

A Study on the Meaning of The First Slam Dunk Based on Text Mining and Semantic Network Analysis

Kyung-Won Byun *

**Assistant Professor, Department of graduate school of business Administration, Dankook University, Korea
changewon125@ dankook.ac.kr*

Abstract

In this study, we identify the recognition of 'The First Slam Dunk', which is gaining popularity as a sports-based cartoon through big data analysis of social media channels, and provide basic data for the development and development of various contents in the sports industry. Social media channels collected detailed social big data from news provided on Naver and Google sites. Data were collected from January 1, 2023 to February 15, 2023, referring to the release date of 'The First Slam Dunk' in Korea. The collected data were 2,106 Naver news data, and 1,019 Google news data were collected. TF and TF-IDF were analyzed through text mining for these data. Through this, semantic network analysis was conducted for 60 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 the study, the keyword with the high frequency in relation to the subject in consideration of TF and TF-IDF appeared 4,079 times as 'The First Slam Dunk' was the keyword with the high frequency among the frequent keywords. Next are 'Slam Dunk', 'Movie', 'Premiere', 'Animation', 'Audience', and 'Box-Office'. Based on these results, 60 high-frequency appearing keywords were extracted. After that, semantic metrics and centrality analysis were conducted. Finally, a total of 6 clusters(competing movie, cartoon, passion, premiere, attention, Box-Office) were formed through CONCOR analysis. Based on this analysis of the semantic network of 'The First Slam Dunk', basic data on the development plan of sports content were provided.

Keywords: *Big Data Analysis, Slam Dunk, Basketball Cartoon, Animation, Nostalgia*

1. Introduction

The domestic sports industry ecosystem is directly affected by the national crisis of population decline. The number of people continuously seeking exercise for health is increasing, but the trend of the participating population varies depending on sports events. There are also sports that create a sense of crisis about existence.

Skiing, which is typically called the flower of winter sports, peaked at 6.86 million in the 2011-2012

season, but fell to one-third of 2.13 million in 2021-2022 [1]. In the case of viewing sports, there is a worrisome trend. In the case of Korea's most popular professional baseball, the pre-COVID-19 record peaked at 8,713,420 spectators in 2017, falling to about 8.4 million in 2018 and about 7.54 million in 2019.

In addition, in the case of professional football, the number of spectators continued to decline from 2.42 million in 2012, and decreased to 1.57 million in 2018. In 2019, the number of spectators increased to 2,376,923 people, but it is judged to be a temporary effect due to fierce competition in which the winning team is determined in the last game [2]. In fact, professional baseball and professional football have been playing games without spectators and games with limited spectators due to the influence of the COVID-19 pandemic, and it is not easy to increase spectators even after the transition to the Endemic in 2023. This is because the younger generation is avoiding participation in sports due to the growth of the smartphone game market [3].

The aging of the fan base, centering on viewing sports, and the immersion in contents such as games, entertainment, and entertainment of the younger generation act as enormous risks to the growth and development of the overall sports industry. Popular sports content worldwide, such as soccer leagues in Europe and major leagues in the United States, has grown rapidly since the 1980s, but the spread of code-cutting culture has begun to change the order of the broadcasting market [4]. This phenomenon intensified in the 2000s due to the emergence of OTT and new media such as Netflix, Amazon Prime, and YouTube, but it may be a more serious problem for the younger generation to lose interest in sports.

Amid the crisis of the sports industry, the success of 'The First Slam Dunk' at home and abroad has great implications. 'The First Slam Dunk' is a theatrical version of the cartoon 'Slam Dunk', which sold 120 million copies worldwide, creating a basketball boom in the 1990s. It is the first work directed by the original author Takehiko Inoue, and it was released in Japan in December 2022, 26 years after the end of the series, and topped the box office list in Heung Japan, beating 'Avatar 2' [5]. It was also released in Korea on January 4, 2023, and is expected to rank first among Japanese animations released in Korea as of the end of February 2023 [6].

The success of 'The First Slam Dunk' also has great implications for the domestic sports industry ecosystem. For this reason, basketball has been successful in cartoons, but as mentioned earlier, it has made a movie out of cartoons that ended 26 years ago, drawing keen attention from both the older and younger generations of Korea, not from their own countries. It makes the generation who have encountered actual comics feel nostalgia. Generation Z, who have not been exposed to cartoons, has often been exposed to memes and short videos, so they think it is not very unfamiliar content. Accordingly, the researcher intends to analyze the perception of the 'first slam dunk' that is causing syndrome across times, regions, and generations based on big data provided on social media. Through this, we intend to provide implications for domestic sports content.

