

Big Data Analysis of the Women Who Score Goal Sports Entertainment Program: Focusing on Text Mining and Semantic Network Analysis.

Hyun-Myung Kim¹, Kyung-Won Byun²

¹Ph D. Student of Physical Education, Dankook University, Korea
gianna85@naver.com

²Assistant Professor, Department of graduate school of business Administration, Dankook University, Korea
changewon125@dankook.ac.kr

Abstract

The purpose of this study is to provide basic data on sports entertainment programs by collecting data on unstructured data generated by Naver and Google for SBS entertainment program 'Women Who Score Goal', which began regular broadcast in June 2021, and analyzing public perceptions through data mining, semantic matrix, and CONCOR analysis. Data collection was conducted using Textom, and 27,911 cases of data accumulated for 16 months from June 16, 2021 to October 15, 2022. For the collected data, 80 key keywords related to 'Kick a Goal' were derived through simple frequency and TF-IDF analysis through data mining. Semantic network analysis was conducted to analyze the relationship between the top 80 keywords analyzed through this process. The centrality was derived through the UCINET 6.0 program using NetDraw of UCINET 6.0, understanding the characteristics of the network, and visualizing the connection relationship between keywords to express it clearly. CONCOR analysis was conducted to derive a cluster of words with similar characteristics based on the semantic network. As a result of the analysis, it was analyzed as a 'program' cluster related to the broadcast content of 'Kick a Goal' and a 'Soccer' cluster, a sports event of 'Kick a Goal'. In addition to the scenes about the game of the cast, it was analyzed as an 'Everyday Life' cluster about training and daily life, and a cluster about 'Broadcast Manipulation' that disappointed viewers with manipulation of the game content.

Keywords: Big data, Women Who Score Goal, Kick a Goal, Text mining, Semantic Network, CONCOR Analysis

1. Introduction

Recently, various sports entertainment programs have been aired with many sports events as content. Various sports such as Korean wrestling, badminton, curling, and table tennis are being produced as entertainment programs, as well as popular domestic sports such as soccer, basketball, golf, and baseball. In the keyword of sports, it fills various stories centering on sports such as past sports stars, sports star's family, and sports competition of entertainers, and plans and produces sports entertainment programs to give viewers

Manuscript Received: January. 22, 2023 / Revised: January. 24, 2023 / Accepted: January. 25, 2023

Corresponding Author: changewon125@dankook.ac.kr

Tel: +82-31-8005-3444

Assistant Professor, Department of Graduate School of Business Administration, Dankook University, Korea

pleasure [1]. Among various sports entertainment programs, 'Women Who Score Goals (hereafter referred to as Kick a Goal)' is a South Korean futsal competition that has been broadcast on SBS TV since June 16, 2021. It has been regularly scheduled since June 16, 2021, receiving favorable responses from countless viewers [2]. As the broadcaster announced, it is a 200% authentic program that women who are sincere in soccer make with Korean legend soccer stars [3]

In 2021, pilot broadcasting began as a special program for lunar new year. With high ratings and interest, it was immediately regularized. Regular programming aired from June 16, 2021 to September 14, 2022. Six teams participated in the regular broadcasting from four pilot teams. Immediately after the regular broadcast, it was decided on a season basis, and 10 teams participated in season 2. As of 2023, in season 3, 12 teams are operating leagues divided into upper and lower leagues. The main success factor of such a 'Kick a Goal' program is a case in which the direction of content has been positively changed and established by female performers, not program writers or directors. Even though the directors manipulated the game process with editing and replaced it, 'Kick a Goal' is still airing with the passion and sincerity of the cast [4].

Understanding the public's perception and interest in the 'Kick a Goal' program and what the context of this interest is believed to be able to inform the continuous success strategy of the 'Kick a Goal' program. In addition, it is believed that meaningful information can be provided to entertainment programs produced based on sports contents in the future.

Previously, qualitative surveys were mainly conducted through quantitative surveys using questionnaires or in-depth interviews with subjects to identify factors of perception and interest of viewers and the public. However, this study aims to utilize big data that is widely used in academic and practical fields. In other words, we will provide 'Kick a Goal' program through big data analysis on unstructured data such as portal articles, blogs, and cafe postings.

