Industry Development Institute, Survey of Inviting Foreign Patients on 2013, 2014.

[7] K. S. Han, S. M. Kim, Understanding International tourism, Beak San, Seoul, 2007.

[8] H.Y. Lee, "The Effect of Perceived Medical Tourism Factors on Intention to Visit: Focused on Food Culture & Perceived Risk," Ph.D. dissertation, The Graduate School of Sejong University, 2011.

[9] J. Connell, "Medical tourism: Sea, sand, and surgery," *Tourism Management*, Vol. 27, No. 6, 2006, pp. 1093-1100.

[10] J. N. Goodrich, "Health-care tourism - an exploratory study," *Tourism Management*, Vol. 8, No. 3, 1987, pp. 217-222.

[11] C. M. Hall, Adventure, Sport, and Health Tourism, In B Weiler & C. M Hall(Ed). Special-Interest Tourism, Bellhaven Press, London, 1992.

[12] E. Law, Health and the International Tourist, Health Tourism: A Business Opportunity Approach. Ins, Cliff & S. J. Page (eds.), London: Routledge, 1996.

[13] Joung, Junegak, and Kwangsoo Kim, "Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data," *Technological Forecasting and Social Change*, Vol. 114, 2017, pp. 281-292.

[14] Moehrle, M.G., "Measures for textual patent similarities: a guided way to select appropriate approaches." *Scientometrics*, Vol. 85, 2010, pp. 95-109.

[15] J.W. Hong, and J. Y. Moon, The Category Effect of Keyword of Medical Tourism, *International Conference on Computing Convergence and Applications 2016 (January)*, Busan, Korea, 2016.