

**A study on the analysis of water management issues using text mining based on
government press release and online news data**

By

YOO, Seungyeol

CAPSTONE PROJECT

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC MANAGEMENT

2022

**A study on the analysis of water management issues using text mining based on
government press release and online news data**

By

YOO, Seungyeol

CAPSTONE PROJECT

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC MANAGEMENT

2022

Professor Kim, Seong-han

**A study on the analysis of water management issues using text mining based on
government press release and online news data**

By

YOO, Seungyeol

CAPSTONE PROJECT

Submitted to

KDI School of Public Policy and Management

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF PUBLIC MANAGEMENT

Committee in charge:

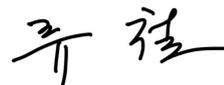
Professor Kim, Seong-han, Supervisor



Professor Lee, Junesoo



Professor Liu, Cheol



Approval as of May, 2022

ABSTRACT

A Study on the Analysis of Water Management Issues using Text Mining based on Government Press Release and Online News Data

By

YOO, Seungyeol

This research paper aims to check whether there are differences in government attention and social attention on major water management issues and to explore how to derive the differences, using text mining techniques based on government press release and online news data.

For this research, keywords on water management issues were derived from press release of the Ministry of Environment (MOE) and online news articles from 2016 to 2020, and the frequency and content of the keywords were compared in terms of governmental and social side. In this process, the Bigram and Word2Vec methods were used to derive keywords that could be sufficiently expressed information of issues on water management. As a result of this research, it was confirmed that the extent of the governments attention and social attention may depending on the water management issues during same period, and that what issues differed by year.

Finally, I think this research result is meaningful in laying the foundation for more research and development in a more advanced direction, not a methodology that dramatically improves the analysis of policy issues.

TABLE OF CONTENTS

1. Introduction	1
2. Literature review	3
3. Problem statement	4
4. Methodology of study	6
4.1 Procedure	6
4.2 Data collection	8
4.3 Data pre-processing	9
4.4 Bigram keyword extraction	10
4.5 Identification of main keyword	11
4.6 Identification of related word using Word2Vec	12
4.7 Chi-Square test	13
4.8 Frequent keywords comparison of press release and online news articles	14
5. Result and discussion	16
5.1 Result	16
5.1.1 Keywords on water management issue	16
5.1.2 Chi-Square test	16
5.1.3 Frequent keywords comparison	22
5.2 Discussion	25
6. Conclusion	30
6.1 Main conclusion of study	30
6.2 Suggestions for follow-up research	31
References	32
Appendix	33
1. Configuration information for data collection and analysis	
2. Detailed analysis result of keywords on water management issues	

LIST OF TABLES

Table 1. Numbers of the MOE press release related to Water Management	8
Table 2. Numbers of Bigram keyword extracted from the MOE press release	10
Table 3. Form of cross tabulation for Chi-Square test	14
Table 4. Main Keywords on water management issue in 2016	17
Table 5. Main Keywords on water management issue in 2017	18
Table 6. Main Keywords on water management issue in 2018	19
Table 7. Main Keywords on water management issue in 2019	20
Table 8. Main Keywords on water management issue in 2020	21
Table 9. Result of Chi-square test	22
Table 10. Comparison table of frequent keywords on press release and online news articles in 2020	24
Table A1-1. Department related to water management works of the MOE	33
Table A1-2. User dictionary list	34
Table A1-3. Stopwords list	36
Table A1-4. A List of keywords that cannot be selected as main keyword	38

LIST OF FIGURES

Figure 1. Procedure of this study	7
Figure 2. CBOW and Skip-gram architecture of Word2Vec	13
Figure 3. Concept of comparison graph	15
Figure 4. Comparison graph of frequent keywords on press release and online news articles in 2020	25
Figure A2- 1. Analysis result of keywords on water management issues in 2016	40
Figure A2-2. Comparison graph of frequent keywords on press release and online news articles in 2016	41
Figure A2-3. Analysis result of keywords on water management issues in 2017	42
Figure A2-4. Comparison graph of frequent keywords on press release and online news articles in 2017	43
Figure A2-5. Analysis result of keywords on water management issues in 2018	44
Figure A2-6. Comparison graph of frequent keywords on press release and online news articles in 2018	45
Figure A2-7. Analysis result of keywords on water management issues in 2019	46
Figure A2-8. Comparison graph of frequent keywords on press release and online news articles in 2019	47
Figure A2-9. Analysis result of keywords on water management issues in 2020	48
Figure A2-10. Comparison graph of frequent keywords on press release and online news articles in 2020	49

1.Introduction

For Korean domestic public enterprises, identifying government policy trends and social requirements periodically is one of the significant tasks. Domestic public institutions shall establish and submit mid- to long-term management goals for five fiscal years including the following year, to the government in accordance with Article 46 (1) of the Act on the Management of Public Institutions every year. Accordingly, domestic public enterprises submit management goals every year, and most of them establish additional management strategies to achieve management goals. One of the most important procedures in establishing management goals and management strategies every year is to identify government policy trends and social requirements. Since government policies and social requirements are external factors that have the greatest influence on the management of public enterprises. According to Kim (2014), in the case of public interest-oriented public enterprises, the government's policy decisions and policy projects are direct and practical explanatory factors for expanding the business area of the public enterprise. And, looking at the individual laws that are the basis for the establishment of domestic public enterprises, it is revealed that the purpose of enacting the law and establishing the public enterprises is to contribute to the promotion of public welfare such as people's life and the national economy.

In the domestic water management field, for the following two reasons, the difficulty of identifying government policy trends and social requirements is aggravated. First, the business scope of public enterprises related to water management is quite wide. According to Article 9 of the Korea Water Resources Corporation Act, K-water's business scope includes traditional water management business such as water resource construction and operation, waterworks construction and management, water environment management, in addition,

water reuse, developing industrial complexes, new and renewable energy establishment, basic investigations, diagnosis, technology development and education related to water management. Second, related research cases are insufficient. Various studies have been conducted on the fields of analyzing laws and institutional problems and suggesting improvement measures for water management, but it is difficult to find research cases in areas that derive major issues so that policy or social issues can be overview on water management.

Despite the above difficulties, government press releases and online news data can be useful to identify policy and social issues on water management. Since government press releases are produced to promote and guide government policies, the press releases of government ministries in charge of water management will reflect the government's policy attention in water management issues. And, since domestic online news media are updating press releases in real time on issues that are attracting social attention, it is believed that online news articles will reflect social attention in water management issues. In addition, it will be very useful to use text mining to understand the contents of water management issues from these government press releases and online news articles data. This is because text mining has been widely used in the work of extracting keywords representing the content of the text from text data sets, through recent technical advances.

The purpose of this study is to explore how to derive differences in major water management issues in policy and social aspects using text mining techniques based on government press releases and online news articles. The result of this study is expected to help public enterprises in charge of water management establish management goals and management strategies by allowing them to identify policy and social issues on water management to be actively responded to.

2. Literature Review

Big data analysis of complaints using text mining techniques: Gangnam-gu (Park, 2020)

Park (2020) attempted to derive keywords for civil complaints by analyzing civil complaints received in Gangnam-gu, Seoul from 2014 to 2019 using text mining techniques, and to analyze the types and causes of civil complaints through the derived keywords. Park reveals that the following two data pre-processing processes are important in analyzing text data.

First, a separate user dictionary consisting of compound nouns should be constructed. As a pre-processing of Korean text data, the python package is used to extract nouns based on Korean part of speech, and the basic python package has the disadvantage of not being able to properly distinguish new words or compound nouns. Therefore, in order not to exclude meaningful words from the pre-processing process, a separate user dictionary must be configured and reflected in the python package.

Second, after extracting Korean nouns, words that do not fit the purpose of analysis should be designated as stopword. For example, general words such as 'this', 'that', and 'the head of the district' are not directly related to the purpose of the study, but are feared to select as major keywords due to their high frequency of appearance in text data. These words should be excluded from the text data set in pre-processing step.

However, in this study, despite analyzing the network relationship between keywords after deriving major keywords from civil complaints data, it is confirmed that it may be difficult to accurately grasp what the original text data means if keywords are extracted in one word unit. For example, as a result of network analysis of the keyword "hospital", "care", "happiness", and "patient" were presented as highly relevant keywords, but it was difficult to

intuitively grasp which contents of complaint derive such keywords.

Natural Language Processing and Computational Linguistics (Srinivasa-Desikan, 2018)

Srinivasa-Desikan (2018) introduced the contents and trends of data analysis and processing technologies, from the most basic data pre-processing to the latest technologies, deep learning and chatbot. According to Srinivasa-Desikan (2018), the following two techniques suggest that keywords extracted from text data will help convey information about the meaning of the context.

First, the keyword should be extracted by Bigram method. Bigram means not extracting keywords in units of one word, but extracting keywords in the form of a phase consisting of two words. Srinivasa-Desikan (2018) argued that Bigram is the most commonly used language model in recent years and helps prevent context loss problems.

Second, it is to find a pair of words related to keywords by Word2Vec method. Word2Vec converts the word into a vector form, and it is possible to find words with similar meanings in consideration of the meaning of the word. Srinivasa-Desikan (2018) introduces the gensim library that can be used in python programs to utilize Word2Vec, and introduces the Similarity function of the library as a way to find meaningful word pairs.

From these results, if keywords for water management issues are derived from text data such as government press releases, it is judged that Bigram and Word2Vec will be helpful to minimize the loss of context meaning.

3. Problem Statement

This study attempted to explore how to derive differences in major water management issues in policy and social aspects using text mining techniques based on

government press releases and online news articles, and selected the following research problem.

[Research Problem] Can be confirmed the difference between government attention and social attention on water management issues using text mining analysis based on the Ministry of Environment(hereafter, MOE) press release and online news articles?

If there is a water management issue that differs in the government's attention and social attention, the direction of establishing management goals or management strategies of public enterprise in charge of water management may vary depending on the difference. For example, if the government's attention in any negative water management issue is higher than social attention, it will be necessary to prepare a new policy proposal to address the social needs of the issue. Conversely, if the social attention in a positive water management issue is lower than the government's attention, a plan should be prepared so that the public and society can feel the purpose or achievement of the government's policy on the issue more. This is because public enterprises are organizations that realize government policies and at the same time are organizations established to pursue the public attention of society. Therefore, grasping the difference between government-side attention and social-side attention in water management issues was judged to be a meaningful topic that can contribute to the process of establishing management goals and management strategies of public corporations in charge of water management.

In order to solve the research problem, the MOE press release is analyzed as text mining using Bigram and Word2Vec to derive keywords representing water management issues, and check the frequency of appearance from press release and online news articles for those keywords. It is intended to that the frequency in the press release of the MOE is

regarded as government attention, and the frequency from online news articles is regarded as social attention rest.

Government press releases are usually produced to promote government policies and guide government policies on policy issues. And according to Han (2019), the MOE, which is in charge of preservation of the water environment and conservation and use and development of water resources under the current government organization law, is a key ministry in charge of water management in Korea. Therefore, the MOE press release is suitable for grasping domestic water management issues, and it can be seen that the frequency of water management issue keywords in the MOE press release reflects the government attention in the issue.

Park (2014) claimed that media reporting plays a decisive role in growing an issue into a social issue by attracting the general public's attention. Therefore, it can be seen that the frequency of water management issue keywords appearing in online news articles reflects the social attention in the issue.

Taken together, it is decided that the government attention and social attention in the issue can be confirmed respectively through the analysis of the frequency of keywords on water management issues appearing in the MOE press release and online news articles.

