

BIG DATA ANALYSIS PLATFORM TEXTOM

TEXTOM MANUAL

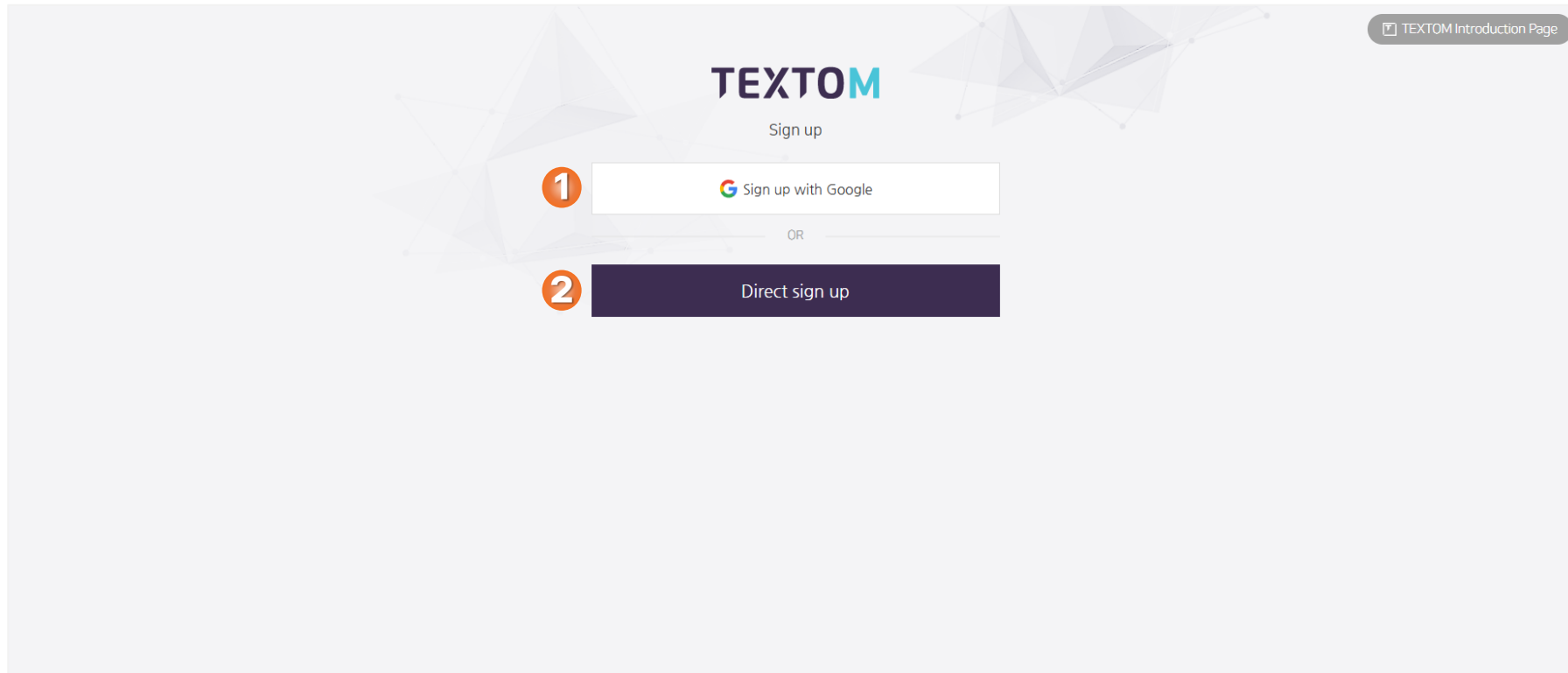


TEXTOM

CONTENTS

Click the table of contents to go to the page.
Review the manual before use for easier navigation of TEXTOM

INTRO. Table of contents	→ Click
Sign-up & Log in	→ Click
Dashboard	→ Click
Collection	→ Click
Refinement	→ Click
Analysis	→ Click
Visualization	→ Click
My page	→ Click



- ① **Sign up with Google**
Easily sign up using your Google account.
- ② **Direct Sign up**
Sign up by entering your personal information.