Big data is a term that means vast amounts of data, and it is from the 2012 World Economic Forum report that it was officially used. Big data is a large amount of data that is typically structured or unstructured, which means data that is difficult to process with typical DB, S/W technologies [5]. Due to the development of information and communication technology, the importance of unstructured data analysis such as text, videos, and images is increasing as various structured, semi-structured, and unstructured data such as search records, consumer access locations and logs are accumulated and SNS such as Instagram, Facebook, and Twitter are activated [6].

As a result, interest in the value of big data and various ways to utilize it is increasing worldwide, and various attempts and studies to explore and analyze big data are increasing [7]. Big data analysis methods for these unstructured data include text mining analysis, opinion mining analysis, semantic network analysis, topic modeling, and sentiment analysis [8]. The purpose of this study is to understand the public's perception and interest in the 'Kick a Goal' program using text mining and semantic network analysis. Through this analysis, it is intended to provide information on production or broadcasting of sports entertainment programs.

2. Research method

2.1. Data collection and analysis method

In order to collect big data on the 'Kick a Goal' program, data on Naver and Google portal sites including related keywords were collected using Textom, a big data solution service. 27,911 cases of data accumulated over the 16 months from June 16, 2021 to October 15, 2022, when season 1 of 'Kick a Goal' In order to extract information from unstructured texts, data were collected using text toms, and unnecessary words, adverbs, conjunctions, and articles that were not related to the research topic were refined using Note ++. In addition, in order to find meaningful keywords, data mining was conducted to analyze the frequency of keywords that can

identify how many texts are mentioned, and TF-IDF analysis, which grasped the importance of keywords in consideration of the frequency of specific words (TF) in documents and the frequency of specific words (DF) in various documents, 80 key keywords related to ‘Kick a goal’ were derived. In this process, the cast appearing in the program was excluded. Semantic network analysis was conducted to analyze the relationship between the top 90 keywords analyzed and connected. NetDraw of UCINET 6.0 was used to derive degree centrality through the UCINET 6.0 program to understand the characteristics of the network and to visualize and clearly express the connection relationship between keywords. CONCOR analysis was conducted to derive a cluster of words with similar characteristics based on the semantic network.

2.2. Analysis process

The big data analysis process for ‘Kick a goal’ first collects data using Textom and extracts appropriate keywords through data refinement. After that, simple frequency analysis (TF) and TD-IDF are analyzed to confirm the high frequency of appearance. Next, connectivity centrality is confirmed using the UCINET 6.0 program, and semantic network analysis is conducted using UCINET 6.0 and NetDraw. CONCOR analysis is conducted for categorization and visualization of key keywords. This process is shown in Figure 1 below.

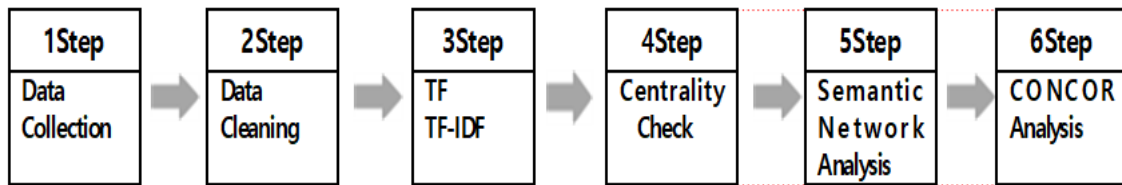


Figure 1. Analysis process

3. Results

3.1. Top keyword frequency analysis TF-IDF analysis through text mining

Text mining is the extraction of meaningful information and knowledge from large-scale text data based on natural language processing technology [5]. 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, ‘Kick a Goal’ appeared the most at 25,888 times based on the keyword appearance frequency. Next, ‘Entertainment Program(13,887)’, ‘Broadcasting(9,905)’, ‘Appear (7,253)’, ‘Soccer(6,896)’, ‘Season(4,641)’, ‘Game(4,321)’, ‘Team (3,424)’, ‘FC GUCHEOK(3,293)’. Next, the rankings of TF-IDF are ‘Kick a Goal(16,772.41)’ and ‘Entertainment Programs(11,796.81)’ ‘Broadcasting(11,794.98)’, ‘Soccer(11,566.9)’, ‘Appear(10,452.08)’, ‘Season(9,777.37)’, ‘Game(8,900.00)’, ‘FC GUCHEOK(8,233.034)’, ‘Team(7,834.84)’.