4. Methodology of study

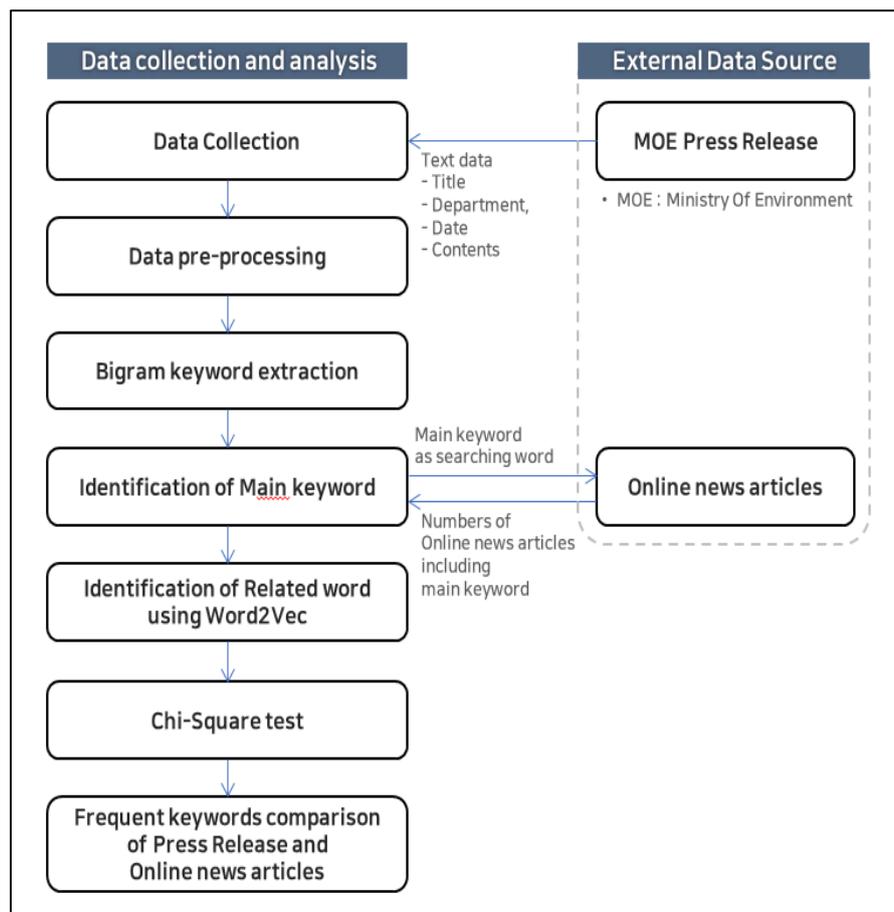
4.1 Procedure

The procedure for this study is shown in Figure 1. First, press releases from the MOE were collected, and pre-processing of the press release text data was performed. Next, the Bigram keyword was extracted from the press release text data, and the main keyword

representing the water management issue was selected among the extracted keywords. In addition, the frequency from press releases and online news articles on the main keyword was confirmed. Next, the related word for the main keyword was selected using Word2Vec. After that, a chi-square test was conducted to check whether there was a significant difference in the frequency distribution of the main keyword according to the difference between the MOE press release and the online news articles, and finally, the contents of the frequent Main keyword were compared.

Basically, the python program was used to collect and analyze data to derive results, and the SPSS program was used for the chi-square test.

Figure 1. Procedure of this study



4.2 Data collection

As the first step in this study, press releases from the MOE were collected using the web scraping method. First, information on the title, department in charge, and registration date of the MOE press release from 2016 to 2020 was collected from the MOE website. Next, only press release information registered by the department in charge of water management-related affairs was selected. A total of 47 departments were designated as departments of the MOE in charge of water management-related tasks (see Appendix 1, Table A1-1). Finally, additional the body data of the press release was collected from the MOE website for the selected press releases related to water management.

As a result of collecting press release data from the MOE, the number of press releases is shown in Table 1. From 2016 to 2020, there were a total of 5,051 press releases by the MOE, of which 757 press releases related to water management. In this study, the work of extracting and analyzing keywords related to water management was performed based on 757 text body data of the press release related to water management.

Table 1. Numbers of the MOE press release related to water management

Category	2016	2017	2018	2019	2020	Total
Numbers of the MOE press release	833	893	890	1,235	1,200	5,051
Numbers of the MOE press release related to water management	114	153	134	164	192	757

4.3 Data pre-processing

Data pre-processing was performed as a preparation process for the analysis of text data. Text data pre-processing is a step in which original text data, which is a mixture of sentences, words, and special characters, is converted into a form suitable for data analysis. Data pre-processing was carried out in the order of POS (part of speech) tagging, tokenization, and stopword processing.

POS tagging and tokenization refer to changing original text data in sentences or phrases into a set of words necessary for analysis. In this study, Konlpy, a Korean only tagging and tokenization library of the Python program, was used. Konlpy's open Korean text module provides the function of separately classifying only nouns from text data. Using the Okt module in Konlpy library, a text dataset consisting only of nouns and numbers was generated from the original text data of the press release related to the MOE. However, the Okt module automatically recognized and processed common nouns stored in advance in its dictionary, but did not properly recognize complex nouns in which two or more common nouns were combined in single word unit. For example, for compound nouns recently used as one Korean word in the domestic water management field, such as "수생태계(means Aquatic Ecosystem)", the Okt module recognized them in two words: "수(means Aqua)" and "생태계(means Ecosystem)". As another example, the Okt module recognized "물산업(means Water Industry)" in two words: "물(means Water)" and "산업(means Industry)". Therefore, in order for the Okt module to properly recognize the compound nouns mainly used in single word unit, a user dictionary consisting of 212 compound nouns (see Appendix 1, Table A1-2) was configured and added to the Okt module's own list of nouns.

Stopword processing is an operation of removing meaningless characters in advance, including special characters (e.g., parentheses, #, :, etc.). In this study, all other special characters except Korean, alphabet, and number were removed as basic stop word processing. In addition, 294 nouns that are not directly related to water management were selected among the words included in the press release to form a separate list of stopwords (see Appendix 1, Table A1-3), and the stopwords were excluded from the text dataset.

4.4 Bigram keyword extraction

Keywords in the form of Bigram were extracted from the text data set of the press release of the MOE that was pre-processed. Bigram means a method of processing two sets of words into one keyword. In most existing research cases, in the case of extracting keywords through text mining, a Unigram method was used to process one word into one keyword. However, it was difficult to infer what purpose and meaning the keyword was used in the original document because the unigram method has limitations in information that the keyword can express. Therefore, in this study, the Bigram method was applied so that keywords for water management policy issues can express more specific information. Table 2 shows the number of Bigram keywords extracted from the text data of the MOE press release by year in this study

Table 2. Numbers of Bigram keyword extracted from the MOE press release

Category	2016	2017	2018	2019	2020	Total
Numbers of Bigram keyword extracted from the MOE press release	16,341	21,859	25,342	27,404	31,066	122,012

4.5 Identification of main keyword

The main keywords representing the water management issues was classified from the Bigram keyword set, and the frequency of appearance in the MOE press release and online news articles of the main keyword was calculated.

To classify the main keyword representing the water management issue, first, keywords with high frequency of appearance but not directly related to water management were selected in advance so that the keyword was not selected as the main keyword. For example, Bigram keywords such as 'Plan Establishment' and 'Promotion Plan' have a high frequency of appearance, but the keywords cannot identify the association with water management intuitively. In this way, 192 Bigram keywords that cannot directly represent the water management issue were selected (see Appendix 1, Table A1-4), and these keywords were first excluded from the main keyword identifying, and only were selected as related words that serve as supplementary explanations for the main keyword. In addition, Bigram keywords with too little frequency of appearance in press releases of the MOE and online news articles were excluded from the main keyword identifying and processed and only were selected as related words. In this study, the lower limit of the frequency of appearance was designated as 1% of the maximum frequency of appearance in press releases of the MOE and online news articles. Among the Bigram keywords from 2016 to 2020, the maximum frequency of press release was 110, and the maximum frequency of online news articles was 4,000. Therefore, Bigram keywords less than one frequency of appearance in the press release of the MOE or less than 40 frequency of appearance in online news articles were excluded from selecting as main keyword.

The search target of online news articles was limited to articles searched through the

Naver portal website. Naver and Daum are the representative portal sites in Korea where can search for online news articles, and Naver provides up to 4,000 news articles for news search keywords and Daum up to 800 news articles. Therefore, it was decided to use the Naver Portal, where can check more news articles. And, using the Naver news search option function, only articles containing a phrase that exactly matches the main keyword of Bigram form were counted as frequency of main keyword in online news articles.

4.6 Identification of related word using Word2Vec

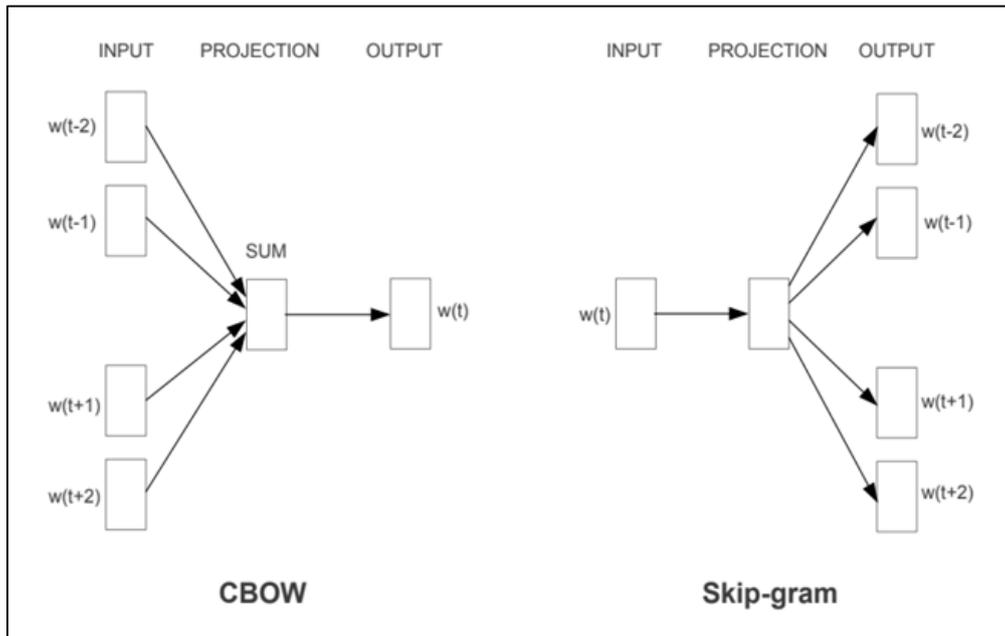
For each main keyword, Word2Vec method was used to select three related words that were highly related to the main keyword.

Word2Vec is an algorithm that enables to find keywords that are highly related to a specific keyword in a text dataset. Mikolov et al. (2013) revealed that Word2Vec is implemented in two models, CBOW and Skip-gram, respectively, as shown in Figure 2, and among them, Skip-gram was more accurate in finding semantic relationships between words. Therefore, Skip-gram was applied in this study. Skip-gram proceeds with learning by predicting which words can be located around the target word for a text dataset.

In this study, Gensim, a python library, was used to utilize Word2Vec. After converting the Bigram keyword set into a word vector using the Gensim library, words with high cosine similarity to the Main keyword were extracted as related words through the Gensim's Similarity function. The cosine similarity has a value between 0 and 1, and as the association with the main keyword increases, a value close to 1 is calculated. Here, a high correlation with the main keyword means that the word vector and cosine distance of the main keyword are close, which means that there is a high probability of being placed close to the main keyword within the original text data. In this study, to prevent keywords that are not

related to the main keyword from being selected as related words, keywords that did not appear in the same document with the main keyword or keywords with a cosine similarity calculation value of less than 0.90 were excluded from selecting as related word.

Figure 2. CBOW and Skip-gram architecture of Word2Vec



[Source : Mikolov et al. (2013)]

4.7 Chi-Square test

A chi-square test was performed using the SPSS program to check whether there was a statistically significant difference in the frequency distribution of the main keywords on water management issue according to the difference between media such as the MOE press release and the online news articles. For each year from 2016 to 2020, it was created a cross tabulation for chi-square test for the top 30 main keywords based on the frequency in the MOE press release as shown in Table 3, and the significance level was determined to 0.05.