TEXTOM

You need to login to access TEXTOM.

1

KOR ENG CHN EDU

ID

Please enter your ID

Password


Please enter your Password

Save ID

2

Login

OR

 Sign in with Google

[Find ID/Password](#) → [Sign up](#) →

TEXTOM is optimized for the latest version of Chrome, 1280*1024 or higher resolution.

- ① **Select Service**
Log in to the desired service :
TEXTOM, TEXTOM Global, TEXTOM
China or TEXTOM Edu.
Tip. Existing members can use email
authentication to access all three
services (Korean, English, and
Chinese).
Tip. To log in to TEXTOM Edu, you
must first have a TEXTOM
account.
- ② **Log in with Google**
You can easily Log in using your
Google account.

TEXTOM

Please write down the information you wrote when you signed up for a member.

1 Find ID **2** Reset password

Name

Please enter your name

Email

Please enter your email

Find ID

TEXTOM

Reset password

The password has been initialized to be the same as the ID.
The initialized password is [redacted]

3 Go to login page

- ① Find ID**
You can find your TEXTOM ID by name and email.
- ② Reset password**
You can reset your existing password by entering your ID, name and email address.
- ③ Reset new password**
You can reset your password to a new one on the [My Page]-Edit Member Information page.
Tip. The reset password will be the same as your current TEXTOM ID.

TEXTOM Dashboard 백데이터팀 ENG LOG OUT

1 Total Collection/Analysis status Based on the past month (Completed / In Progress)

24 / 24 Portal Data 2 / 2 Upload Data 4 / 4 Refinement List

2 Collection List Only the last 5 collection lists in progress are shown. More details >

Data Name	Collection Type	Collection Status	Collection Date
reddit_test.txt	Upload Data	Done	24.05.29
test.xlsx	Upload Data	Done	24.05.29
[REDACTED]	Portal Data	Done	24.05.29
"e-learning trends"	Portal Data	Done	24.05.29
"text mining techniques"	Portal Data	Done	24.05.29

3 Refinement List Viewing the last 5 refined lists registered. More details >

Data Name	Version	Modified Words	Refinement Status	Refinement Date
"text mining techniques"	1	0Cases	Done	24.05.30
"e-learning trends"	1	0Cases	Done	24.05.30
re-[REDACTED]	1	0Cases	Done	24.05.29
test.xlsx	1	0Cases	Done	24.05.29

4 백데이터팀 User View Capacity Details >

Remaining Capacity: 102.16 MB / 0.00 KB

5 Notice More details >

- Notice ★Service Suspension Notice for Upgrade (April 29th t... 24.03.12
- Notice Collect more than 1,000 data on each channel of the ... 24.02.07
- Notice Guide to changes in permission for collection units 24.01.31
- Notice Is there a way not to repeat the same tablet multiple ti... 24.01.31
- Notice All information regarding the price of TEXTOM 24.01.30

FAQ: Frequently Asked Questions Q&A: Feedback and Improvement Suggestions

채팅문의

① **Total Collection/Analysis status**
The current status of data collection and refinement can be checked for each channel : Portal, Upload Data.
Tip. Clicking the icon will take you to that page.

② **Collection Progress List**
Check the collection progress of the last five registered data.
Tip. Clicking [More details] button will take you to that page.

③ **Data Preprocessing List**
Check the refinement progress of the last five registered data.
Tip. Clicking [More details] button will take you to that page.

④ **Remaining Capacity**
Check the remaining data.
Tap [Capacity details] to go 'My page'.

⑤ **Notice**
View service information such as announcements, manuals, and utilization papers.

