International Journal of Internet, Broadcasting and Communication Vol.15 No.3 211-218 (2023)
http://dx.doi.org/10.7236/IJIBC.2023.15.3.211

IJIBC 23-3-26

Big Data Analysis on the Perception of Home Training According to the Implementation of COVID-19 Social Distancing

Hyun-Chang Keum[1], Kyung-Won Byun[2]

[1]*Ph D. Student of Department of Sport Interaction Science, The Graduate School of Sungkyunkwan University, Korea*
lafirst@gmail.com
[2]*Assistant Professor, Department of graduate school of business Administration, Dankook University, Korea*
changewon125@ dankook.ac.kr

Abstract

Due to the implementation of COVID-19 distancing, interest and users in 'home training' are rapidly increasing. Therefore, the purpose of this study is to identify the perception of 'home training' through big data analysis on social media channels and provide basic data to related business sector. Social media channels collected big data from various news and social content provided on Naver and Google sites. Data for three years from March 22, 2020 were collected based on the time when COVID-19 distancing was implemented in Korea. The collected data included 4,000 Naver blogs, 2,673 news, 4,000 cafes, 3,989 knowledge IN, and 953 Google channel news. These data analyzed TF and TF-IDF through text mining, and through this, semantic network analysis was conducted on 70 keywords, big data analysis programs such as Textom and Ucinet were used for social big data analysis, and NetDraw was used for visualization. As a result of text mining analysis, 'home training' was found the most frequently in relation to TF with 4,045 times. The next order is 'exercise', 'Homt', 'house', 'apparatus', 'recommendation', and ‘diet'. Regarding TF-IDF, the main keywords are 'exercise', 'apparatus', 'home', 'house', 'diet', 'recommendation', and 'mat'. Based on these results, 70 keywords with high frequency were extracted, and then semantic indicators and centrality analysis were conducted. Finally, through CONCOR analysis, it was clustered into ‘purchase cluster’, ‘equipment cluster’, ‘diet cluster’, and ‘execute method cluster’. For the results of these four clusters, basic data on the 'home training' business sector were presented based on consumers' main perception of 'home training' and analysis of the meaning network.

Keywords: *Home Training, Homt, COVID-19 Social Distancing, Big Data Analysis, Text Mining, Semantic Network Analysis, CONCOR Analysis*

1. Introduction

The unprecedented COVID-19 pandemic has been dealt with untact policies such as blockades between regions, social distancing, restrictions on use, telecommuting, and non-face-to-face education, which has

Manuscript Received: July. 19, 2023 / Revised: July. 23, 2023 / Accepted: July. 27, 2023
Corresponding Author: changewon125@dankook.ac.kr
Tel: +82-31-8005-3444, Fax: +82-31-8021-7137
Assistant Professor, Department of Graduate School of Business Administration, Dankook University

brought many changes in all areas of society, including economy, society, culture, and education. Because the degree of change is so great, it is divided into BC(Before Corona) and AD(After Disease) based on before and after COVID-19 [1]. In March 2020, the Korean government implemented strong social distancing. As a result, it also had a significant impact on sports activities. Due to concerns about infectious diseases, outdoor sports such as golf, mountain climbing, walking, tennis, and climbing have been in the spotlight. In addition, consumption of home training, an exercise method that maintains health at a low cost and hygienically without being tied to time and space, has increased [2].

Home training was a new emerging stock in the fitness industry because of the high accessibility of content, without incurring extra high costs even before the pandemic of infectious diseases such as COVID-19 [3]. This home training business sector has made a big difference due to the global COVID-19 pandemic. In order to follow non-face-to-face quarantine rules, people were restricted from sports activities enjoyed in their daily lives, and naturally, more and more cases of non-face-to-face home training at home through TV programs and various online content channels [4]. Due to this increase in interest and participation in home training, home training-related industries have also grown. Home training-related content has become easier to access through video sharing sites, Internet TV, and smartphone applications, and sales of various related products such as home training sportswear and body fat scales have increased [5]. In the distribution industry, a home training shop was set up as a health-related product group, or a home training equipment sales category was operated on the online sales page [6].

In this industry trend, research from related academia is also continuously accumulating after social distancing. Looking at the trends in this research trend, it can be divided into first, online content and application-related research for home training, second, research on exercise effects, and third, research on home training participants and athletic leaders. In this situation in the home training business field and academia, the study aims to collect and analyze big data related to home training generated on social media from March 2020 to recently when COVID-19 distancing was strongly implemented to understand consumer perceptions, important keywords, issues, meanings, and clusters. Through this, we would like to provide basic data for home training-related business and research.