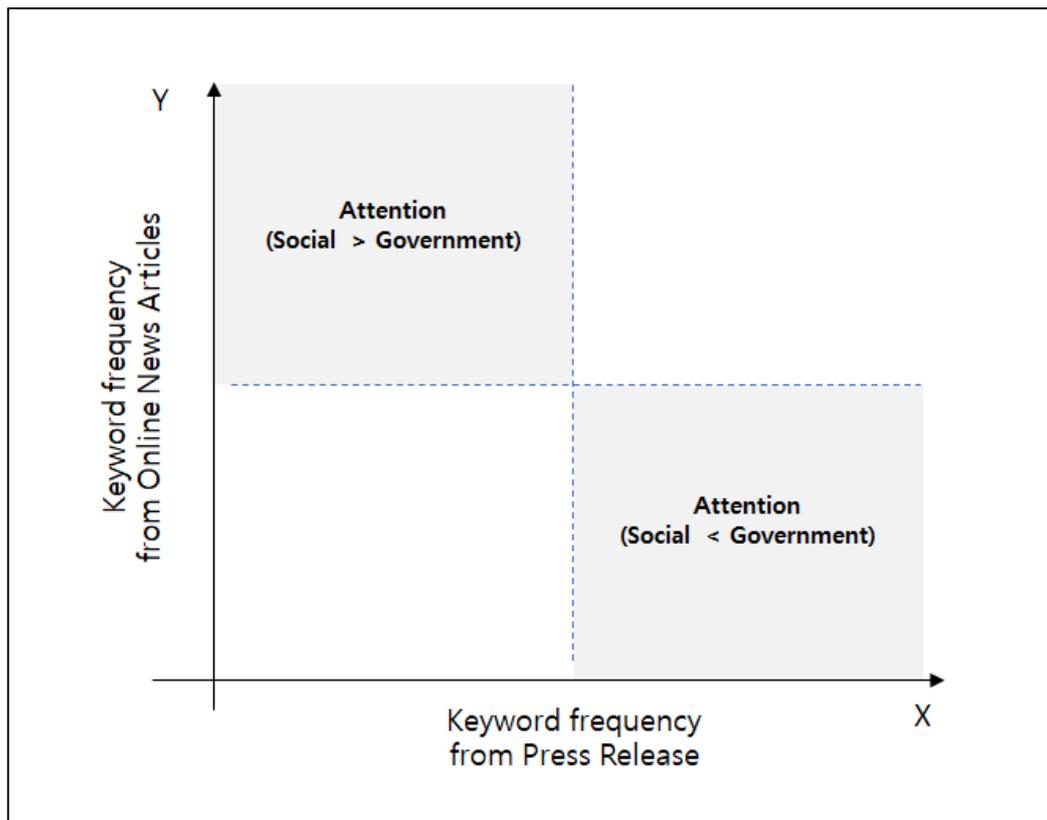
Table 3. Form of cross tabulation for Chi-Square test

Category		Main Keyword				
		Keyword #1	Keyword #2	~	Keyword #29	Keyword #30
Meida	Keyword frequency from the MOE press release			~		
	Keyword frequency from online news articles			~		

4.8 Frequent keywords comparison of press release and online news articles

Finally, for the main keyword representing water management issues, the frequency of press releases by the MOE and online news articles was compared in the form of table and graph. In particular, it is shown as a comparative graph as shown in Figure 3 so that the difference of frequency of keyword appearance in the two media can be visually confirmed. The X-axis is the frequency in the MOE press release for the Main keyword, and the Y-axis is the frequency of online news articles for the Main keyword. In the upper left area of the comparison graph, main keywords with a relatively higher frequency of online news articles than the frequency of press releases from the MOE are located. The water management issue of Main Keyword, located in this area, can be estimated that social attention is higher than that of the government. Conversely, in the lower right area of the comparison graph, main keywords with a relatively higher frequency of press releases from the MOE than the frequency of online news articles are located. The water management issue represented by the main keyword located in this area can be estimated that the government attention is higher than social attention.

Figure 3. Concept of comparison graph



5. Result and discussion

5.1 Result ¹

5.1.1 Keywords on water management issue

Tables 4~8 show the results of analyzing the main keyword on water management based on the MOE press release and online news articles from 2016 to 2020. The tables are presented the top 30 main keywords on the basis of frequency of press release (see also Appendix 2).

5.1.2 Chi-Square test

Table 9 shows the result of the chi-square test to determine whether the frequency distribution of the main keyword on the water management issue is different due to the difference between medias such as the MOE press release and online news article. From 2016 to 2020, the p-value of the chi-square test for each case was all less than 0.001. Therefore, it was confirmed that there was a statistically significant difference in the frequency distribution of the main keyword on the water management issue between the press release of the MOE and the online news articles.

¹ In this study, all processes of extracting the Bigram keyword from text data were performed based on Korean characters. The extracted Bigram keywords consist of two Korean words, but in the process of translating Korean into English to display the analysis results in this report, the number of words translated into English and the number of words translated into English sometimes changed. For example, if the Bigram keyword "수질 개선" consisting of two Korean words was translated into English, the keyword was written to "Water quality improvement" and required three English words. As in this example, even if the keyword extraction results in this report were expressed in three or more English words, it was not a bad performance of Bigram keyword extraction, but just an inevitable result of the translation process because the language systems of Korean and English are different.

Table 4. Main keywords on water management issue in 2016

No.	Main Keyword	Frequency from the MOE press release	Frequency from online news articles	Remark (Main Keyword in Korean)
1	Water industry Promotion	35	500	물산업 육성
2	Strategic Environmental impact assessment	33	1010	전략 환경영향평가
3	International Water week	32	80	국제 물주간
4	Water industry Cluster	28	450	물산업 클러스터
5	Water circulation Leading	24	320	물순환 선도
6	Ecology Restoration project	22	820	생태 복원사업
7	Green algae Occurrence	22	1250	녹조 발생
8	Boryeong-dam Intake pipeline	20	430	보령댐 도수로
9	Pollution Load management	18	320	오염 총량관리제
10	Algae occurrence	18	510	조류 발생
11	World Water day	17	1490	세계 물의날
12	Smart Water grid	16	250	스마트 워터그리드
13	Soil Groundwater	13	220	토양 지하수
14	4 Major rive Project	13	3830	4 대강 사업
15	Tap water Production	12	940	수돗물 생산
16	Non point Pollution	12	1500	비점 오염
17	Nakdonggang river Catchment area	12	900	낙동강 수계
18	Draught Occurrence	12	320	가뭄 발생
19	Water source Conservation area	11	2180	상수원 보호구역
20	Water Meter	11	2120	수도 계량기
21	Seawater desalination Plant	11	580	해수담수화 플랜트
22	Mini Environmental impact assessment	10	210	소규모 환경영향평가
23	Gunnam Flood control reservoir	10	350	군남 홍수조절지
24	Aquatic ecosystem Preservation	10	100	수생태계 보전
25	Domestic Water industry	10	250	국내 물산업
26	Water industry Technology	10	90	물산업 기술
27	Water source quality	9	700	상수원 수질
28	Water circulation Improvement	9	110	물순환 개선
29	Drought Warning level	9	70	가뭄 주의단계
30	Water supply Adjustment	8	80	용수공급 조정

Table 5. Main keywords on water management issue in 2017

No.	Main Keyword	Frequency from the MOE press release	Frequency from online news articles	Remark (Main Keyword in Korean)
1	Drought Prevention	46	1960	가뭄 대비
2	Drought Measures	43	4000	가뭄 대책
3	Water Supply	38	4000	용수 공급
4	Alternative Supply	35	420	대체 공급
5	Drought Warning	33	70	가뭄 경보
6	Boryeong-dam Intake pipeline	31	530	보령댐 도수로
7	LID Method	30	100	LID 기법
8	National Multi-purpose dam	28	390	전국 다목적댐
9	International Water week	27	140	국제 물주간
10	Micro Plastics	27	520	미세 플라스틱
11	Aquatic ecosystem Preservation	26	160	수생태계 보전
12	Well Development	23	2890	관정 개발
13	Non point Pollution	23	1550	비점 오염
14	Local waterworks Modernization	23	1320	지방상수도 현대화
15	Drought Area	23	1780	가뭄 지역
16	Water Reserve	22	160	용수 비축
17	Water Supply	21	510	상수도 보급
18	Agriculture Reservoir	20	1610	농업 저수지
19	Small Waterworks facility	19	650	소규모 수도시설
20	Water Shortage	18	2630	용수 부족
21	Aquatic ecosystem Health	18	200	수생태계 건강
22	Common Drought	18	550	상습 가뭄
23	Opening Water level	17	380	개방 수위
24	Soil Groundwater	17	230	토양 지하수
25	District Drought	17	1140	지역 가뭄
26	Green algae Status	17	1190	녹조 현상
27	Agricultural water Supply	16	3280	농업용수 공급
28	National Drought	16	140	전국 가뭄
29	Wide waterworks Business	16	470	광역상수도 사업
30	Multi-purpose dam Capacity	15	60	다목적댐 저수량

Table 6. Main keywords on water management issue in 2018

No.	Main Keyword	Frequency from the MOE press release	Frequency from online news articles	Remark (Main Keyword in Korean)
1	Harmful Blue-green Algae	54	780	유해 남조류
2	Water management Unification	53	1430	물관리 일원화
3	Algae Warning	52	1490	조류 경보
4	PFAS	37	1570	과불화 화합물
5	Green algae Occurrence	36	1400	녹조 발생
6	Naturalness Restoration	30	690	자연성 회복
7	Nakdonggang river Catchment area	29	1400	낙동강 수계
8	Opening Monitoring	27	170	개방 모니터링
9	Groundwater Pollution	25	1080	지하수 오염
10	Green algae Response	21	270	녹조 대응
11	Opening Period	21	610	개방 기간
12	Soil Groundwater	19	140	토양 지하수
13	Algae Toxin	19	180	조류 독소
14	Water management Technology	19	780	물관리 기술
15	Alternative Supply	18	330	대체 공급
16	Water Shortage	17	1530	용수 부족
17	Hangang river Nakdonggang river	16	50	한강 낙동강
18	Geumgang river Catchment area	16	910	금강 수계
19	Groundwater Use	16	490	지하수 이용
20	Products for Waterworks	16	50	수도용 제품
21	Non point Pollution	14	1520	비점 오염
22	World Water day	14	1610	세계 물의날
23	Water velocity Increasement	13	60	유속 증가
24	Governance	13	90	협치 거버넌스
25	Drought Measures	13	2180	가뭄 대책
26	Pusan Eco-delta city	13	1210	부산 에코델타시티
27	Rainfall Radar	12	60	강우 레이더
28	Local waterworks Modernization	12	1330	지방상수도 현대화
29	Aquatic ecosystem Health	12	210	수생태계 건강
30	Ecological Space	12	1050	생태 공간

Table 7. Main keywords on water management issue in 2019

No.	Main Keyword	Frequency from the MOE press release	Frequency from online news articles	Remark (Main Keyword in Korean)
1	Geungang river Yongsangang river	79	50	금강 영산강
2	Water industry Cluster	70	400	물산업 클러스터
3	Green algae Occurrence	65	1190	녹조 발생
4	Water management Committee	56	1530	국가 물관리위원회
5	Basin Water management Committee	40	620	유역 물관리위원회
6	Naturalness Restoration	39	1260	자연성 회복
7	National Water management	39	340	국가 물관리
8	Water quality Ecology	37	60	수질 생태
9	Incheon Tap water	33	1110	인천 수돗물
10	Water management Basics	33	280	물관리 기본
11	Aquatic ecosystem Health	31	270	수생태계 건강
12	Water gate Opening	30	1280	수문 개방
13	4 Major river Project	29	3780	4 대강 사업
14	Water quality Improvement	28	4000	수질 개선
15	Non point Pollution	27	2300	비점 오염
16	Water environment Preservation	26	90	물환경 보전
17	Opening Period	26	1300	개방 기간
18	Water Supply	23	4000	수돗물 공급
19	Nakdonggang river Upstream	22	870	낙동강 상류
20	Well Development	21	1220	관정 개발
21	Green algae Response	21	340	녹조 대응
22	Tab water Quality	21	1920	수돗물 수질
23	Rainfall Radar	21	170	강우 레이더
24	Groundwater Level	21	550	지하 수위
25	Hangang river Nakdoggang river	20	90	한강 낙동강
26	Water industry Promotion	20	310	물산업 진흥
27	Micro Plastics	19	4000	미세 플라스틱
28	Water management Unification	18	730	물관리 일원화
29	Andong-dam Upstream	18	170	안동댐 상류
30	Bottling Tap water	18	510	병입 수돗물