The screenshot shows the 'Data Collection' page in the TEXTOM interface. The page has a dark sidebar on the left with navigation options: ENG, KOR, CHN, Add Capacity, Dashboard, Collection (selected), Data Collection, Collection List, Cleansing, Data preprocessing, Refinement List, Analysis, and Customizing. The main content area is titled 'Data Collection' and has three tabs: 'Portal/Channel' (selected), 'Upload Data', and 'Request Channel'. At the top right, there are user and language settings (백데이터팀, ENG) and a 'LOG OUT' button.

Four numbered callouts highlight key features:

- 1 Collected keywords:** A text input field with a '+' button and an 'Operators' dropdown. Below it, a tip states: 'Using the [add keyword](#) feature, you can create multiple [collection lists](#) at once (with the same collection conditions but different lists of keywords).'.
- 2 Period:** Two date pickers showing '2024-05-22' and '2024-05-29'. A tip below reads: 'The collection period can be set for up to 3 years. * If you require long-term, large-scale collection, please request it through the appropriate channels. It is not possible to set the collection period for YouTube, Yahoo, Quora and Reddit.'
- 3 Channel:** A section with logos for Google, YouTube, yahoo!, Quora, and reddit, each with a checkbox. Under Google, there are sub-options: 'Google Overall', 'News', 'Google Facebook', and 'Web documents'. A note says: 'Google Facebook refers to Facebook posts whose results are collected by Google.'
- 4 Create collection list >>** A large blue button at the bottom of the form.

At the bottom right of the page, there is a purple button labeled '채팅문의' (Chatting).

- ① **Collect Keywords**
Enter the keywords you want to collect. Use the [+], [-] buttons to create keyword sets and **search operators** to improve data quality.
Tip. Instead of broad keywords like "economy" and "tourism," use specific ones like "California economy" and "Napa Valley tourism" for better data quality and reduced usage.
- ② **Collection Period**
Set the desired data collection period. Yahoo, Quora, and Reddit do not support setting a collection period.
Tip. Set the collection period for up to 3 years. For longer or larger collections, please request through 'Request Channel'.
- ③ **Channel**
Select the collection portal/channel.
Tip. If you select Google channel, you cannot collect data alongside other channels. For Yahoo, Quora, and Reddit, please set up separate collection requests.
- ④ **Run the data collection according to the selected options.**

TEXTOM Data Collection

ENG KOR CHN

Add Capacity

Dashboard

Collection

Data Collection

Collection List

Cleansing

Data preprocessing

Refinement List

Analysis

Customizing

Information Page Site

백데이터팀 ENG LOG OUT

Portal/Channel Upload Data Request Channel

1 Data Upload

Only `.txt`, `.xlsx`, and `.csv` files can be uploaded.
You can select up to 5 files at a time.
※ Register as a separate collection list.

For spreadsheet files, you can upload data that matches the specified **columns** to utilize additional analysis features such as time series analysis.

※ A Column - Title, B Column - Body Text, C Column - Channel, D Column - Date
※ Upload File Example

	Title Column	Body Text Column	Channel Column	Date Column
	A	B	C	D
1	Title	Body Text	Channel	Date
2	Micron upgraded to ...	Morgan Stanley is ...	news	2024-05-24
3	Veteran or outdated? ...	Samsung Electronics ...	news	2024-05-24

If the spreadsheet data does not match the above format, please enter the data to be analyzed in **column A**.

2

3 Create collection list >>

채팅문의

- ① **Data Upload**
Upload your own data. Click the [+]
button to select up to 5 files.
Tip. To use additional analysis features
like time series analysis, upload
Excel files in the sample file format.
Otherwise, enter the data for
analysis in Column A.

- ② **Sample Format**
Click the button to download the
sample format. Please refer to it and
upload your data in the specified
format.

- ③ Run the data collection according to
the selected options.

TEXTOM

Add Capacity

Dashboard

Collection ▾

Data Collection

Collection List

Cleansing ▾

Data preprocessing

Refinement List

Analysis

Customizing

Information Page Site

Data Collection

빅데이터팀 ▾ ENG ▾ LOG OUT

Portal/Channel

Upload Data

Request Channel

Feel free to request any channels and items you'd like special collection for!
Please request the items you desire, and we'll promptly review them and provide you with information on collection costs and time frames.