2. Research method

2.1 Data collection and analysis method

Keywords for data collection for this study were 'home training' and 'Homt'. Naver and Google were selected as collection channels. Naver Channel collected data from blogs, cafes, news, and Knowledge IN sections, while Google Channel collected data from news sections, and Textom version 6.0, a big data analysis solution, was used to collect data. The collection period was collected on an annual basis for about three years and four months, starting on March 22, 2020, when social distancing was strengthened due to COVID-19. The collected data included 4,000 Naver blogs, 2,673 news, 4,000 cafes, 3,989 Knowledge IN, and 953 Google channel news. Keyword extraction was limited to 70 cases in consideration of the frequency of appearance, and data collection information is shown in <Table 1>.

Table 1. Analyze data information

Item	Content
Collection period	Naver(blogs, cafes, news and Knowledge In) , Google(news)
Collection channel(collection unit)	March 22, 2020 - July 20, 2023(annual basis)
Search word(70item)	Home Training, Homt
Collection tools/ Analysis / Visualization tool	TEXTOM / Ucinet 6.0 NetDraw

2.2. Investigation tools and data processing

In this study, data were collected and analyzed using TEXTOM, a social matrix program, and among the collected text data, words, adverbs, and conjunctions that are not necessary for the research topic were purified, and unrelated contents were deleted. A semantic network analysis was conducted to determine the correlation and association between extracted keywords, and three centralities were derived through the Ucinet 6 program using Netdraw, and the connection between keywords was visualized and clearly expressed. CONCOR analysis was conducted to derive clusters of words with similar characteristics based on semantic network analysis.

3. Results

3.1. Text mining analysis

Text mining is the extraction of meaningful information and knowledge from large-scale text data based on natural language processing technology [7]. In this study, the frequency of words, TF-IDF were analyzed through text mining. Word frequency refers to the frequency of the extracted word, and TF-IDF is the multiplication of the keyword frequency (TF) and the reciprocal of the document frequency (IDF), indicating how important a word is in a specific document.

As a result of text mining analysis, the top 10 keywords based on the frequency of keyword appearance were 'home training' the most at 4,045 times based on the keyword appearance frequency. Next, exercise(3842), Homt(2968), house(1064), apparatus(1041), recommendation(773), diet(671), time(414), home(373) and mat(367) appeared in order. Next, TF-IDF ranks exercise(1698.93), apparatus(1564.03), Homt(1417.07), house(1229.3 8), diet(1201.17), recommendation(1189.07), mat(986.52), set(908.8), eat(895.2) and time (886.59) appeared in order.

Table 1. Text mining analysis results

No	Word	TF	TF-IDF	No	Word	TF	TF-IDF
1	Home Training	4,045	441.78	36	Stretching	209	574.87
2	Exercise	3,842	1698.93	37	Weight	203	551.92
3	Homt	2,968	1417.07	38	Product	188	548.73
4	House	1,064	1229.38	39	Weight Reduction	186	530.54
5	Apparatus	1,041	1564.03	40	Indoor	185	551.64
6	Recommendation	773	1189.07	41	Use	177	502.36
7	Diet	671	1201.17	42	Plank	166	512.02
8	Time	414	886.59	43	Online	164	521.34
9	Home	373	844.68	44	Lower Body	164	483.77
10	Mat	367	986.52	45	Question	162	459.78
11	Effect	348	800.73	46	Push-Up	159	494.79
12	Eat	339	895.20	47	Price	156	435.25
13	Dumbbell	326	882.29	48	Man	154	461.72

14	Fitness Center	325	720.87	49	Shoulder	153	500.65
15	Set	321	908.80	50	Abs	152	481.69
16	Aerobic	311	762.47	51	Whole Body	151	460.41
17	Training	305	690.69	52	Eye Body	151	522.95
18	Squat	305	767.34	53	Height	150	462.67
19	Yoga	304	790.72	54	Dinner	147	452.10
20	Band	296	859.48	55	Gym	141	472.54
21	Muscular Strength	296	732.86	56	Pull-Up	140	469.19
22	Muscle	291	712.05	57	Support	137	408.51
23	Method	286	695.75	58	Record	135	420.10
24	Body	275	667.71	59	Fit	134	457.20
25	Sale	268	780.20	60	Morning	134	415.76
26	Woman	266	699.37	61	Machine	134	467.65
27	Video	265	670.80	62	Dumbbell	130	422.59
28	Routine	264	695.63	63	Postural	125	403.71
29	COVID-19	257	620.42	64	Homtmom	123	459.23
30	Health	254	649.62	65	Melkin	122	447.36
31	YouTube	249	612.86	66	Need	116	364.24
32	Wellness	248	632.96	67	Arm	115	390.95
33	Pilates	244	706.66	68	Equipment	112	408.90
34	Day	230	622.47	69	Walking	112	382.14
35	Menu	228	622.76	70	Beginner	112	383.55