Table 8. Main keywords on water management issue in 2020

No.	Main Keyword	Frequency from the MOE press release	Frequency from online news articles	Remark (Main Keyword in Korean)
1	Naturalness Restoration	81	1100	자연성 회복
2	Water industry Cluster	72	80	물산업 클러스터
3	Hydrothermal Energy	65	1790	수열 에너지
4	Water resource Satellite	51	120	수자원 위성
5	Green algae Occurrence	40	930	녹조 발생
6	Rainfall Radar	38	250	강우 레이더
7	Floating Debris	37	1210	부유 쓰레기
8	Flood Damage	36	4000	홍수 피해
9	Algae Warning	35	740	조류 경보
10	Green New deal	34	4000	그린 뉴딜
11	Tap water Larvae	33	4000	수돗물 유충
12	Tap water Supply	32	4000	수돗물 공급
13	Water management Committee	32	780	국가 물관리위원회
14	Flood Response	29	920	홍수 대응
15	City Water circulation	28	220	도시 물순환
16	Waterworks Pipe network	27	740	상수도 관망
17	National Water management	25	170	국가 물관리
18	Water management Unification	25	710	물관리 일원화
19	Basin Water assistance center	24	390	유역 수도지원센터
20	Decrepitude Waterworks	24	1940	노후 상수도
21	Water environment Preservation	23	80	물환경 보전
22	Seomjingang river Basin	23	1470	섬진강 유역
23	Spill Groundwater	22	210	유출 지하수
24	Non point Pollution	22	2170	비점 오염
25	Nakdonggang river Basin	21	3780	낙동강 유역
26	Flood Control	21	4000	홍수 통제
27	Hangang river Catchment area	21	1990	한강 수계
28	Smart Waterworks	20	960	스마트 상수도
29	Artificial intelligence AI	20	880	인공지능 AI
30	River Use	19	50	하천 사용

Table 9. Result of Chi-square test

Category	2016	2017	2018	2019	2020
Pearson Chi-Square value	852.7	1659.4	819.3	3438.4	3375.4
df	29	29	29	29	29
p-value	<0.001	0.000	<0.001	0.000	0.000
Ratio of cells which have expected count less than 5 (%)	13.3	13.3	10.0	6.7	8.3
Minimum expected count	1.68	1.61	1.74	2.58	1.51

5.1.3 Frequent keywords comparison

Keywords with high frequency based on press releases of the MOE and online news articles were compared with each other. Table 10 shows the results of comparing the top 10 keywords by sorting the data on the basis of frequency of each main keyword for the MOE press release and online news articles in 2020. And the comparison graph of the main keyword with the frequency in the press release of the MOE as the X-axis and the frequency of online news articles as the Y-axis are shown in Figure 4.

According to Table 10, in 2020, the MOE press release showed the highest frequency of the keyword "Naturalness Restoration", and from the related words, it was confirmed that this keyword was related to the 4 major rivers. Next of "Naturalness Restoration", the frequency of keywords for each issue was high in the order of Water industry cluster, Hydrothermal energy, Water resource satellites, Green algae occurrence, Rainfall radar, Floating debris, Flood damage, Algae warning, and Green new deal. On the other hand,

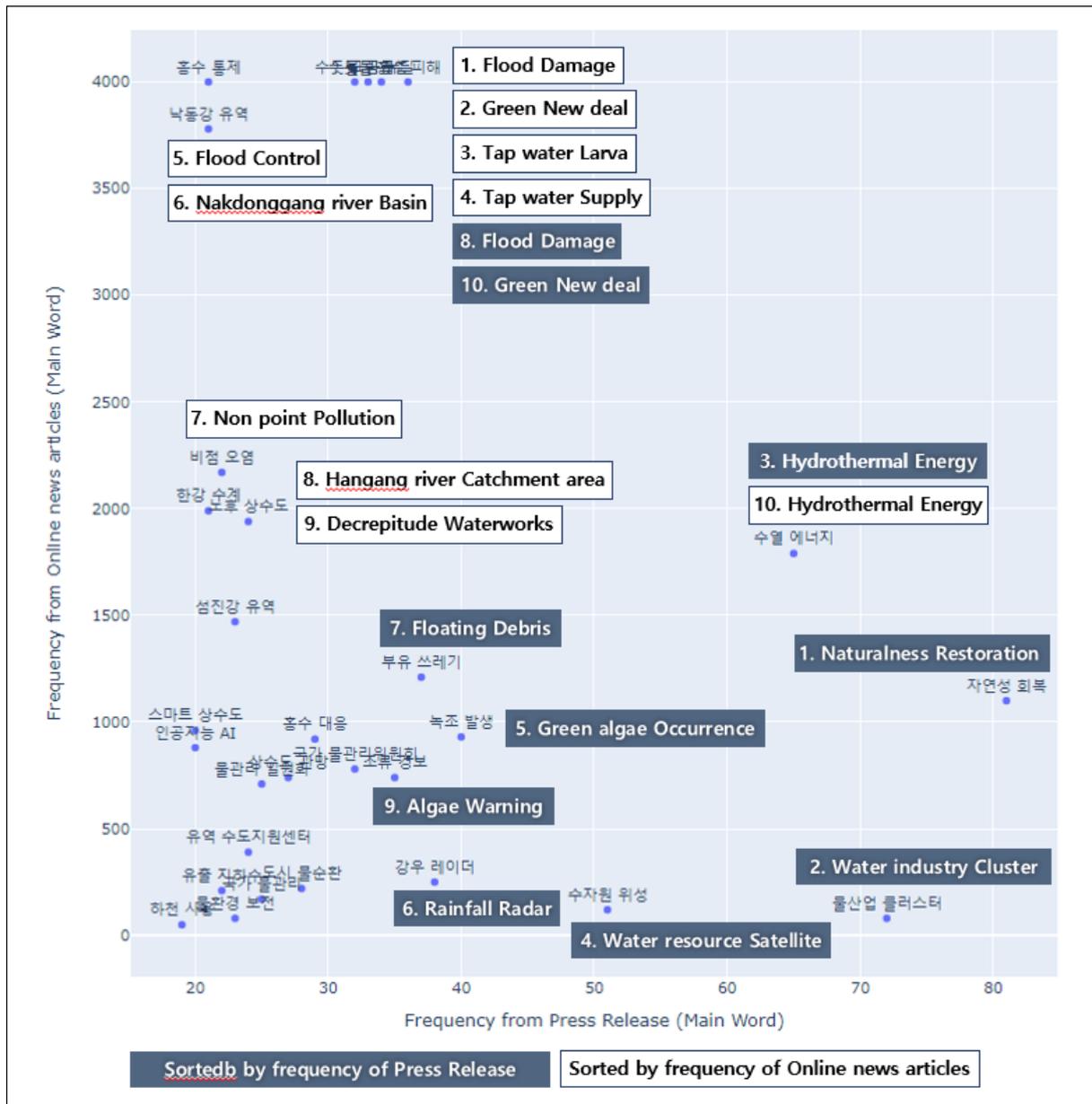
online news articles showed the highest frequency of the keyword "Flood Damage", and related words confirmed that the keyword was related to the Damage area, Damage recovery, and prevention measures. Next of "flood damage", the frequency of keywords was high in the order of Green new deal, Tap water larvae, Tap water supply, Flood control, Nakdonggang river basin, Non point pollution, Hangang river catchment area, Decrepitude waterworks, and Hydrothermal energy.

Among the top 10 keywords sorted by the frequency of press releases of the MOE, seven keywords, such as Naturalness restoration, Water industry cluster, Water resource satellites, Green algae occurrence, Rainfall radar, Floating debris, and Algae warning, were not included in the top 10 keywords by online news articles. In addition, among the top 10 keywords sorted by the frequency of online news articles, 7 keywords such as Tap water larvae, Tap water supply, Flood control, Nakdonggang river basin, Non point pollution, Hangang river catchment area, Decrepitude waterworks were not included in the top 10 keywords in the MOE press release. The difference in these keywords can also be confirmed through the comparison graph of Figure 4. At the bottom right of the comparison graph, keywords with a relatively more dominant frequency of the MOE press release criteria are displayed. It includes Naturalness restoration, Water industry cluster, Hydrothermal energy, and Water resource satellites. In the upper left corner of the comparison graph, keywords with a relatively more dominant frequency of online news articles criteria are displayed. Flood damage, Green new deal, Tap water larvae, Tap water supply, Flood control, and Nakdonggang river basin. In the lower left corner of the comparison graph, keywords that were not frequently displayed on both the MOE press release and the online news articles were displayed.

Table 10. Comparison table of frequent keywords on press release and online news articles in 2020

No	Sorted by frequency of MOE press release		Sorted by frequency of online news articles	
	Main Keyword (frequency)	Related words	Main Keyword (frequency)	Related words
1	Naturalness Restoration (81)	4 Major river Naturalness 4 Major river Investigation Restoration Plan	Flood Damage (4,000)	Damage Area Damage Recovery Prevention Measure
2	Water industry Cluster (72)	Nation Water industry Cluster Company Cluster Business	Green New deal (4,000)	Development Business Activation Plan Plan Water resource
3	Hydrothermal Energy (65)	Eco-friendly Water heat Complex Cluster Energy Complex	Tap water Larvae (4,000)	Larvae Occurrence Larvae Detection Hygienic Management
4	Water resource Satellite (51)	Satellite Development Satellite Utilization Flood Draught	Tap water Supply (4,000)	Hygienic Management, Management Comprehensive countermeasure, People Reassurance
5	Green algae Occurrence (40)	Harmful Blue-green algae Algae Warning system Mulgeum Maeri	Flood Control (4,000)	Flood Information Damage Prevention Management Measure
6	Rainfall Radar (38)	Flood Special report River Water level Sudden Flood	Nakdonggang river Basin (3,780)	Nakdonggang river integrated water management, Upstream Area, Integrated water management Plan
7	Floating Debris (37)	Garbage Collection National River Cleaning Activity	Non point Pollution (2,170)	Installation Operation Performance Test Pollution Facility
8	Flood Damage (36)	Damage Area Damage Recovery Prevention Measure	Hangang river Catchment area (1,990)	Flood control Capacity Flood Prevention Multi-purpose dam Operation
9	Algae Warning (35)	Harmful Blue-green algae Analysis Result Algae Warning system	Decrepitude Waterworks (1,940)	Waterworks Facility Waterworks Management system, Waterworks Maintenance business
10	Green New deal (34)	Development Project Activation Plan Plan Water resource	Hydrothermal Energy (1,790)	Eco-friendly Water heat Complex Cluster Energy Complex

Figure 4. Comparison graph of frequent keywords on press release and online news articles in 2020



5.2 Discussion

In this study, it was attempted to confirm the difference between the government's attention in water management issues and social attention through text mining analysis of press releases from the MOE and online news articles.

First, it was intended to confirm whether there is a difference between the government's attention and social attention in water management issues. In this regard, through the chi-square test, it was confirmed that there is a statistically significant difference of frequency distribution of main keyword on water management issues between press releases from the MOE and online news articles (see Table 9). As mentioned in the definition of problem statement, it can be said that the government's attention in the issue was reflected in the frequency of policy issue keywords shown in the press release of the MOE, a government agency in charge of water management in Korea. In addition, it can be said that the frequency of water management issue keywords appearing in online media reports reflected the social attention in the issue. Therefore, through the results of the chi-square test conducted in this study, it was confirmed that the size of the government's attention and social attention in the same water management issue may differ within a unit period of one year.