Table 1. Text mining analysis results

NO	Keyword	TF	TF-IDF	NO	Keyword	TF	TF-IDF
1	Kick a Goal	25888	16772.41435	41	Victory	900	3181.014387
2	Entertainment Program	13887	11796.81047	42	Goalkeeper	897	3160.275305
3	Broadcasting	9905	11794.98244	43	Legend	886	3069.956895
4	Appear	7253	10452.08007	44	Controversy	843	3239.270809
5	Soccer	6896	11566.89654	45	Celebrity	819	2898.869937
6	Season	4641	9777.371975	46	Announcer	817	3081.127731
7	Game	4321	8900.002496	47	Footballer	803	2892.704424
8	Team	3424	7834.838377	48	Progress	802	2871.946318
9	FC GUCHEOK	3293	8233.034194	49	Singer	791	2942.375161
10	Activities	3273	7064.176256	50	Viewer	771	2775.347558
11	FC BULNABANG	2629	7065.9734	51	Ability	763	2787.746342
12	FC WONDER WOMAN	2605	6962.273787	52	Advance	739	2801.295439
13	FC TOPGIRL	2590	7100.875456	53	Profile	715	2728.033214
14	FC GAVENGERS	2458	6820.370993	54	Editing	710	2834.615162
15	Director	2352	6483.56876	55	Challenge League	709	2768.342271
16	Member	2285	6055.21245	56	League	702	2689.078138
17	FC ANACONDA	2271	6669.071808	57	Championship	666	2605.333726
18	FC ACTIONISTAR	1840	5674.079264	58	Top	655	2556.29898
19	Model	1809	5306.081414	59	Ace	646	2523.536239
20	Player	1789	5046.221499	60	Production Crew	646	2610.812444
21	Wednesday	1754	4968.229022	61	Tears	645	2513.750123
22	FC NATIONAL TEAM FAMILY	1743	5413.234681	62	Futsal	620	2610.252861
23	Watching	1556	4959.042641	63	Interest	618	2378.958835
24	Actor	1525	4667.531019	64	Fan	605	2356.761634
25	Super League	1455	4659.385196	65	Ranking	592	2386.398003
26	Issue	1373	4291.495447	66	Stopover	592	2462.298713
27	Instagram	1342	4085.452537	67	All-Star Game	584	2438.746391
28	FC WORLD CLASS	1302	4359.562312	68	Filming	579	2323.282485
29	Manipulation	1286	4586.564037	69	Charm	572	2245.035742
30	Love	1207	4225.63185	70	Information	569	2269.41761
31	Sports	1193	3949.196594	71	Rerun	565	2547.34671
32	Women's football	1034	3693.20376	72	Beautiful	559	2265.091946
33	Exercise	985	3448.423672	73	Younger	556	2236.684138
34	Injury	982	3475.812305	74	New team	555	2254.794373
35	Result	970	3404.235837	75	Proud	543	2148.715041
36	COVID-19	961	3489.545206	76	Entertainer	539	2144.433262
37	Fun	949	3289.367986	77	Passion	536	2136.734817
38	Star	944	3246.774574	78	Video	533	2146.357376

39	Battle	929	3232.174711	79	Cast	533	2181.638128
40	Popularity	908	3172.2237	80	Captain	527	2126.569223

3.2. Semantic network analysis and CONCOR analysis

In the semantic network analysis, the degree centrality of the keywords was confirmed and CONCOR analysis for clustering was conducted. Connection centrality refers to the degree to which nodes words are connected to other nodes on the network. The more connected nodes, the higher the connection centrality. The means that it has a great influence on the network because it affects other words [9]. The results of the connection centrality analysis are shown in Table 3 below.

The results of connection centrality analysis are shown in Table 3. In simple frequency analysis and TF-IDF analysis, 'Kick a Goal', 'Entertainment Program', 'Broadcasting', 'Soccer', 'Season', 'Game', 'Appear', and 'FC GUCHEOK', which were also high in frequency analysis, are also found to play a significant role in connecting with other keywords to form meaning and context. Considering connection centrality. The network is shown in figure 2.