3.2. Semantic network analysis and CONCOR analysis

Through text mining, 70 keywords were extracted through TF and TF-IDF results, and 1-mode matrix analysis was performed to convert them into matrices. For these analysis results, semantic network analysis and visualization were performed using the Ucinet6 program. Although keywords according to the frequency of occurrence are important keywords, the frequency of occurrence is low, which may be overestimated or underestimated. To this end, three centrality analyses were conducted. The three centrality analysis analyzed degree centrality, closeness centrality, and betweenness centrality. The specific results of the centrality analysis are shown in Table 2.

Considering the three centrality analysis comprehensively, the top 15 keywords show the same results, and as in TF, the keyword with the highest centrality is 'home training'. Considering the three centralities, important keywords are 'exercise', 'home', 'house', 'apparatus', 'recommendation', 'diet', 'time', 'aerobic', 'musical strength', 'method', 'video', 'YouTube', 'wellness', and 'plank'. Next, to visualize the network, we generated network data with Ucinet 6.0 and visualized it using Netdraw as shown in Figure 1. Ucinet 6.0 and Netdraw are software that visualize and represent networks between words, visually showing the strength of connections between key words and the role of specific words in the network as a whole.

Table 2 Centrality analysis results

No	Word	Degree	Closeness	Betweenness	No	Word	Degree	Closeness	Betweenness
1	Home Training	.086	1	.122	36	Stretching	.006	.986	.083
2	Exercise	.101	1	.122	37	Weight	.006	.958	.073
3	Homt	.064	1	.122	38	Product	.005	.972	.046
4	House	.028	1	.122	39	Weight Reduction	.005	.945	.060
5	Apparatus	.026	1	.122	40	Indoor	.004	.945	.040
6	Recommendation	.023	1	.122	41	Use	.005	.945	.037
7	Diet	.019	1	.122	42	Plank	.004	1	.122
8	Time	.012	1	.122	43	Online	.003	.873	.048
9	Home	.009	.986	.116	44	Lower Body	.005	.986	.083
10	Mat	.008	.958	.071	45	Question	.006	.920	.019
11	Effect	.010	.972	.046	46	Push-Up	.005	.920	.023
12	Eat	.012	.986	.112	47	Price	.004	.932	.070
13	Dumbbell	.011	.972	.046	48	Man	.005	.958	.041

14	Fitness Center	.009	.986	.083	49	Shoulder	.006	.958	.038
15	Set	.010	.972	.075	50	Abs	.005	.958	.065
16	Aerobic	.011	1	.122	51	Whole Body	.005	.972	.072
17	Training	.007	.986	.083	52	Eye Body	.003	.683	.006
18	Squat	.009	.986	.083	53	Height	.005	.986	.110
19	Yoga	.008	.986	.083	54	Dinner	.005	.932	.057
20	Band	.008	.972	.046	55	Gym	.003	.920	.068
21	Muscular Strength	.011	1	.122	56	Pull-Up	.004	.945	.040
22	Muscle	.009	.972	.072	57	Support	.004	.920	.056
23	Method	.008	1	.122	58	Record	.004	.908	.058
24	Body	.008	.986	.118	59	Fit	.003	.972	.072
25	Sale	.005	.896	.06	60	Morning	.005	.945	.096
26	Woman	.007	.972	.079	61	Machine	.004	.932	.068
27	Video	.008	1	.122	62	Dumbbell	.004	.986	.083
28	Routine	.009	.945	.065	63	Postural	.003	.958	.041
29	COVID-19	.006	.986	.083	64	Homtmom	.002	.683	.005
30	Health	.007	.986	.083	65	Melkin	.003	.831	.016
31	YouTube	.007	1	.122	66	Need	.003	.972	.078
32	Wellness	.006	1	.122	67	Arm	.004	.972	.075
33	Pilates	.006	.932	.067	68	Equipment	.002	.852	.046
34	Day	.007	.972	.046	69	Walking	.004	.896	.048
35	Menu	.007	.986	.110	70	Beginner	.004	.958	.038