2. Research method

2.1 Data Collection and Analysis Method

Naver and Google were selected as collection channels for the data collection for this study, and news covering 'The First Slam Dunk' was collected from these channels. Textom version 6.0, a big data analysis solution, was used for data collection. For data collection, 'The First Slam Dunk' was used as the collection keyword. The data collection period considered that the film was released in Korea on January 4, 2023. Therefore, data were collected from January 1, 2023 to February 15, 2023. The data collection unit was set as a 'weekly unit' in consideration of garbage data and data preprocessing. The data collected were 2,106 from Naver News and 1,019 from Google News. Keyword extraction was limited to 60 considering the

frequency of appearance. The analysis data information is shown in Table 1.

Table 1. Analyze data information

Item	Content
Collection channel	Naver news, Google news
Collection period(weekly unit)	January 1, 2023 - February 15, 2023
Search word(60item)	The First Slam Dunk
Collection tools	TEXTOM
Analysis & Visualization tool	UCINET 6.0, NetDraw

2.2. Investigation Tools and Data Processing

In this study, data collection and analysis were conducted using the social matrix program TEXTOM. The data refinement process for the collected data was carried out as shown in <Table 2>. Text mining was performed on the refined data to analyze keyword TF, TF-IDF. In addition, to understand the context of the relevance between the extracted keywords, the semantic connection structure and connection centrality between words were analyzed using UCINET 6, a network software program. Network visualization and CONCOR analysis were performed using Netdraw function.

Table 2. Example of data cleaning

Keyword	Cleaning
The First Slam Dunk, First Slam Dunk	The First Slam Dunk
Avatar 2, Avatar The Way of Water	Avatar 2
Inoue, Inoue Takehiko	Inoue Takehiko
Buksan High School, Buksan	Buksan High School
Meaningless words and numbers, Postpositional particle	Deletion

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.

According to the text mining analysis, 'The First Slam Dunk the most at 4,079 times based on the keyword appearance frequency. Next, 'Slam Dunk(2,433)', 'Movie(1,637)', 'Premiere(1,055)', 'Animation(977)', 'Audience(878)', 'Box-Office(858)', 'Breakthrough(833)', 'Cartoon(644)', 'Hit(615)'. Next, the rankings of TF-IDF are 'The First Slam Dunk(785.83)' and 'Slam Dunk(1657.19)', 'Movie(1350.61)', 'Premiere(1298.19)', 'Animation(1230.70)', 'Audience(1224.87)', 'Box-Office(1345.07)', 'Breakthrough(1197.84)', 'Cartoon(1096.02)', 'Hit(1065.06)'.

Table 1. Text mining analysis results

No	Word	TF	TF-IDF	No	Word	TF	TF-IDF
1	The First Slam Dunk	4079	785.83	31	Preview	211	582.90
2	Slam Dunk	2433	1657.19	32	Generation	195	567.73
3	Movie	1637	1350.61	33	Basketball Team	179	504.28
4	Premiere	1055	1298.19	34	Recent	179	504.28
5	Animation	977	1230.70	35	Ghost	176	522.47
6	Audience	878	1224.87	36	Holiday	175	531.33
7	Box-Office	858	1345.07	37	National Champion	172	490.46
8	Breakthrough	833	1197.84	38	Ticket	169	484.88
9	Cartoon	644	1096.02	39	All Time	165	499.82
10	Hit	615	1065.06	40	Work	164	489.00
11	Japan	558	995.70	41	Open	164	496.79
12	Avatar	439	921.51	42	Titanic	164	522.64
13	Cumulative Audience	421	622.33	43	Accumulate	158	494.77
14	Bargaining	399	891.29	44	Ticketing	157	505.52
15	Continuity	387	888.25	45	Memory	155	473.89
16	Avatar2	367	818.66	46	Seoul	155	486.58
17	Original	326	758.17	47	Challenge	145	438.30
18	Theatrical Version	322	733.74	48	Dream	135	417.65
19	Director	321	762.37	49	Passion	134	415.56
20	Weekend	287	768.01	50	Song Tae-seop	133	441.15
21	Inoue Takehiko	276	664.61	51	New Year	130	416.42
22	Theater	274	673.42	52	Comic Book	130	438.63
23	Domestic	263	659.06	53	Picture	128	410.01
24	Integrated Computer Network	252	622.33	54	Sale	126	428.89
25	Cinema	251	633.16	55	Basketball	126	420.28
26	Buksan High School	244	617.55	56	Korea	125	408.95
27	Hero	243	661.18	57	Top	125	408.95
28	Mobilization	242	621.86	58	Kang Baek-ho	120	396.92
29	Film Promotion Committee	237	599.84	59	New Work	116	381.58
30	Craze	225	602.10	60	Word of Mouth	116	380.54