Table 2 Centrality analysis results

NO	Keyword	Degree	NO	Keyword	Degree	NO	Keyword	Degree	NO	Keyword	Degree
1	Kick a Goal	0.126	21	Wednesday	0.011	41	Victory	0.006	61	Tears	0.004
2	Entertainment program	0.058	22	FC NATIONAL TEAM FAMILY	0.011	42	Goalkeeper	0.004	62	Futsal	0.002
3	Broadcasting	0.056	23	Watching	0.009	43	Legend	0.005	63	Interest	0.003
4	Appear	0.023	24	Actor	0.006	44	Controversy	0.005	64	Fan	0.002
5	Soccer	0.030	25	Super League	0.01	45	Celebrity	0.003	65	Ranking	0.003
6	Season	0.027	26	Issue	0.006	46	Announcer	0.004	66	Stopover	0.003
7	Game	0.027	27	Instagram	0.005	47	Footballer	0.003	67	All-Star Game	0.004
8	Team	0.021	28	FC WORLD CLASS	0.007	48	Progress	0.004	68	Filming	0.003
9	FC GUCHEOK	0.019	29	Manipulation	0.008	49	Singer	0.003	69	Charm	0.002
10	Activities	0.013	30	Love	0.004	50	Viewers	0.003	70	Information	0.002
11	FC BULNABANG	0.014	31	Sports	0.005	51	Ability	0.004	71	Rerun	0.002
12	FC WONDER WOMAN	0.015	32	Women's Football	0.004	52	Advance	0.005	72	Beautiful	0.002
13	FC TOP GIRL	0.015	33	Exercise	0.003	53	Profile	0.003	73	Younger	0.002
14	FC GAVENGERS	0.014	34	Injury	0.005	54	Editing	0.005	74	New Team	0.003
15	Director	0.013	35	Result	0.006	55	Challenge League	0.004	75	Proud	0.002
16	Member	0.012	36	COVID-19	0.004	56	League	0.004	76	Entertainer	0.002

17	FC ANACONDA	0.014	37	Fun	0.004	57	Championship	0.004	77	Passion	0.003
18	FC ACTIONISTAR	0.011	38	Star	0.004	58	Top	0.004	78	Video	0.002
19	Model	0.007	39	Battle	0.007	59	Ace	0.004	79	Cast	0.003
20	Player	0.009	40	Popularity	0.004	60	Production Crew	0.004	80	Captain	0.003