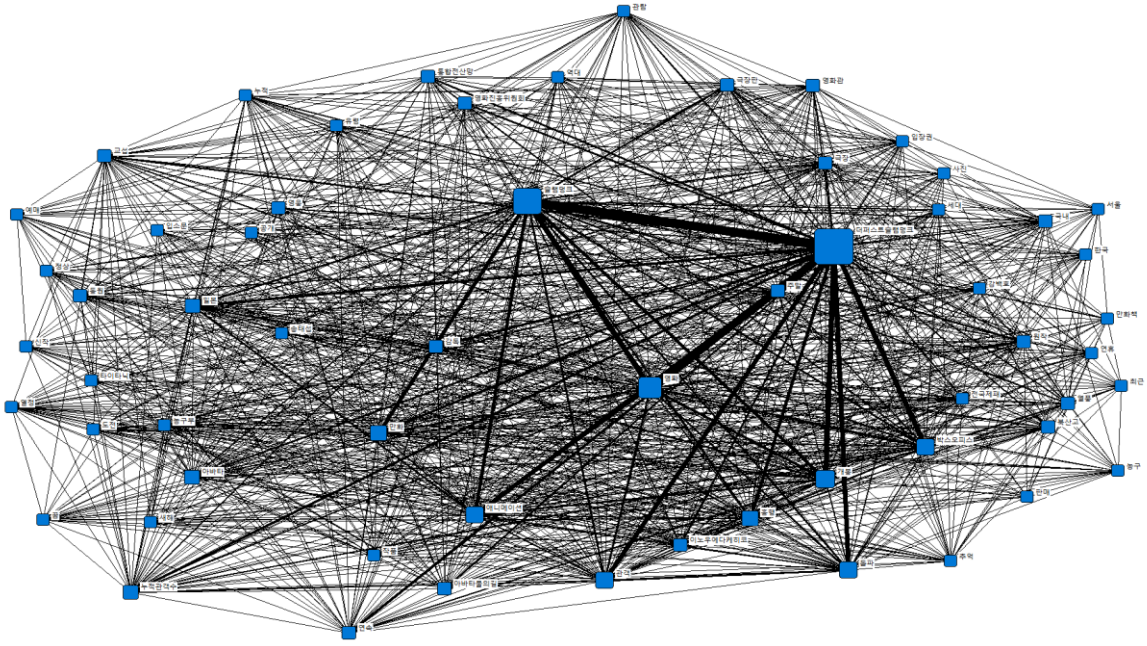

Figure 1. Result of network visualization

Next, the convergence of interrupted correlation of repetitive correlation was performed. CONCOR is an analysis technique that identifies word blocks and identifies relationships between blocks based on Pearson correlations of concurrent appearance matrices between words, and words within a block are considered to have something in common with a particular topic. Therefore, it is a useful analytical technique for understanding the topics discussed across the entire network [8], namely, clustering large networks into multiple small networks to easily grasp the contextual implications of each keyword [9]. In this study, we constructed a semantic network based on the top 70 keywords and determined the appropriate number of clusters by referring to dendrograms as classification criteria in CONCOR analysis.

The determination of the number of clusters was based on the top 70 keywords to establish a semantic network, and the appropriate number of clusters was determined by referring to the dendrogram in CONCOR

analysis [10]. As a result of semantic network analysis and CONCOR analysis, four clusters were formed. The first is the purchase cluster. It is a cluster of words 'sale', 'record', 'machine', 'price', and 'eye body'. The second is the equipment cluster. It has words such as 'gym', 'indoor', 'use', 'homtmom', 'online', and 'pilates'. The third is the diet cluster. It has words such as 'weight', 'height', 'eat', 'menu' and 'weight reduction'. The fourth is the exercise method cluster. It has words such as 'home training', 'exercise', 'Homt', 'effect', 'set', 'aerobic', 'training', 'squat', 'muscular strength', 'method', 'plank', ‘lower body’, 'push-up', abs' and pull-up'.

The categorization and visualization according to the results of CONCOR analysis based on the semantic network are shown in Figure 2 and Table 3 below.