3.2. Semantic Network Analysis and CONCOR Analysis

Semantic network analysis is a method of applying social network analysis that values the relationship between actors to communication messages when grasping the system structure of society [8]. To do this, 60 keywords were extracted through TF and TF-IDF results through text mining, 1-mode matrix analysis was performed, and converted into a matrix. In response, semantic network analysis and visualization were performed using UCINET 6.0, NetDraw program.

Although the keyword according to the occurrence frequency is an important keyword, the occurrence frequency is low, and thus may be overexpressed or underestimated. To this end, three centrality analyses were conducted. The three centrality analyses are degree centrality, closeness centrality and betweenness centrality were analyzed. The specific results of centrality analysis are shown in Table 2.

In the three connection centrality analysis, the top 10 keywords with high degree of centrality show the same results. As in TF, the highest centrality keyword is 'The First Slam Dunk', followed by 'Slam Dunk', 'Movie', 'Premiere', 'Animation', 'Audience', 'Box-Office' 'Breakthrough', 'Cartoon', and 'Hit'. Next, network data was generated with UCINET to visualize the network, and it was visualized as shown in Figure 1 using NetDraw. UCINET and NetDraw are software that visualizes and expresses the network between words and visually shows the strength of the connection between key words and the role of certain words in the entire network.

Table 2 Centrality analysis results

No	Word	Degree	Closeness	Betweenness	No	Word	Degree	Closeness	Betweenness
1	The First Slam Dunk	0.170	1	0.057	31	Preview	0.010	0.983	0.041
2	Slam Dunk	0.107	1	0.057	32	Generation	0.009	1	0.057
3	Movie	0.078	1	0.057	33	Basketball Team	0.012	0.983	0.048
4	Premiere	0.054	1	0.057	34	Recent	0.008	0.922	0.038
5	Animation	0.051	1	0.057	35	Ghost	0.009	0.983	0.052
6	Audience	0.045	1	0.057	36	Holidays	0.008	0.967	0.052
7	Box-Office	0.050	1	0.057	37	National Champion	0.012	1	0.057
8	Breakthrough	0.042	1	0.057	38	Ticket	0.012	0.922	0.024
9	Cartoon	0.035	1	0.057	39	All Time	0.010	0.967	0.037
10	Hit	0.031	1	0.057	40	Work	0.008	1	0.057
11	Japan	0.031	1	0.057	41	Open	0.004	0.983	0.041
12	Avatar	0.023	1	0.057	42	Titanic	0.008	0.952	0.031
13	Cumulative Audience	0.023	1	0.057	43	Accumulate	0.008	0.967	0.050
14	Bargaining	0.020	1	0.057	44	Ticketing	0.007	0.952	0.031
15	Continuity	0.023	1	0.057	45	Memory	0.008	0.983	0.054
16	Avatar2	0.017	0.983	0.052	46	Seoul	0.006	0.894	0.034
17	Original	0.017	1	0.057	47	Challenge	0.010	0.967	0.045
18	Theatrical Version	0.017	1	0.057	48	Dream	0.010	0.922	0.031
19	Director	0.018	1	0.057	49	Passion	0.010	0.967	0.045

20	Weekend	0.018	1	0.057	50	Song Tae-seop	0.006	0.983	0.054
21	Inoue Takehiko	0.016	1	0.057	51	New Year	0.007	0.952	0.039
22	Theater	0.014	1	0.057	52	Comic Book	0.006	0.937	0.044
23	Domestic	0.014	1	0.057	53	Picture	0.006	0.952	0.046
24	Integrated Computer Network	0.017	0.967	0.035	54	Sale	0.006	0.831	0.020
25	Cinema	0.015	0.952	0.043	55	Basketball	0.006	0.937	0.041
26	Buksan High School	0.015	1	0.057	56	Korea	0.005	1	0.057
27	Hero	0.011	0.967	0.037	57	Top	0.008	0.937	0.028
28	Mobilization	0.015	0.937	0.024	58	Kang Baek-ho	0.005	0.937	0.044
29	Film Promotion Committee	0.016	0.983	0.041	59	New Work	0.006	0.967	0.051
30	Craze	0.011	0.983	0.054	60	Word of Mouth	0.005	0.937	0.029