Next, as a text analysis method to understand the contents of water management issues with differences in government attention and social attention, this study used Bigram, Word2Vec, comparison tables and comparison graphs. From the press release of the MOE, all keywords for water management issues were derived in the form of Bigram, and the top 30 main keywords were selected on basis of frequency. In addition, using the Word2Vec method, up to three related words with high cosine similarity among keywords appearing in the same document as the main keyword were presented. In addition, a comparison table (see Table 10) and a comparison graph (see Figure 4) for the top 10 keywords were presented based on the frequency of the main keywords of each of the MOE press releases and online news articles. First, it is determined that the use of the Bigram method and Word2Vec has the effect of increasing the information transfer power of the keyword. For example, in Table 10,

the keyword with the highest frequency based on the MOE press release was "Naturalness Restoration", and if keywords were extracted in the Unigram method without using Bigram and Word2Vec as in previous studies, these keywords would have been divided into two words, "Naturalness" and "Restoration", and no related keywords would have been presented. In such Unigram case, it is difficult to grasp which object is natural with only the word "Naturalness." This is because not only rivers but also mountains, seas, and soils can be the subject of the word natural. And similarly, it is difficult to understand what kind of object the word "Restoration" means. This is because the object of recovery may not be natural objects, but old facilities or laws. However, in this study, the keyword was expressed as "Naturalness Restoration" by utilizing Bigram and Word2Vec, and information from related words allowed us to infer that this keyword is related to natural recovery for the four major rivers. According to Srinivasa-Desikan (2018), Bigram helps to prevent the loss of contextual information, and Word2Vec helps to express the meaning of keywords. As Srinivasa-Desikan said, the results of this study confirmed that Bigram and Word2Vec can improve information amount of keyword compared to previous studies using only Unigram. Next, the results of expressing the top 10 keywords in comparison graphs with comparison tables based on the frequency of each of the MOE press releases and online news articles made it possible to intuitively recognize keywords with high government attention and high social attention.

It is determined that the comparison table and comparison graph of water management keywords as above may provide appropriate insights to public enterprises such as K-water in charge of water management to establish management goals and management strategies.

For instance, through the information on keywords that showed relatively much higher government attention in 2020, it can be used to review the background and direction of

the establishment of management strategies as follows.

•○ "Naturalness Restoration" : It can be seen that the main keyword "Naturalness Restoration" means the issue of naturalness restoration for the 4 major rivers through the information of related words. The "Naturalness Restoration" keyword had been appearing as the main keyword since 2018 (see Table 6-7), but social attention was still relatively low in 2020. Therefore, policy efforts and promotional measures are needed to enable the people and society to sympathize with the need for the naturalness restoration of 4 major rivers and to experience the results.

•○ "Water industry Cluster" : It can be seen that the main keyword "Water industry Cluster" means an issue for the National Water Industry Cluster through the information of related words. The National Water Industry Cluster is a facility operated at the national level to foster the domestic water industry. Nevertheless, as social attention in water industry clusters was relatively low, it is necessary to check the current status of water industry clusters contributing to fostering the domestic water industry and to allow the public and society to feel the achievement.

•○ "Water resource Satellite" : It can be seen that the main keyword "Water resource Satellite" means an issue of utilizing satellite in relation to flood and drought. Since it appeared as the main keyword for the first time in 2020 and showed low social attention, policy efforts are needed to present specific measures to utilize water resource satellite and to enable the people and society to sympathize with such measures.

As another example, through the information on keywords that showed relatively

higher social attention in 2020, it can be used to review the background and direction of the establishment of management strategies as follows.

- "Tap water Larvae" : It can be seen that the main keyword "Tap water Larvae" means an issue of larvae occurring in tap water through the information of related words. Since this is a negative issue, it is necessary to come up with measures to prevent recurrence of tap water larvae in order to resolve society's negative perception of this issue.

- "Flood Control" : It can be seen that the main keyword "Flood Control" means an issue of damage prevention and management measures against floods through the information of related words. In 2020, both the government's attention in flood damage and social attention were high. Therefore, measures are needed to prevent such flood damage and strengthen management against flood.

- "Nakdonggang river Basin" : It can be seen that the main keyword "Nakdonggang river Basin" means an issue of the Nakdonggang River Integrated Water Management Plan through the information of related words. The Nakdonggang River Integrated Water Management Plan refers to a plan under consideration to resolve water disputes between a number of local governments in the Nakdonggang River basin that have continued for decades. As social attention in this issue was very high, policy solutions to persuade a number of local governments involved to resolve water disputes in the Nakdonggang River basin should be included in the Nakdonggang River Integrated Water Management Plan.

Taken together, the results of text analysis on the MOE press release and Internet news articles conducted in this study confirmed the difference between the government's

attention in water management issues by year and social attention. The results of this study are expected to contribute to public enterprise establishing management strategies and management goals by reflecting both major water management issues on the governmental and social aspects.

However, as a limitation of this study, it was confirmed that an additional clustering step was necessary. Looking at the keywords sorted based on the MOE press release in Table 10, No. 5 "Green algae occurrence" and No. 9 "Algae warning" are keywords that are very related to each other. "Green algae" is a type of "algae", and the fact that most of the related words of "Harmful blue-green algae" and "Algae warning system" coincide also supports this. In addition, looking at the keywords sorted based on online news articles in Table 10, No. 1 "Flood damage" and No. 5 "Flood control" are keywords that are very related to each other. If these highly related main keywords are grouped together and multiple keywords are integrated into any one keyword, more diverse issues will appear at the analysis result list, which will eventually provide analysts with information on more water management issues.

6. Conclusion

6.1 Main Conclusion of Study

The purpose of this study was to use text mining techniques based on government press releases and online news releases to check whether there is a difference in government attention and social attention in major water management issues, and to explore how to derive the difference. In this study, the frequency of keyword appearance of the issue in the MOE press release was regarded as the government's attention, and the frequency of keyword appearance of the issue in the online news press release was regarded as social attention.

The conclusions obtained through the study are as follows. First, as a result of analyzing the MOE press releases and online news articles from 2016 to 2020, it was confirmed that the size of the government's attention and social attention may differ depending on water management issues within a year. Second, the method of presenting the main keyword for the water management issue and related words using Bigram and Word2Vec was more helpful in grasping the meaning of the issue than the method of deriving the keyword using the existing Unigram method. Third, the MOE press release on major water management issue keywords and comparison tables and comparison graphs according to frequency in online news articles were effective in intuitively grasping the contents of water management issues with different government attention and social attention.

This study can be said to be meaningful in notifying public companies in charge of water management that there is a need to grasp the difference between the government's attention and social attention in water management issues, as well as suggesting ways to confirm such differences. Through this, it is expected that public enterprises in charge of water management will contribute to establishing management goals and management strategies by reflecting both government, social, and water management issues.

6.2 Suggestions for Follow-up Research

In order to obtain better analysis results in the future, it is necessary to add a keyword clustering step before finally selecting the main keyword. When keywords with high relevance are grouped together and multiple keywords are integrated into any one keyword, keyword information on more diverse issues may appear in the list as a result of sorting the main keywords based on frequency.

References

- Han, S.W. (2019). *Yuyeokbyeol Tonghampulgwanri Chegye Guchukeul Wihan Beopje Gaeseon Bangan* [Measures to improve legislation to establish an integrated water resources management system for each watershed]. In Jung, W.T., Choi, B.S., Jung, H., Kim, J.W., Kim, C.S., Son, J.S., Kwon, J.H., & Won, K.H. (Eds.), *Gukmini Wonhaneun Tonghampulgwanri* [Integrated Water Resources Management that people want] (1st ed., pp.310~345). Pieona.
- Kim, J.K. (2014). *Gonggieop Jeongchaekron* [Introduction to state-owned enterprise policy]. Goyang-si: Moonwoosa.
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Proceedings of Workshop at *The International Conference on Learning Representations*, arXiv:1301.3781v3
- Park, K.C. (2020). *Textmining Gibeopeul Hwalyonghan Minwon Bigdata Bunseok: Gangnamgu Pyeon* [Big data analysis of complaints using text mining techniques: Gangnam-gu]. Seoul: Seoul Digital Foundation
- Park, K.M. (2014). A Development and Application of the Measurement Model of Social Attention Increased on Public Issue. *Korea Public Administration Quarterly*, 26(2), 163-189
- Srinivasa-Deskian, B. (2018). *Natural Language Processing and Computational Linguistics*. Birmingham, UK: Packt Publishing Ltd.

Appendix 1. Configuration information for data collection and analysis

Table A1-1. Department related to water management works of the MOE

개방모니터링팀	유역계획과
금강유역환경청	유역총량과
기획총괄팀	유역총량연구과
낙동강유역환경청	유역협력소통팀
물관리위원회기획단	토양지하수과
물관리위원회지원단	토양지하수연구과
물관리정책실	평가총괄팀
물산업클러스터 운영준비단	한강물환경연구소
물산업클러스터추진기획단	한강유역환경청
물산업협력과	한강홍수통제소
물이용기획과	홍수대책기획단
물정책총괄과	환경부 4 대강 자연성회복을 위한
물환경공학연구과	조사·평가단
물환경정책과	환경영향평가과
물환경평가연구과	4 대강 자연성회복을 위한 조사·평가단
보개방 상황실	4 대강 조사평가단
보개방상황팀	4 대강 조사평가단 유역협력소통팀
상수원관리과	
상하수도연구과	
수도정책과	
수돗물 안심지원단	
수생태관리과	
수생태보전과	
수자원개발과	
수자원관리과	
수자원정보센터	
수자원정책과	
수질관리과	
심의지원소통팀	
영산강물환경연구소	
예보통제과	

Table A1-2. User dictionary list

강변여과수	기술개발	보호구역	수계기금
강정고령보	낙동강유역본부	복원사업	수도미터
강정보	낙동강유역환경청	부유물	수도법
경계단계	남부지역	북부지역	수도사업
경보제	내진보강	북한지역	수도사업자
고도정수처리	내진울	분석결과	수도시설
고령보	넌스크라	브롬산염	수도요금
고순도	농어촌공사	비점오염원	수도용
공급량	농업용수	사업자	수도정비
공도교	뉴딜	산업단지	수도지원센터
공론화	단일화	산업통상자원부	수생태
공업용수	달성보	삼성서울병원	수생태계
공업용수도	대행업	삼성엔지니어링	수요량
과불화	데이터센터	삽교호	수질개선
과불화옥탄산	도수로	상수관	수질검사
과불화hex산술폰산	동부지역	상수관로	수질관리
관계기관	두산중공업	상수도	수질기준
관리체계	롯데월드	상수도관	수질오염
관심단계	맛냄새	상수원	숙련도
광역상수도	맞춤형	상하수도	스마트시티
광역원수	매수토지	상황실	승촌보
국립환경과학원	머신러닝	생태벨트	시범사업
국비	물관리	생태환경	식품안전
국제인증	물관리위원회	생활용수	신뢰도
그린도시	물산업	서부지역	신뢰성
금강보	물주간	선제적	실수요량
금강유역본부	물환경	선진화	안동댐
금강유역환경청	미세물질	성과공유제	안심지원단
급수체계	배수지	세종보	에코델타시티
기본계획	배출권	세포수	연구단
기수생태계	배출량	소상공인	영산강유역환경청
기수역	보령댐	솔로몬제도	영섬유역본부

Table A1-2. User dictionary list (Continued)

오염물질	종합상황실	하천유지용수	
오염원	주민지원	한강유역본부	
요구량	주의단계	한강유역환경청	
용수공급	죽산보	한국경제	
용수비축	지방비	한국농어촌공사	
용수원	지방상수도	한국상하수도협회	
운영효율화	지역주민	한국수력원자력	
워터그리드	지원단	한국수자원공사	
월드워터챌린지	지하수	한국환경공단	
위탁업자	진흥법	함안보	
위험지도	창녕보	합천보	
유수율	처리시설	현대화	
유지용수	체류시간	현실화	
유해물질	총량관리제	홍수기	
의견수렴	총량제	홍수위	
이행계획	총유기탄소량	홍수조절지	
인공지능	취수장	홍수주의보	
일부지역	측정법	환경대응용수	
임하댐	치수능력	환경영향평가	
자동측정	탄소배출권	환경협력	
자동측정기기	탄소배출량	- 끝 -	
자연성	탄소중립		
잔류염소	테스트베드		
적접화	통합물관리		
절감량	특별교부세		
정보통신	폐수종말처리시설		
정비사업	폐수처리시설		
정수장	풍수해		
정수처리	필승교		
제시안	하수처리		
제한수위	하수처리시설		
종합대책	하천시설		