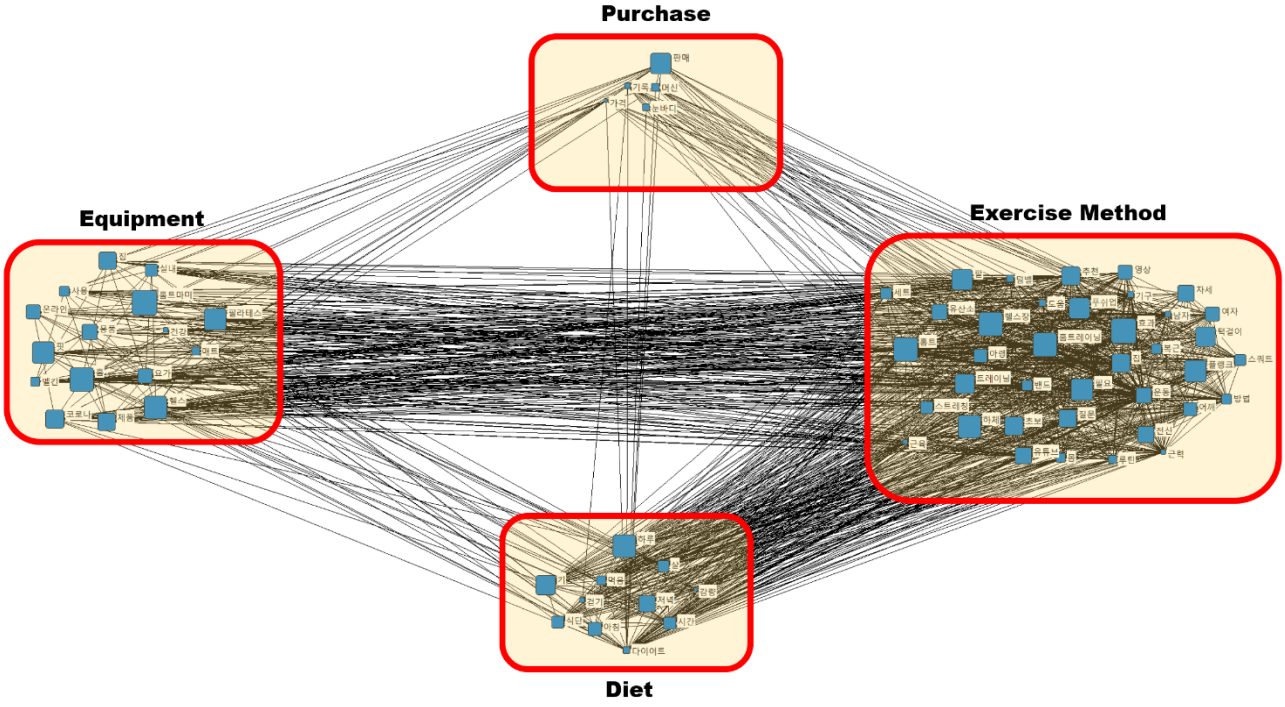

Figure 2. CONCOR analysis results of semantic network

Table 3. Result of categorization

Cluster	Keyword
Purchase	Sale, Record, Machine, Price, Eye Body
Equipment	Gym, Indoor, Use, Homtmom, Online, Pilates, Equipment, Wellness, Fit, Mat, Melkin, Home, Yoga COVID-19, Product, Health
Diet	Day, Weight, Height, Eat, Morning, Dinner, Walking, Menu, Diet, Time, Weight Reduction
Exercise Method	Home Training, Exercise, Homt, House, Apparatus, Recommendation, Effect, Dumbbell, Set Fitness Center, Aerobic, Training, Squat, Band, Muscular Strength, Muscle, Method, Body Woman, Video, Routine, YouTube, Stretching, Plank, Lower Body, Question, Push-Up, Man Shoulder, Abs, Whole Body, Pull-Up, Support, Dumbbell, Postural, Need, Arm, Beginner

4. Conclusion

As a result of text mining analysis, except for the compound word 'home training' and the abbreviation 'Homt', the main emerging keywords are 'exercise', 'apparatus', 'recommendation', 'diet', 'time', 'mat', 'effect', and 'eat'. From a TF-IDF perspective, they are 'exercise', apparatus', 'diet', 'recommendation', 'mat', 'set', 'eat', and 'time'. This means that these keywords are frequently mentioned and have a high proportion in the document. Reflecting this, 70 keywords with high frequency of appearance were extracted and semantic network analysis was performed, and three (Degree, Closeness, and Betweenness) centrality was confirmed. As for the centrality analysis results, major emerging keywords such as TF and TF-IDF survey results show high centrality. Next, the CONCOR analysis result was clustered into four clusters.

First, 'purchse cluster' is related to purchasing supplies or machines necessary for home training. In other words, the main word is about the price and sales of what is needed, and the meaning of whether these products can be improved or confirmed with the eyes is connected. It is judged that home training goods companies should have a process of accurately and appropriately delivering sales and purchase convenience and purchase information on these products.