SNS(X)

YOUTUBE

NAVER News/Comments

Smart store Review

Cafe, Dcinside

Channel Description

Channel URL

Collection Description

Period ~

Request Channel Inquiry >>

채팅문의

Data Collection Request

If you want to collect data outside of the channels provided by TEXTOM, please register as a request channel.

- Based on the contents of the request, we will check whether the data can be collected from the requested site and provide a quotation according to the calculation method.
- The collection method through the request channel is to receive only the collected data or to build it within the request channel page.
- The collected data is delivered as an EXCEL(.xlsx) file or txt file.
- Once built, you can use it in the same way as the existing channels provided in the collection service.

TEXTOM Collection List

Add Capacity

Dashboard

Collection

Data Collection

Collection List

Cleansing

Data preprocessing

Refinement List

Analysis

Customizing

Information Page Site

Portal/Channel Upload Data

Keyword Search

Collected data will be deleted after 30 days.

1 cases / Total 3 cases

Delete

	Keyword	Channel	Period	Creation Date	Capacity	Collection Status
<input type="checkbox"/>	"text mining techniques"	google(news,web)	2024-01-01~2024-05-29	2024-05-29	84.48 KB	Done
<input type="checkbox"/>	"e-learning trends"	youtube yahoo quora	2024-01-01~2024-05-29	2024-05-29	91.88 KB	Collecting
<input checked="" type="checkbox"/>	"e-learning trends"	google(news,web)	2024-01-01~2024-05-29	2024-05-29	86.90 KB	Done

Select Data

"e-learning trends"
2024-01-01 ~ 2024-05-29

Channel	Section	Collection Quantity (records)	Capacity	Original Text
Google	Web	216	56.27 KB	Preview
Google	News	106	30.63 KB	Preview

Preprocess Data»

(Estimated Deduction Size : 86.90 KB)

Please check remaining capacity!

Remaining Capacity : 108.45 MB

Add Capacity

채팅문의

- You can search using keywords.
- View Collected data**
Shows the list of collected data. Click the checkbox to view the information of the corresponding data in the bottom section.
- Delete Data**
Select the checkbox and click the [Delete] button to remove the data from the collection list.
Tip. Data will be automatically deleted 30 days after collection. It is recommended to download the data that the user wants to keep.
- Check Data Information**
Clicking on the data will show the channel, section, and original text in the preview area.
Tip. Previewing the original text helps in filtering spam keywords during the data processing stage.
- Process the data.
- You can check the estimated user usage and residual usage and add usage.

TEXTOM Data preprocessing

1 Select Column All Title Body text
Please select the column for text analysis.

2 Morpheme analysis

3 Analytical language

4 Analyzer
It provides morpheme analysis results based on the CoreNLP library developed by Stanford University.

5 Analysis Part of speech Noun Verb Adjective Foreign Language Number

6 User Dictionary

7

Preprocessing data

"e-learning trends"
2024-01-01 ~ 2024-05-29

Preprocessing history

<input type="checkbox"/>	Function name	Detail History
<input type="button" value="Preprocess Data >>"/>		

- ① **Select Analysis Column**
Select the column to analyze.
Tip. Analyze "title" for quick trends;
analyze entire text for in-depth insights.
- ② **Morphological Analysis**
If you are uploading already processed data or analyzing the raw text as it is, select [Not selected].
- ③ Select a language for analysis.
- ④ TEXTOM provides morphological analysis results based on the CoreNLP library created by Stanford University.
- ⑤ Select Part-of-Speech(POS) tagging criteria.
Tip. Verbs : interaction between subjects
Adjectives : sentiment analysis
Numbers : proper noun differentiation
- ⑥ **User Dictionary**
Use the user dictionary for efficient repeated analysis of specific topics or keywords.
Tip. Register keywords to refine in the user dictionary first. (My Page -> User Dictionary)
- ⑦ Click the [Apply] button to save the selected options in the preprocessing history.