Table A1-3. Stopwords list

가득	권 OO(name)	다음	물놀이	사장
가축	그동안	다짐	미자	사진
각서	근무	단장	미채	사태
간담	금융	당부	바로	사항
감사	기사	대면	바이	삭제
감사원	기상	대부분	박 OO(name)	산업부
감안	기여	대폭	박 OO(name)	상금
강의	기하	대하	박 OO(name)	상황
강정	김 OO (name)	대한	박 OO(name)	상황실
개교	김 OO (name)	대해	박 OO(name)	생중계
개년	김 OO (name)	덧글	반려견	서도
개사	김 OO (name)	더욱	발표	서민
개요	김 OO (name)	도지사	백만	석탄
개월	김 OO (name)	돌말류	백분율	설명
건의	김 OO (name)	동안	번가	세미나
계재	김 OO (name)	동위	법화	세창
경우	나머지	돼지	별로	소강
경험	나비	득표수	보가	송금
계속	내년	또한	보도	송 OO(name)
고용부	내용	라며	보도자료	수가
공무원	노아	류량	보어	수경
과장	논의	리미	복지부	수자원공사
관계자	농가	린다	본격	시장
관련	농림축산식품부	마련	본부	신 OO(name)
광산	농법	마스크	본사	아동
교수	농식품부	마치	부속	아이디어
교육청	농약	만전	부처	약속
국립환경과학원	농어촌공사	만큼	분야	업자
국무총리	누구	매우	비대	여명
국장	누리집	모두	비대면	역할
국토교통부	다각	모든	사옥	열기
국토부	다만	문 OO(name)	사의	열병
국회	다소	문체부	사자	예년

Table A1-3. Stopwords list (Continued)

예정	일보	차질	표명
오전	일부	참고	하량
오지	임대료	채권	하위
오후	작년	청사	학회
올해	작품	초과	한경
용어	잠자리	총력	한국경제
용역	장관	총재	한국농어촌공사
용의	장 OO(name)	최근	한국수자원공사
우리	장 OO(name)	최대	한국환경공단
우리나라	적극	최대한	한편
원장	전후	최선	합동
원평	절기	최소	항목
웹하드	점차	최소한	해명
위원	정 OO(name)	최악	해상
위원장	정수기	최우수	행사
위해	정책	축사	행안부
유 OO(name)	제로	축산	행정안전부
윤 OO(name)	조명	취수하	향후
응답	조선일보	치늬	헤드라인
이 OO(name)	조희	코로나	현안
이내	종은	크게	현재
이번	주로	타인	홍 OO(name)
이사장	중이	태세	화상회의
이상	중인	토론회	환경부
이 OO(name)	즉각	통한	환경청
이재	지난	통해	황 OO(name)
이전	지난해	퇴비	회의
이 OO(name)	지속	특별자치시	회의실
이하	직구	팔랑	흙탕물
이 OO(name)	질감	퍼포먼스	QR
이후	질의	포럼	- ㄷ -
인근	차관	포인트	
인용	차레	포함	

Table A1-4. A List of keywords that cannot be selected as main keyword

감시 기준	국가 조정	민간 기업	안전 관리
강수량 평년	국내 기업	바닷물 유입	안전 기준
개발 사업	국민 안전	발전 방안	안전 점검
개선 방안	규제 혁신	발전 사업	안정 공급
개선 사업	기간 부여	발전 협의	야생 생물
개정 공포	기술 개발	방사성 물질	업무 협약
건설 사업	기술 발전	방안 제시	영향 개발
검사 결과	기술 사업	방안 제시안	영향 평가
검사 기관	기술 인력	배출 시설	완전 개방
경계 단계	기업 기술	부족 대비	우려 지역
경기 총남	기업 지원	분석 결과	운영 관리
경남 지역	기업 해외진출	비상 공급	운영 기준
경보 발령	기획 위원회	사업 계획	원인 조사
경북 경남	기후변화 협약	사업 시행	위생 관리
경제성 분석	건축 운영	사업 추진	위생 안전
계획 수립	남부 지방	산업 육성	유관 기관
공급 시설	논문 대회	상류 지역	유엔 기후변화
공론화 위원회	누적 강수량	선도 도시	유역 지방
관리 강화	단계 도달	선도 사업	의견 수렴
관리 개선	단계 진입	설치 운영	이행 기간
관리 계획	대상 계획	수용 대표	이행계획 제출
관리 기준	대상 시설	시범 도시	입법 예고
관리 대상	대응 방안	시설 개선	자연 방사성
관리 대책	대응 체계	시설 관리	전남 경부
관리 방안	대책 추진	시설 설치	전남 경북
관리 시스템	대표 지점	시스템 구축	전문 위원회
관리 의무	도시 개발	시행 규칙	전문가 교육과정
관리체계 구축	매입 구역	신청서 제출	점검 결과
관심 단계	멸종위기 야생	실증 실험	점검 실시
구성 운영	모니터링 결과	심각 단계	정밀 조사
국가 계획	모의 훈련	쓰레기 수거	정보 센터
국가 시범	문제 해결	아시아 위원회	정보 시스템

Table A1-4. A List of keywords that cannot be selected as main keyword (Continued)

정보 제공	충남 서부지역	-끝-	
정화 활동	충남 전남		
제도 개선	충남 지역		
조사 결과	취약 지역		
조사 실시	측정 기기		
조사 평가	태양광 발전		
조업 정지	토양 정화		
조정 기준	토지 이용		
종합 계획	통합 관리		
종합 고려	투자 사업		
주민 대표	평가 결과		
주민 참여	평가 대상		
주민지원 법률	평년 수준		
주의 단계	품질 검사		
지도 점검	피해 발생		
지리 정보	피해 방지		
지역 단계	하류 지역		
지역 사회	한국 환경		
지역 주민	해결 방안		
지역 주의단계	허용 기준		
지역 중심	혁신 기업		
지원 사업	현대화 사업		
처리 계획	협약 체결		
처리 방안	협의 구성		
체계 구축	협의 절차		
최적 대안	환경 개선		
추가 대책	환경 관리		
추가 이행	환경 영향		
추진 계획	환경 정보		
추진 방안	환경 조사		
추진 현황	후속 조치		
충남 서부	MOU 체결		

Appendix 2. Detailed analysis result of keywords on water management

Figure A2- 1. Analysis result of keywords on water management issues in 2016

2016 Main Word	Related Words	Frequency from Press Release (Main Word)	Frequency from Online news articles (Main Word)
1	물산업 육성 [육성 전략, 스마트 물산업, 국가 조정]	35	500
2	전략 환경영향평가 [환경영향평가 대상, 대상 계획, 의견 수렴]	33	1010
3	국제 물주간 [대한민국 국제, 성공 개최, 개최 세계]	32	80
4	물산업 클러스터 [기업 기술, 국가 물산업, 기술 발전]	28	450
5	물순환 선도 [개발 기법, 도시 선정, 도시 물순환]	24	320
6	생태 복원사업 [수질 수생태계, 수질 개선, 우수 사례]	22	820
7	녹조 발생 [녹조 대응, 조류 독소, 녹조 현상]	22	1250
8	보령담 도수로 [분기 관로, 조정 기준, 주민 불편]	20	430
9	오염 총량관리제 [삼교호 수계, 총량관리제 시행, 특대 지역]	18	320
10	조류 발생 [조류 대응, 냄새 물질, 조류 독소]	18	510
11	세계 물의날 [물의날 기념, 기념 개최, 기업 학계]	17	1490
12	스마트 워터그리드 [기술 개발, 정보통신 기술, 부족 문제]	16	250
13	토양 지하수 [세계 토양, 토양 환경, 기념 개최]	13	220
14	4대강 사업 [조사 평가, 상류 지역, 조사 결과]	13	3830
15	수돗물 생산 [수돗물 공급, 모의 훈련, 세계보건기구 WHO]	12	940
16	비점 오염 [비점오염원 관리, 지도 점검, 시설 관리]	12	1500
17	낙동강 수계 [한강 수계, 금강 수계, 한강 낙동강]	12	900
18	가뭄 발생 [전국 강수량, 경기 충남, 충남 서부]	12	320
19	상수원 보호구역 [배출 시설, 관리 실태, 관리 지역]	11	2180
20	수도 계량기 [계량기 동파, 주민 불편, 동파 사고]	11	2120
21	해수담수화 플랜트 [해수담수화 기술, 스마트 물관리, 해수담수화 사업]	11	580
22	소규모 환경영향평가 [평가 대상, 계획 환경영향평가, 협의 절차]	10	210
23	군남 홍수조절지 [운영 관리, 임진강 유역, 임진강 건설]	10	350
24	수생태계 보전 [보전 법률, 유역 지방, 복원 계획]	10	100
25	국내 물산업 [해외진출 지원, 성장 동력, 경쟁력 확보]	10	250
26	물산업 기술 [기술 발전, 발전 협의, 사업 해외진출]	10	90
27	상수원 수질 [특별 대책, 특대 지역, 폐수 이용]	9	700
28	물순환 개선 [빗물 유출, 업무 협약, 빗물 침투]	9	110
29	가뭄 주의단계 [평년 예상, 전망 가뭄, 전망 생활]	9	70
30	용수공급 조정 [조정 기준, 주민 불편, 부족 대비]	8	80

Figure A2-2. Comparison graph of frequent keywords on press release and online news articles in 2016