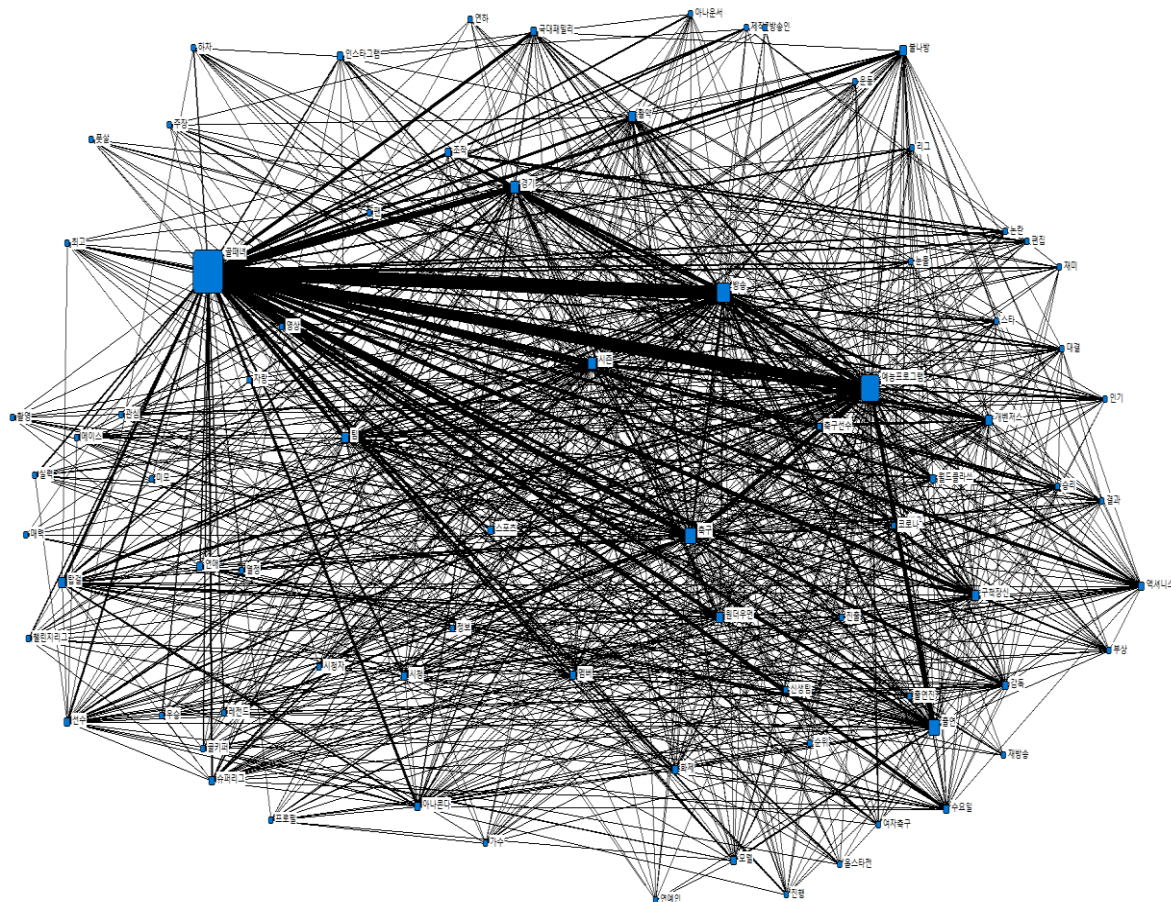


Figure 2. Result of network visualization

Next, convergent correlation analysis (CONCOR) was conducted to identify specific issues by classifying keywords with similar appearance patterns through cluster analysis of semantic network. CONCOR analysis is a type of structural equivalence analysis that measures through the probability that a particular word shares a third word by identifying whether each keyword has a similar type of relationship to each other. The purpose of structural equivalence analysis is to easily understand the contextual meaning of each keyword by clustering large-scale networks into several small-scale networks based on the correlation analysis between each keyword [10].

CONCOR analysis was conducted after constructing semantic network focusing on the top 80 keywords, and it is necessary to reduce each keyword to a small number of clusters through an appropriate number of classification criteria. At this time, the criteria for determining the number of clusters were determined in four clusters by referring to the dendrogram, which is a data that expresses the process of forming clusters by each

keyword in a tree-type graph. As a result of semantic network analysis and CONCOR analysis, four groups were formed. First, the first cluster was named 'Program' as a group of words such as 'Kick a Goal', 'Broadcasting', 'Season', 'Captain', and 'FC GUCHEOK' and 'FC BULNABANG'. Second, the cluster was named 'Soccer' as the words of 'Entertainment Programs', 'Soccer', 'Popularity', 'Footballer', 'Futsal', and 'Women's Football'. Third, the cluster was named 'Everyday Life' with words such as 'Appear', 'Activities', 'Love', 'Celebrity', 'Beautiful Face', 'Proud', 'Younger', 'Instagram', and 'COVID-19'. Fourth, the cluster is called 'Broadcast Manipulation' with words such as 'Manipulation', 'Controversy', 'Production Crew' and 'Editing'.

The categorization and visualization according to the results of CONCOR analysis based on the semantic network are shown in Figure 3 and Table 4 below.