TEXTOM Data preprocessing

Morpheme analysis | **Data Preprocessing**

1 Data Preprocessing Not selected

2 Duplicate deleted Not selected
 Select Column All Title Body text

3 Spam filtering Not selected
 Select Column All Title Body text
 Keyword Please enter the keywords to filter.

Filter data based on the selected column to include or exclude specific words.

- Include keyword: Only use data containing the specified word for analysis.
- Exclude keyword: Exclude data containing the specified word from analysis.
- AND: Filter data to meet 'all' the specified include and exclude keyword conditions.
- OR: Filter data to meet 'any' of the specified include and exclude keyword conditions.

4

Preprocessing data

5

Preprocessing history

<input type="checkbox"/>	Function name	Detail History
<input type="checkbox"/>	Select Column	title, contents
<input type="checkbox"/>	Analytical language	en
<input type="checkbox"/>	Analyzer	CoreNLP
<input type="checkbox"/>	Analytical history	noun, adjective
<input type="checkbox"/>	User Dictionary	Not selected

7

- ① **Data Cleansing Option Selection**
You can use cleansing features such as duplicate removal and spam filtering.
Tip. It is recommended to use these features to reduce data volume and enhance analysis efficiency.
- ② **Duplicate Removal**
Remove duplicate data based on the selected column(s). When multiple conditions are set, identical data within each column will be removed.
- ③ **Spam Filtering**
Filter data using keywords to include or exclude records based on the selected column.
- ④ Click the [Apply] button to save the selected options in the preprocessing history.
- ⑤ **Data Preview**
Preview the data to be analyzed.
Tip. Reviewing the original text beforehand aids in spam keyword filtering and helps reduce data usage.
- ⑥ Check the box and click [Delete] to remove the analysis options.
- ⑦ Proceed with data preprocessing with the selected analysis option.

Cleansing Data

Check and manage the keyword refinement status.

- <13> -

TEXTOM Refinement List

Keyword Refining List

0% / Total 1 cases

1 [Edit] "e-learning trends"

2 Keyword Refining List

3 Edit

4 Download

5

6 Refining Upload

7 Delete

8 Refresh

Preprocessing information

Collection information

Keyword	"e-learning trends"
Collection channel	google(news.web)
Collection Unit	Month unit
Collection Capacity	85.90 KB