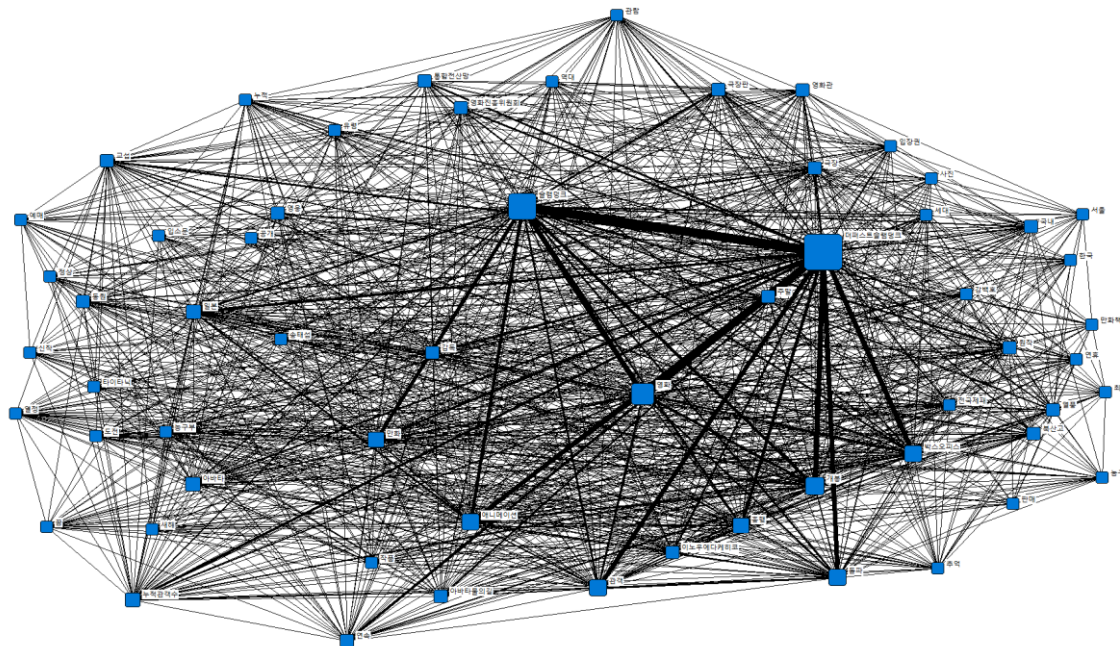


Figure 1. Result of network visualization

Next, the Convergence of iterated correlation(CONCOR) analysis was conducted. CONCOR is an analytical technique that identifies blocks of words based on the Pearson correlation of simultaneous emergence matrices between words and grasps the relationship between blocks. Words in a block are considered to have something in common with a particular subject, and jointly constitute an event or subject. Therefore, it is a useful analytical technique to understand what topics are discussed in the whole network [9]. In other words, the purpose is to easily grasp the contextual meaning constituted by each keyword by clustering large networks into several small networks [10]. In this study, a semantic network was established based on the top 60 keywords, and the appropriate number of clusters was determined by referring to the dendrogram as the classification criterion in the CONCOR analysis.

As a result of semantic network analysis and CONCOR analysis, six clusters were formed. The first is a cluster of words ‘Competing Movie’, ‘Avatar’, ‘Avatar 2’, ‘Hero’, ‘Ghost’, ‘Holiday’, and ‘New Work’. The second is a cluster called ‘Cartoon’ with words such as ‘Slam Dunk’, ‘Cartoon’, ‘Original’, ‘Thermal

Version’, ‘Crazy’, ‘Generation’, ‘Etc’. The third is a cluster called ‘Passion’ with words such as ‘Basketball Team’, ‘Challenge’, ‘Dream’, ‘Passion’. The fourth is a cluster called ‘Premiere’, which is a group of words such as ‘Premiere’, ‘Animation’, ‘Japan’, ‘Theater’, ‘Domestic’, ‘Recent’, ‘Etc’. The fifth is the cluster called ‘Attention’, which is a group of words such as ‘The First Slam Dunk’, ‘Movie’, ‘Breakthrough’, ‘Hit’, ‘Cumulative Audience’, ‘Director’, ‘Etc’. The sixth is a cluster called ‘Box-Office’ with words such as ‘Audience’, ‘Box-Office’, ‘Continuity’, ‘Weekend’, ‘Integrated Computer Network’, ‘Cinema’, ‘Etc’.