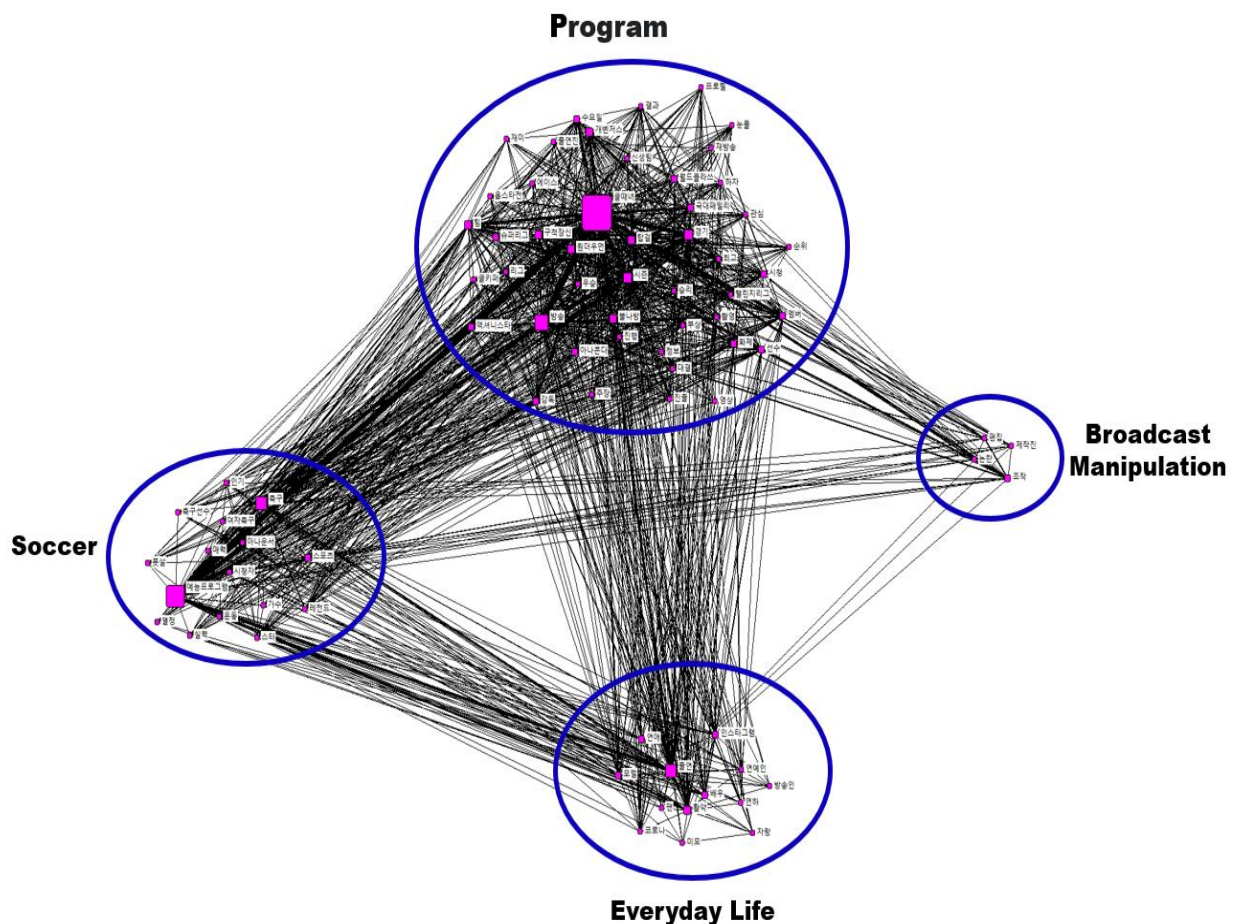


Figure 3. CONCOR analysis results of semantic network

Table 4. Result of categorization

Cluster	Keyword
Program	Kick a Goal, Broadcasting, Season, Captain, FC GUCHEOK, FC BULNABANG, FC WONDER WOMAN, FC TOP GIRL, FC GAVENERS, FC ANACONDA, FC ACTIONISTAR, FC NATIONAL TEAM FAMILY, FC WORLD, CLASS, Director, Member, Player, Wednesday, Watching, Super League, Issue, Injury, Result, Fun, Battle, Victory, Goalkeeper, Progress, advance, Profile Challenge League, League, Championship, Top, Ace, Tears, Interest, Ranking, Stopover, All-Star Game, Filming, Information, Rerun, Newt Team, Video, Cast
Soccer	Entertainment programs, Soccer, Popularity, Footballer, Announcer, Ability, Passion, Futsal, Sports, Women's Football, Exercise, Singer, Legend, Charm, Viewers, Star
Everyday Life	Appear, Activities, Love, Celebrity, Beautiful Face, Proud, Younger, Instagram, COVID-19, Fan Actor, Entertainer, Model
Broadcast Manipulation	Manipulation, Controversy, Production Crew, Editing

4. Conclusion

The main purpose of the study is to analyze the perception of viewers and the public by collecting and analyzing unstructured data on 'goal girls', which are established as a season despite unfavorable factors such as broadcasting manipulation, amid a surge in sports entertainment programs. Data was collected based on unstructured data on the Internet portal, and the collection period is from June 16, 2021, the start date of regular broadcasting, to October 15, 2022. Using Textom, 27,911 data were collected and text mining and CONCOR analysis were performed. The conclusions drawn based on the analysis results are as follows.