Second, it is about the necessary products according to the location and exercise type for home training of the 'equipment cluster'. This cluster is categorized with online brands for home training called 'Melkin' and 'Homtmom', which are considered to be mentioned a lot by consumers. Specifically, 'Melkin' is an online shopping mall and offers 'Homtmom' online diet programs. In addition, it is judged that there is a high interest in filates, Yoga products, and meth products. Accordingly, online companies for programs or products for home training will have to provide programs and related products for each type of home training exercise, and establishing a platform that can clearly present the categories of online shopping and provide prompt feedback on related inquiries can improve the company's performance.

Third, 'diet's cluster' represents the purpose that home training consumers want to get the most. Time keywords such as breakfast, dinner, and day, menus, keywords related to diet such as meals, and exercise effects such as diet and weight reduction have been contextualized. In an untact society, exercise space constraints due to concerns about infectious diseases lead to lack of exercise and weight gain. As a result, dieting is believed to be a major concern. Companies that provide home training exercise programs should focus on these consumer needs.

Fourth, keywords for exercise methods such as dumbbells, chin-ups, aerobics, training, squat, muscle strength exercise, and full-body exercise in the 'exercise method cluster' were contextual, and YouTube and videos were found to find these exercise methods. If the previous diet cluster contains a fantasy about weight loss, the 'exercise method cluster' contextualizes various exercise methods and techniques to obtain more diverse part-by-part or full-body exercise effects. Related business industries need to provide motivational content such as various exercise methods, exercise posture, exercise principles, exercise routines, and challenge missions so that they can form a belief that home training can raise health awareness and exercise awareness, which are internal motivations. In addition, it is believed that it will be effective to provide a way to continue until the exercise effect is achieved without giving up on home training. In addition, it is judged that it will be effective to provide a way to continue until the exercise effect is obtained without giving up on home training.

References

[1] M. G. Kim, “Social Changes and Social Policy Paradigm in Times of Covid-19,” Health and welfare policy forum, Vol. 290, pp. 6-19, 2020.

[2] I. J. Hwang, and H. J. Cheon, “People who Exercise Alone at Home: Characteristics by Types and their Continuous

Intention," Journal of Leisure Studies, Vol. 20, No. 3, pp. 55-79, 2022. DOI: https://doi.org/10.22879/slos.2022.20.3.55

[3] N. K. Lee, "A Structural Analysis on the Effects of Consumer Value, Satisfaction, and Purchase Intention on Home Training Consumption of Home Training Consumers," (Unpublished doctoral dissertation). Sungshin Women's University, Seoul, Korea, 2022.

[4] H. M. Jeon, and Y. H. Pan, "A Classification of Type and User Interaction Model of Online Home Training Service in the Social Distancing Environment: Based on the Service Design Perspective," Design Convergence Study, Vol. 19, No. 4, pp. 15-30, 2020. DOI: https://doi.org/10.31678/SDC83.2

[5] M. C. Lee, J. S. Lee, D. I. Han, and K. H. Han, "Changes in the Perception of Home Training Using Text Mining Techniques: Comparison between before and after the COVID-19 Pandemic," Journal of Health Informatics and Statistics, Vol. 48, No. 1, pp. 15-25, 2023. DOI: https://doi.org/10.21032/ jhis.2023.48.1.15

[6] Korea Consumer Agency(2019, February). Home Training Hazard Analysis. Safety Report. pp. 1-17, 2019.

[7] A. Hotho, A. Nurnberger, and G. Paaβ, "A brief survey of text mining," LDV Forum, Vol. 20. No. 1, pp. 19–62. 2005.

[8] R. L. Li, and C. N. Jun, "An Analysis of the Marketing Environment in the Chinese Glasses Market: Focus on Utilizing of Big Data," Global Business Administration Review, Vol. 15, No. 3, pp. 1-24, 2018.

[9] S. B. Lee and M. H. Song, "An Analysis of the Age-Downing Issues in the Election of Public Official Election Act: Based on the Semantic Network Analysis of Newspaper Articles," The Study of Election, Vol. 1, No. 13, pp. 5-35. 2021.

[10] K. W. Byun, "A Study on the Meaning of The First Slam Dunk Based on Text Mining and Semantic Network Analysis," International Journal of Advanced Smart Convergence Vol. 12, No.1, pp. 164-172. 2023. DOI: http://dx.doi.org/10.7236/IJASC.2023.12.1.164