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.

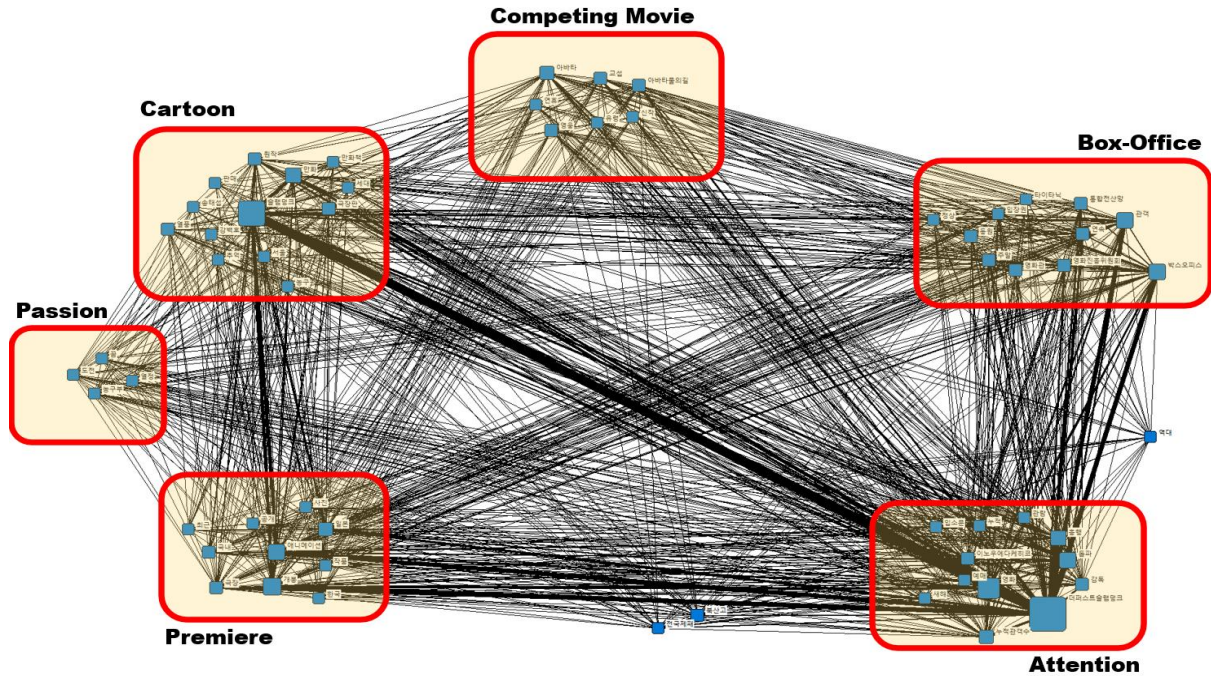


Figure 2. CONCOR analysis results of semantic network

Table 4. Result of categorization

Cluster	Keyword
Competing Movie	Avatar, Bargaining, Avatar2, Hero, Ghost, Holiday, New Work
Cartoon	Slam Dunk, Cartoon, Original, Theatrical Version, Craze, Generation, Memory, Seoul, Song Tae-seop, Comic Book, Sale, Basketball, Kang Baek-ho
Passion	Basketball Team, Challenge, Dream, Passion
Premiere	Premiere, Animation, Japan, Theater, Domestic, Recent, Work, Open, Picture, Korea
Attention	The First Slam Dunk, Movie, Breakthrough, Hit, Cumulative Audience, Director, Inoue Takehiko, Preview, Accumulate, Ticketing, New Year, Word of Mouth
Box-Office	Audience, Box-Office, Continuity, Weekend, Integrated Computer Network, Cinema, Mobilization, Film Promotion Committee, Ticket, Titanic, Top