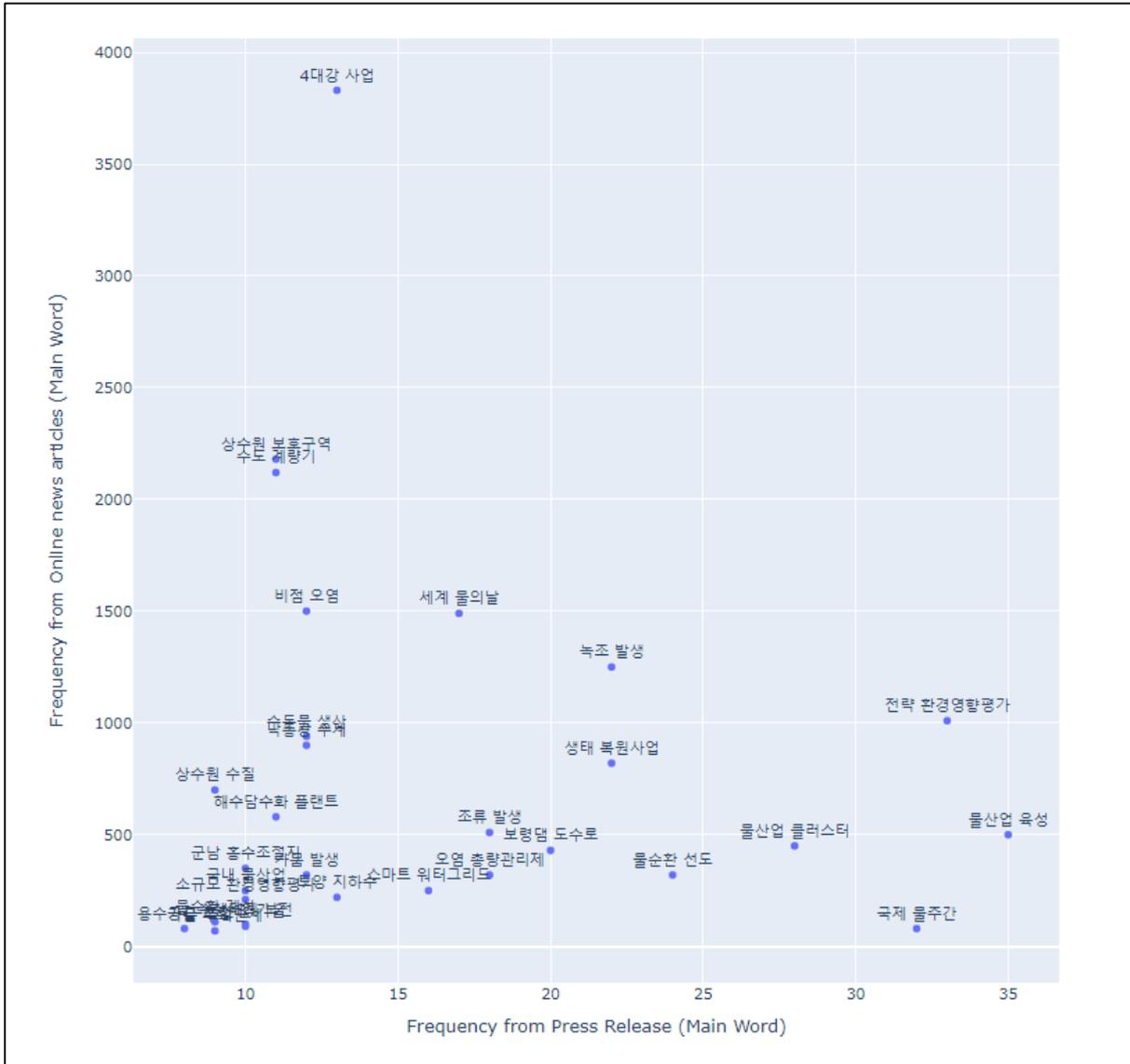


Figure A2-3. Analysis result of keywords on water management issues in 2017

	2017 Main Word	Related Words	Frequency from Press Release (Main Word)	Frequency from Online news articles (Main Word)
1	가뭄 대비	[강수량 부족, 추가 대책, 공급 지장]	46	1960
2	가뭄 대책	[가뭄 발생, 국민 안전, 가뭄 대응]	43	4000
3	용수 공급	[강우 부족, 공급 위주, 정상 공급]	38	4000
4	대체 공급	[급수체계 조정, 급수 지역, 보령댐 저수량]	35	420
5	가뭄 경보	[용수원 개발, 충남 지역, 가뭄 해소]	33	70
6	보령댐 도수로	[도수로 가동, 단계 관리, 하천유지용수 감량]	31	530
7	LID 기법	[빗물 관리, 개발 LID, 도시 물순환]	30	100
8	전국 다목적댐	[공업용수 공급, 강수량 부족, 공급 지장]	28	390
9	국제 물주간	[대한민국 국제, 해결 방안, 컨벤션 센터]	27	140
10	미세 플라스틱	[실태 조사, 물질 함유, 국제 기구]	27	520
11	수생태계 보전	[보전 법률, 실태 점검, 설치 운영]	26	160
12	관정 개발	[용수원 개발, 추진 가뭄, 저수지 준설]	23	2890
13	비점 오염	[보전 법률, 도시 물순환, 오염 시설]	23	1550
14	지방상수도 현대화	[수돗물 공급, 선도 사업, 수돗물 생산]	23	1320
15	가뭄 지역	[가뭄 피해, 저수지 준설, 대응 방안]	23	1780
16	용수 비축	[단계 도달, 심각 단계, 단계 관리]	22	160
17	상수도 보급	[수돗물 공급, 농어촌 지역, 농촌 지역]	21	510
18	농업 저수지	[가뭄 전망, 저수지 경영, 가뭄 전국]	20	1610
19	소규모 수도시설	[지하수 관정, 방사성 물질, 자연 방사성]	19	650
20	용수 부족	[긴축 운영, 용수공급 조정, 대비 용수공급]	18	2630
21	수생태계 건강	[물환경 관리, 하천 호소, 환경 상태]	18	200
22	상습 가뭄	[우려 지역, 부족 지역, 발생 지역]	18	550
23	개방 수위	[수위 도달, 금강 공주보, 목표 수위]	17	380
24	토양 지하수	[환경 관리, 국가 부지, 토양 환경보전]	17	230
25	지역 가뭄	[가뭄 전망, 경기 강원, 충북 충남]	17	1140
26	녹조 현상	[낙동강 수계, 녹조 발생, 원격 모니터링]	17	1190
27	농업용수 공급	[공급 문제, 개방 모니터링, 개방 대상]	16	3280
28	전국 가뭄	[가뭄 완화, 전국 강수량, 수준 전국]	16	140
29	광역상수도 사업	[대청 단계, 대산 임해, 단계 광역상수도]	16	470
30	다목적댐 저수량	[수계 다목적댐, 성진강 수계, 한강 금강]	15	60

Figure A2-4. Comparison graph of frequent keywords on press release and online news articles in 2017

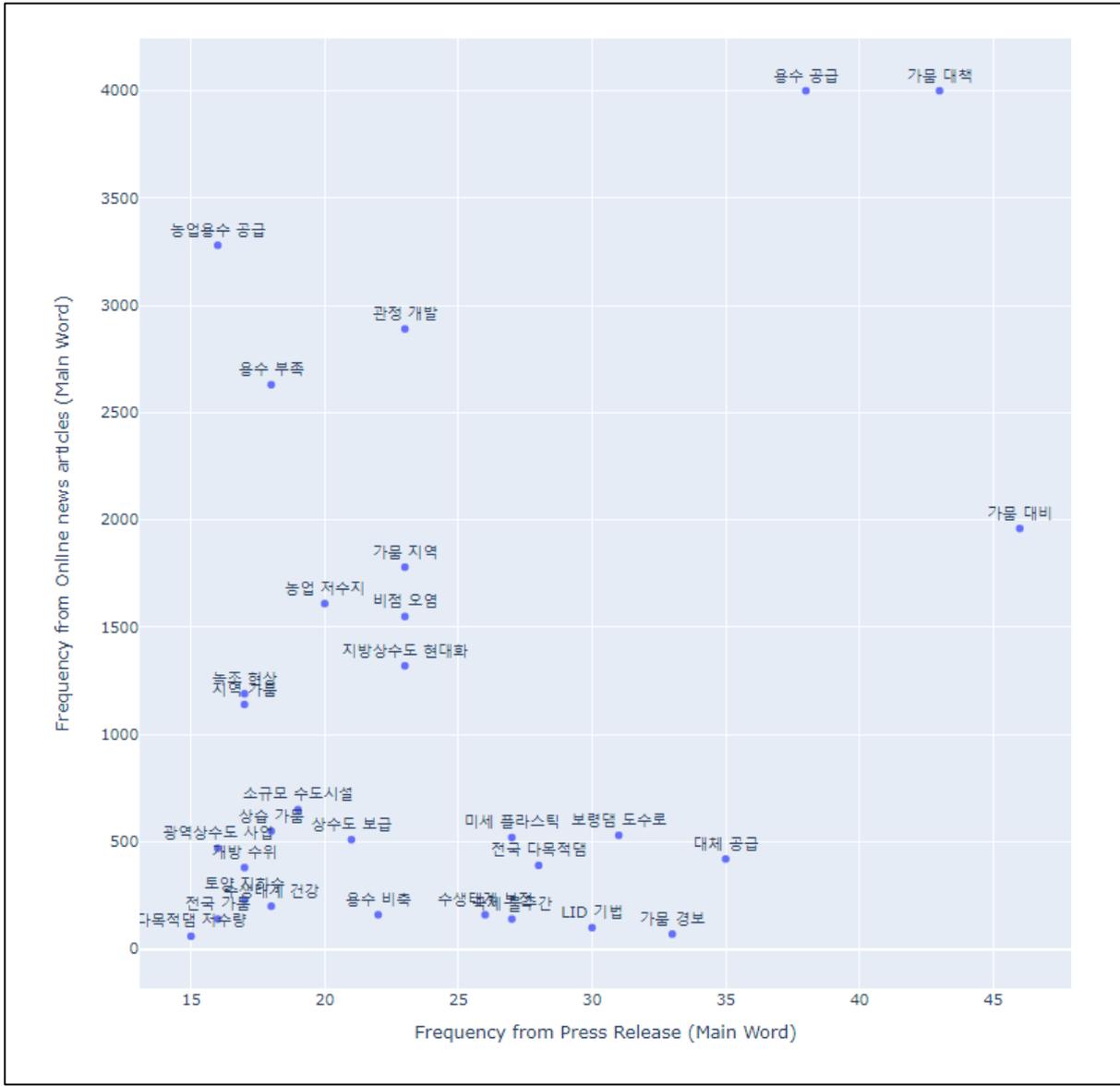


Figure A2-5. Analysis result of keywords on water management issues in 2018

2018 Main Word	Related Words	Frequency from Press Release (Main Word)	Frequency from Online news articles (Main Word)
1 유해 남조류	[분석 결과, 남조류 세포수, 조류 경보제]	54	780
2 물관리 일원화	[수량 수질, 국제 물주간, 통합 물관리]	53	1430
3 조류 경보	[경보 발령, 창녕 합안, 녹조 남조류]	52	1490
4 과용화 확립됨	[조사 결과, 조사 실시, 권고 기준]	37	1570
5 녹조 발생	[분석 결과, 조류 경보제, 영양 염류]	36	1400
6 자연성 회복	[4대강 자연성, 국가물관리위원회, 4대강 조사]	30	690
7 낙동강 수계	[조사 결과, 감시 기준, 수질 감시]	29	1400
8 개방 모니터링	[개방 계획, 개방 영향, 4대강 사업]	27	170
9 지하수 오염	[시설 설치, 유발 시설, 오염 유발]	25	1080
10 녹조 대응	[수돗물 공급, 여름철 녹조, 대응 관리]	21	270
11 개방 기간	[수위 저하, 지하 수위, 수위 회복]	21	610
12 토양 지하수	[조사 실시, 시설 설치, 토양 환경]	19	140
13 조류 독소	[조류 발령, 냄새 물질, 검사 결과]	19	180
14 물관리 기술	[물산업 진흥, 물관리 기본, 공모 시행]	19	780
15 대체 공급	[용수 비축, 시설 건설, 계통 광역상수도]	18	330
16 용수 부족	[부족 대비, 용수 비축, 가뭄 대비]	17	1530
17 한강 낙동강	[영산강 수계, 4대강 사업, 처리 계획]	16	50
18 금강 수계	[세종보 공주보, 수위 EL, 최저 수위]	16	910
19 지하수 이용	[수위 개방, 개방 수위, 수막 재배]	16	490
20 수도용 제품	[관리 강화, 제품 위생, 시행 개정안]	16	50
21 비점 오염	[비점오염원 관리, 비점 오염물질, 지도 점검]	14	1520
22 세계 물의날	[가뭄 홍수, 물의날 기념, 고양시 킨텍스]	14	1610
23 유속 증가	[유량 유속, 개방 세종보, 합천창녕보 창녕합안보]	13	60
24 협치 거버넌스	[핵심 전략, 거버넌스 구축, 유역 순회]	13	90
25 가뭄 대책	[가뭄 정보, 전국 다목적댐, 가뭄 대응]	13	2180
26 부산 에코델타시티	[세종 생활권, 기본 구상, 사업 시행]	13	1210
27 강우 레이더	[홍수 대응, 체계 확립, 도시 침수]	12	60
28 지방상수도 현대화	[용수 확보, 상수 가뭄, 지방상수도 확충]	12	1330
29 수생태계 건강	[건강 평가, 동물 부착, 현상 조사]	12	210
30 생태 공간	[양수장 가뭄, 서식 환경, 이프로 개방]	12	1050

Figure A2-6. Comparison graph of frequent keywords on press release and online news articles in 2018