Morpheme analysis

Analyzer	CoreNLP
Analysis Part of speech	noun, adjective
User Dictionary	-

no	title	body	channel	sub-channel	start date	end date	url	refined data
1	Online vs	https://ww.google	news		2024-01-0	2024-01-3	This study person high education e	
2	10 more p	https://ecc.google	news		2024-01-0	2024-01-3	From AI to more prediction retail m	
3	Creatorpr	https://ww.google	news		2024-01-0	2024-01-3	A recent i creatorpreneur upcomin	
4	4.7 Best Ed.	https://ww.google	news		2024-01-0	2024-01-3	Education good education stock	
5	5 MOOCs	h/https://ww.google	news		2024-01-0	2024-01-3	PRNewswi mooc market used north	
6	How to Cl	https://ww.google	news		2024-01-0	2024-01-3	This complr software system comp	
7	7 50 Creativ	https://sm.google	news		2024-01-0	2024-01-3	Creativity creative business idea	
8	8 The 5 Top	https://sm.google	news		2024-01-0	2024-01-3	As techno top virtual assistant servi	
9	9 55 Amazir	https://sm.google	news		2024-01-0	2024-01-3	There are amazing home business	
10	10 The Top eLearning	T google	web		2024-01-0	2024-01-3	https://ww.top elearning trend learr	
11	11 11 E-Learn	11 E-Learn	google		2024-01-0	2024-01-3	https://ww.e learning trend predicti	
12	12 Discover t	The world	google		2024-01-0	2024-01-3	https://ww.e trend edunext world e	
13	13 Trends in	Additional	google		2024-01-0	2024-01-3	https://ww.trend management syste	
14	14 Trends & I	February 1	google		2024-01-0	2024-01-3	https://ele.trend forecast archive cr	
15	15 Category: The	evolu	google		2024-01-0	2024-01-3	https://kai.category online educatio	
16	16 Blended L	1:11 - The	google		2024-01-0	2024-01-3	https://mj.blended learning future	
17	17 The Scienc	Visit our v	google		2024-01-0	2024-01-3	https://ww.science success psycholo	
18	18 Personaliz	For more i	google		2024-01-0	2024-01-3	https://ww.learning balance more ir	
19	19 Selecting	How AI &	google		2024-01-0	2024-01-3	https://ww.ideal firm business ml so	
20	20 Full Stack	7. Continu	google		2024-01-0	2024-01-3	https://getfull stack developer platf	
21	21 Hire the b	Constantly	google		2024-01-0	2024-01-3	https://ww.good ux designer serbia	
22	22 Look 1 un	Check out	google		2024-01-0	2024-01-3	https://ww.check fact e learning tre	
23	23 Here's wh	Check out	google		2024-01-0	2024-01-3	https://ww.target fact e learning tre	
24	24 Graph &	(https://tov	google		2024-01-0	2024-01-3	The post i graph geometric ml next	
25	25 AI trends	https://sec	google		2024-01-0	2024-01-3	With the trend close look machin	
26	26 Learning i	https://ww	google		2024-01-0	2024-01-3	As learnin digital age	
27	27 What the	https://ecc	google		2024-01-0	2024-01-3	Here are stend deck ai	
28	28 A Positive	https://ww	google		2024-01-0	2024-01-3	As faculty, positive environment enc	

Copy Download

채팅문의

- Click the [Edit] button to modify the data name.
- Keyword Refinement List**
You can check the history of keyword editing.
- Real-Time Edit**
Correct keywords directly on the web in real time without uploading a file.
- Data Download**
Download an Excel (.xlsx) file with raw data, refined data, dates, and related information.
- Leave a brief note on the refinement version and history.
- Upload Refined File**
You can upload a new version by downloading the file in ④ and refining it yourself.
Tip: Ensure the file format matches the one downloaded in ④.
- Delete Refinement History**
Select the refinement history to delete using the checkbox, then click [Delete].
- Refresh**
Checks the refinement status in real time.

TEXTOM Refinement List

Keyword Refining List

0/1 Total 1 cases

	Data Edit	Version	Download	Refinement Date	Modified Words	User Dictionary	Memo
<input type="checkbox"/>	Edit	1	Download	2024-05-29 15:32:28	0Cases	Not specified	

1 Preprocessing information

2 Copy Download

Collection information

Keyword	"e-learning trends"		
Collection channel	google(news.web)	Collection Period	2024-01-01 ~ 2024-05-29
Collection Unit	Month unit	Collection Range	Title, Body Text, URL
Collection Capacity	86.90 KB	Collection Date	2024-05-29

Morpheme analysis

Analyzer	CoreNLP
Analysis Part of speech	noun, adjective
User Dictionary	-

Data Preprocessing

Duplicate deleted	Apply(title, contents)
Spam filtering	-

3 Keyword Search

Total 3 cases

Data Name	Refining Date	Version	
test.xlsx	24-05-29	1	Delete
"text mining techniques"	24-05-29	1	Delete
"e-learning trends"	24-05-29	1	Delete

4 Analyze Data »

5

채팅문의

- ① **Processing Information**
View information on data collection and refinement.
- ② Copy or download the preprocessing information in table format.
- ③ **Keyword Search**
Search the refinement list by keyword.
- ④ **Refinement List Management**
Perform management tasks such as viewing and deleting the refinement list.
- ⑤ Analysis is performed on the selected data.