4. Conclusion

The big data analysis process was analyzed by TF and TF-IDF through text mining. Keywords such as ‘The First Slam Dunk’, ‘Movie’, ‘Premiere’, ‘Animation’, ‘Audience’, ‘Box-Office’, ‘Breakthrough’, ‘Cartoon’, and ‘Hit’ were frequently mentioned. Through this, it can be seen that there is a high interest in the release of the ‘Slam Dunk’ cartoon as an animation. Sixty keywords of high frequency appearance were

extracted and semantic network analysis was performed. Accordingly, the analysis of degree centrality, closeness centrality, and betweenness centrality was confirmed. In the three centrality analyses, 'The First Slam Dunk', 'Slam Dunk', 'Movie', 'Premiere', 'Animation', 'Audience', and 'Box-Office', which show high frequency in text mining, show high centrality

Next, the CONCOR analysis results were clustered into six clusters. First, 'competing movie' cluster is about competing movies with 'The First Slam Dunk'. Avatar 2 is the most powerful film in the competition, and works such as Bargaining, Hero, and Ghost have been mentioned. The syndrome of 'The First Slam Dunk' is believed to be a solid story and video technology that can appeal to all generations. Second, the 'cartoon' cluster is related to the original cartoon, and keywords such as the main characters, original, generation, and memory were clustered. I think it is a success factor that reminded me of the emotion I felt in the cartoon while watching the movie. Third, the main story line of the 'passion' cluster is the process of challenging five basketball players from Buksan High School with the dream of winning the national championship. Although Buksan High School is considered an underdog, famous lines such as "If you give up, that moment is the end of the game" and "Your left hand only helps" represent the passion of Buksan High School basketball team and the main character. In this context, movies appeal to the younger generation while providing a message of courage to young people who wander and are in need in modern society. Fourth, the 'premiere' cluster has a cluster of keywords about movie release. Since it was released in Korea about a month after it was released in Japan first, it shows the expectations of slam dunk fans waiting for its release in Korea. Fifth, the 'attention' cluster means the high interest of fans and audiences in the movie. Including director Inoue Takehiko, he is showing interest in ticketing, cumulative audience numbers, and word of mouth. In the case of clubs that are popular in professional teams, not only performance and players, but also the popularity of leaders will affect. As a result, all members of the team are required to communicate and empathize with fans. Sixth, it reflects the high popularity of the 'box office' cluster 'The First Slam Dunk'. The number of 'Nth time watching' movies that have been watched several times is increasing. They are also watching in Korean and Japanese versions, respectively, and more and more people are waiting to cheer and watch IMAX movies. The success factor that shows this syndrome phenomenon, not only those who have directly encountered the original cartoon but also those who have not experienced the original cartoon, is believed to be due to the director's ability and authenticity story.

Acknowledgement

'The present research was supported by the research fund of Dankook University in 2022'

References

- [1] J. h. Kim, "Value Analysis of Skiwear Brand and Positioning Strategy based on Image and Satisfaction," (Unpublished master's dissertation). Dankook University, Cheonan, Korea, 2016.
- [2] Ministry of Culture, Sports and Tourism, *Sports Industry Research 2020*, pp. 96-100, 2022.
- [3] S. S. Shim, "Structural Relationship between Benefit of Ski Wear Brand, Brand Emotion, Brand Satisfaction, Brand Trust, and Repurchase Intention," *International Journal of Advanced Smart Convergence*, Vol. 11, No. 4, PP. 177-184, 2022. DOI.org/10.7236/IJASC.2022.11.4.177
- [4] J. S. Lee. "MLB's Insight Predicts Code Cutting for Broadcasting Rights," *DBR*, No. 256, 2018.
- [5] W. J. Na. (2023, January 3). "Slam Dunk Starring Song Tae-Sup, Ripped out of Comic Book," *The*

JoongAng, Retrieved February 2, 2023, from <https://www.joongang.co.kr/article/25130747>.

- [6] B. L. Kim. (2023, February 28). “‘Slam Dunk’ Surpasses No. 1 in Japanese Animation... ‘Your name’ 190,000 difference,” *OSEN*, Retrieved February 28, 2023, from <https://www.joongang.co.kr/article/25130747>.
- [7] A. Hotho, A. Nurnberger, and G. Paaß, “A Brief Survey of Text Mining,” *LDV Forum*, Vol. 20. No. 1, pp. 19–62, 2005.
- [8] H. S. Kim, “An Exploratory Study on the Semantic Network Analysis of Food Tourism through the Big Data,” *Culinary Science & Hospitality Research*, Vol. 23, No. 4, pp. 22-32, 2018.
DOI: [org/10.20878/cshr.2017.23.4.003](https://doi.org/10.20878/cshr.2017.23.4.003)
- [9] 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.
- [10] 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.