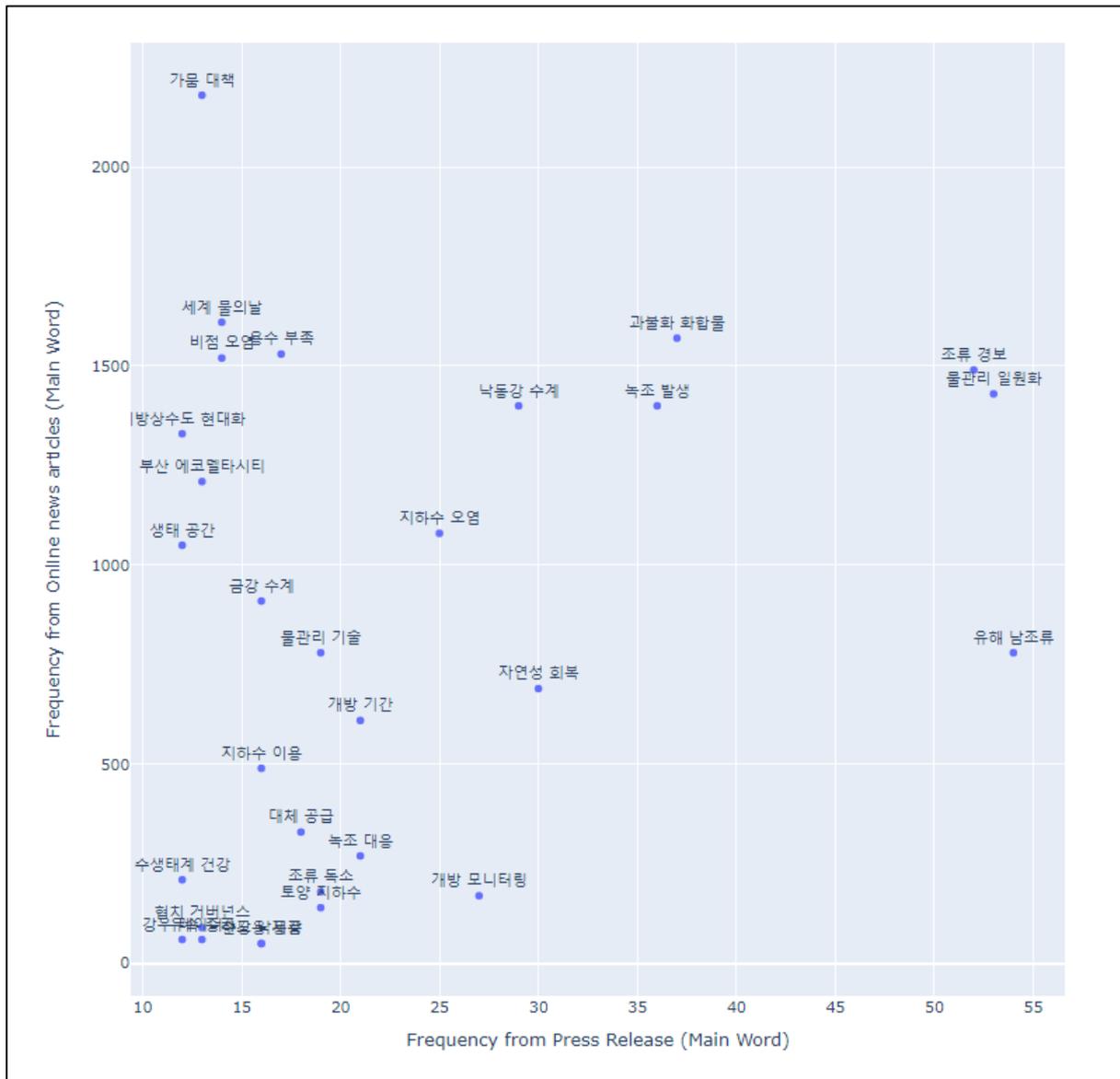


Figure A2-7. Analysis result of keywords on water management issues in 2019

	2019 Main Word	Related Words	Frequency from Press Release (Main Word)	Frequency from Online news articles (Main Word)
1	금강 영산강	[4대강 조사, 전문 위원회, 추가 모니터링]	79	50
2	물산업 클러스터	[국가 물산업, 실증 시설, 기업 지원]	70	400
3	녹조 발생	[유해 남조류, 경보 발령, 조류 경보]	65	1190
4	국가물관리위원회	[의견 수렴, 분쟁 조정, 법정 계획]	56	1530
5	유역물관리위원회	[국가 유역, 분쟁 조정, 유역 위원회]	40	620
6	자연성 회복	[방안 제시안, 이용 대책, 물관리위원회 보고]	39	1260
7	국가물관리	[물관리 기본계획, 기본계획 수립, 유역물관리]	39	340
8	수질 상태	[상태 개선, 종합 고령, 개선 효과]	37	60
9	인천 수돗물	[수질검사 결과, 수돗물 정상화, 급수 구역]	33	1110
10	물관리 기본	[분쟁 조정, 기본 시행, 유역물관리]	33	280
11	수생태계 건강	[평가 결과, 하천 수생태계, 건강 평가]	31	270
12	수문 개방	[실증 실험, 바닷물 유입, 낙동강 하굿둑]	30	1280
13	4대강 사업	[개선 효과, 수질개선 사업, 경제성 평가]	29	3780
14	수질 개선	[개방 수질, 수질개선 사업, 수질 악화]	28	4000
15	비점 오염	[오염 시설, 성능 검사, 시설 성능]	27	2300
16	물환경 보전	[폐수 배출, 배출 허용, 허용 기준]	26	90
17	개방 기간	[개방 수질, 수질 악화, 세종보 공주보]	26	1300
18	수돗물 공급	[수계 전환, 사고 발생, 사고 대응]	23	4000
19	낙동강 상류	[하류 지역, 중금속 오염, 현장 점검]	22	870
20	관정 개발	[지하수 대책, 대체 관정, 개방 지하수]	21	1220
21	녹조 대응	[유관 기관, 여름철 녹조, 국민 안심]	21	340
22	수돗물 수질	[수질 분석, 수돗물 안심지원단, 정상화 원반]	21	1920
23	강우 레이더	[유관 기관, 통합 관리, 레이더 관측]	21	170
24	지하 수위	[개방 지하수, 지하수 이용, 이용 장애]	21	550
25	한강 낙동강	[낙동강 금강, 홍수 통제, 유역 위원회]	20	90
26	물산업 진흥	[물관리 기술, 기술 인증, 기술 발전]	20	310
27	미세 플라스틱	[정보 공유, 협력 방안, 분석 방법]	19	4000
28	물관리 일원화	[기본 시행, 성과 중점, 계획 물관리]	18	730
29	안동댐 상류	[정밀 조사, 영풍 제련소, 중금속 오염]	18	170
30	병입 수돗물	[학교 급식, 급식 지원, 불만 해소]	18	510

Figure A2-8. Comparison graph of frequent keywords on press release and online news articles in 2019

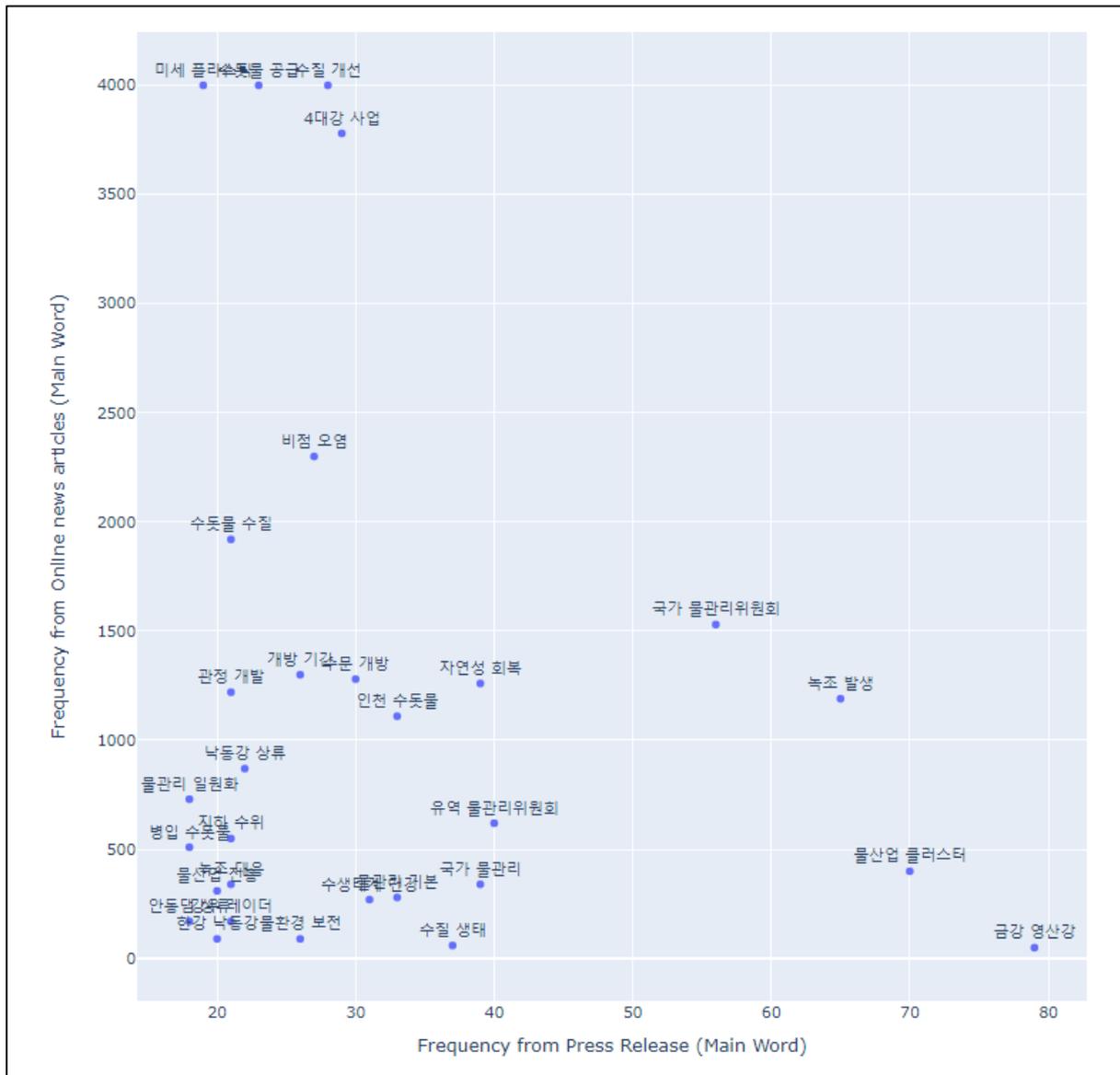


Figure A2-9. Analysis result of keywords on water management issues in 2020

2020 Main Word	Related Words	Frequency from Press Release (Main Word)	Frequency from Online news articles (Main Word)
1 자연성 회복	[4대강 자연성, 4대강 조사, 회복 구상]	81	1100
2 물산업 클러스터	[국가 물산업, 클러스터 기업, 클러스터 사업]	72	80
3 수열 에너지	[친환경 수열, 복합 클러스터, 에너지 복합]	65	1790
4 수자원 위성	[위성 개발, 위성 활용, 홍수 가뭄]	51	120
5 녹조 발생	[유해 남조류, 조류 경보제, 물금 매리]	40	930
6 강우 레이더	[홍수 특보, 하천 수위, 물발 홍수]	38	250
7 부유 쓰레기	[쓰레기 수거, 전국 하천, 정화 활동]	37	1210
8 홍수 피해	[피해 지역, 피해 복구, 방지 대책]	36	4000
9 조류 경보	[유해 남조류, 분석 결과, 조류 경보제]	35	740
10 그린 뉴딜	[개발 사업, 활성화 방안, 계획 수자원]	34	4000
11 수돗물 유출	[유출 발생, 유출 발견, 위성 관리]	33	4000
12 수돗물 공급	[위생 관리, 관리 종합대책, 국민 안심]	32	4000
13 국가물관리위원회	[유역 물관리위원회, 처리 방안, 금강 유역]	32	780
14 홍수 대응	[홍수 대책, 홍수 관리, 홍수 특보]	29	920
15 도시 물순환	[전문가 중심, 전문가 개최, 지표 개발]	28	220
16 상수도 관망	[관리 강화, 시설 운영, 관망 관리]	27	740
17 국가물관리	[물관리 기본계획, 환경 평가, 국민 제감]	25	170
18 물관리 일원화	[물관리 기본계획, 국민 제감, 국민 소통]	25	710
19 유역 수도지원센터	[사고 대응, 현장 수습, 수습 정관]	24	390
20 노후 상수도	[상수도 시설, 상수도 관리체계, 상수도 정비사업]	24	1940
21 물환경 보전	[정기 검사, 과태료 부과, 시행 규칙]	23	80
22 섬진강 유역	[영산강 섬진강, 처리 방안, 의견 청취]	23	1470
23 유출 지하수	[지하수 활용, 모범 사례, 공동 주택]	22	210
24 비점 오염	[설치 운영, 성능 검사, 오염 시설]	22	2170
25 낙동강 유역	[낙동강 통합물관리, 상류 지역, 통합물관리 방안]	21	3780
26 홍수 통제	[홍수 정보, 피해 예방, 관리 대책]	21	4000
27 한강 수계	[홍수조절 용량, 홍수 대비, 다목적댐 운영]	21	1990
28 스마트 상수도	[상수도 관리체계, 공급 과정, 구축 사업]	20	960
29 인공지능 AI	[위기 대응, 안전 관리체계, 스마트 안전]	20	880
30 하천 사용	[시행 개정안, 시행 개정, 사용 허가]	19	50