First, it is the 'goal 'Kick a Goal ' which shows the highest frequency among the collected data. Next, 'entertainment program', 'broadcasting', 'appearance', 'soccer', 'season', 'game', 'team'. In the next TF-IDF, 'Kick a Goal' was the highest, followed by 'entertainment program', 'broadcasting', 'soccer', 'appearance', 'season', 'game', 'FC GUCHEOK'. Regarding the participating teams, 'FC GUCHEOK' was mentioned the most, which was weak in the first appearance of 'FC GUCHEOK', but the players gradually improved their skills by continuing their soccer practice in practice, and the impression of victory and defeat gave viewers a lot of sympathy. As a result, sports-related entertainment programs will be an important point in how to show real scenes to viewers, not screenplays, as sports are 'drama without script'.

Second, it can be divided into four clusters through semantic network analysis and CONCOR analysis. The four clusters are 'program', 'soccer', 'everyday life', and 'broadcast manipulation'. The clustered 'program' is related to the content of a sports entertainment program called 'Kick a Goal'. Keywords related to all participating teams, game management leagues, broadcasting, issues, injuries and fun in the game progress, progress, and watching were categorized. The pros and cons of 'Kick a Goal' are clear. 'Soccer', which is analyzed as a cluster, contains contents about soccer games as a sport. The essence of 'Kick a Goal' is soccer, and future season broadcasts should be produced so as not to make mistakes that overlook this essence. The popularity of 'Kick a Goal' is also interested in the cast and director 'everyday life'. The stories about family, hobbies, and jobs that are projected in the daily life of preparing for the contest of the cast will also be of interest to viewers. 'broadcast manipulation' broadcasted 'Kick a Goal' on December 22, 2021 by changing the scoring order to add dramatic fun. Viewers were angry about this, and there was a case in which PD and CP resigned. It was a complacent production behavior of the production crew that was completely different from the authenticity of the program. It can be seen that authenticity, which is real rather than fake, plays a very important role in entertainment programs, not real sports. Fakes do not last long. Sports is also called a

drama without a script, and authenticity should be the philosophy of the program in the entertainment program based on such sports.

References

- [1] J. Y. Seok and J. M. Lee, ‘Study on trend analysis of sport entertainment contents applying big data algorithm’, *The Korean Journal of Sport*, Vol.20 No.3, pp. 1 ~ 12, 2022. DOI: <https://doi.org/10.46669/kss.2022.20.3.001>
- [2] Wikipedia, ‘Kick a Goal’, Retrieved January 30, 2023, from https://ko.wikipedia.org/wiki/%EA%B3%A8_%EB%95%8C%EB%A6%AC%EB%8A%94_%EA%B7%B8%EB%85%80%EB%93%A4#2023%EB%85%84
- [3] SBS Website, ‘Kick a Goal’, Retrieved January 11, 2023, from <https://programs.sbs.co.kr/enter/kickagoal/about/67983>
- [4] E. J. Kim, “The Narrative and representations of women's sports in the media: An analysis of the SBS entertainment program Women Who Score Goals,” *PNU Journals of Women's Studies*, Vol. 32, No.1, pp. 7-49, 2022. DOI: doi.org/10.22772/pnujws.32.1.202204.7
- [5] T. H. Kim and H. S. Kim, “Suggestions on how to activate the stay type Ulsan tourism using Big Data Analysis,” *Korea Academic Society of Hotel Administration*, Vol. 35, No. 1, pp. 47-76. 2019. DOI : [10.22793/indinn.2019.35.1.003](https://doi.org/10.22793/indinn.2019.35.1.003)
- [6] J. H. Kim and H. R. Lee, “Comparative study on the changes in perceptions of camping in tourists before and after the COVID-19 outbreak using social media big data: Focused on semantic network analysis” *Korean Journal of Hospitality and Tourism*, Vol. 30, No. 5, pp. 117-134. 2021. DOI : <https://doi.org/10.24992/KJHT.2021.7.30.05.117>
- [7] Y. T. Kim, M. S. Seen, and S. K. Oh, “The analysis in the lifestyles of MICE participants using big data,” *Journal of Tourism and Leisure Research*, Vol. 28, No. 10, pp. 21-34. 2016.
- [8] A. E. Jeon and Y. N. Lee, “A study on the passenger perceptions of “no landing flights” using big data,” *International Journal of Tourism and Hospitality Research*, Vol. 36, No. 3, pp. 107-120. 2022. DOI : [10.21298/IJTHR.2022.3.36.3.107](https://doi.org/10.21298/IJTHR.2022.3.36.3.107)
- [9] Y. H. Kim and Y. J. Kim, “Social Network Analysis(4th)”, Seoul: Parkyoungsa.
- [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.