Cleansing Data

Refine keywords directly on web pages, without needing to upload data.

- <15> -

Keyword Refining

Original / Cleansing Data

1

Original Text

Cleansing

- results found / Total 322 results

Search Word

1	person high education effect student achievement recommendation leadership
2	more prediction retail media
3	creatorpreneur upcoming trend creator economy space
4	good education stock
5	moc market usd north america technavio
6	hr software system complete guide
7	creative business idea
8	top virtual assistant service
9	amazing home business idea
10	top elearning trend learning lab lm
11	e learning trend prediction e learning trend prediction mobile learning social learning inevitable rise immersive technology microlearning
12	e trend edunext world e learning promise year exciting new trend e trend
13	trend management system additional future e learning trend top reason moc management system algorithm
14	trend forecast archive craig weiss february e trend forecast surprise big surprise rcat negative impact
15	category online education kaideeducation evolution technology e learning forefront education way knowledge e trend
16	blended learning future e learning consultant elearning future learning ai consultant elearning e learning trend consultant

Analysis Results

2

TF

N-gram

- results found / Total 1017 results

Search Word

1	learning	322
2	trend	321
3	e	304
4	technology	55
5	education	54

3

Word substitution

Exact match, Case sensitive

Enter word directly

AI auto-refined words.

Change Remove

please enter the word to be changed.

+

Please enter the word to be modified.

4

Modification Details

User Dictionary

Delete

Refining Before

Refining After

5

+ Create new version

✓ Apply existing version

- 1 **Raw/Refined Data Tab**
Switch between the raw data tab and the refined data tab to view the data.
- 2 **View Analysis Results**
View TF (Term Frequency) and N-gram analysis results for keyword refinement.
- 3 **Edit Keywords**
Input and change keywords directly or use the [AI automatic refinement words] function.
- 4 **Edit History**
View and delete keyword edit history. Register the current edit history in the user dictionary or load an existing user dictionary.
- 5 **Create New Version**
Add the edit history to a new version.
- 6 **Apply Previous Version**
Apply the edit history to a previous refinement version.
Tip. Once applied, changes cannot be undone. Review edit history and download refined data in advance.

Cleansing Data

Refine your keywords based on the raw/refined data and the results of your analysis.

- <16> -

Keyword Refining

Original / Cleansing Data

1 Original Text

2 Cleansing

Search Word

- results found / Total 322 results

1	person high education effect student achievement recommendation leadership
2	more prediction retail media
3	creatorpreneur upcoming trend creator economy space
4	good education stock
5	mooc market usd north america technavio
6	hr software system complete guide
7	creative business idea
8	top virtual assistant service
9	amazing home business idea
10	top elearning trend learning lab lm
11	e learning trend prediction e learning trend prediction mobile learning social learning inevitable rise immersive technology microlearning
12	e trend edunext world e learning promise year exciting new trend e trend
13	trend management system additional future e learning trend top reason mooc management system algorithm
14	trend forecast archive craig weiss february e trend forecast surprise big surprise rcat negative impact
15	category online education kaideneducation evolution technology e learning forefront education way knowledge e trend
16	blended learning future e learning consultant elearning future learning ai consultant elearning e learning trend consultant

Analysis Results

3 TF

N-gram

Search Word

- results found / Total 1017 results

1	learning	322
2	trend	321
3	e	304
4	technology	55
5	education	54

TF

N-gram

Search Word

- results found / Total 2286 results

1	e	learning	
2	learning	trend	236
3	e	trend	189
4	future	e	48
5	late	e	21

Bigram

Trigram

Tetragram

Search Word

+ Create new version

✓ Apply existing version

① **Raw/Refined Data Tab**
Switch between the raw data tab and the refined data tab to view the data.

② **Keyword Search**
Analyze how specific keywords are used in the raw and refined data, and determine which words are frequently used together with the keywords through N-gram analysis. This can be useful for keyword refinement.
Tip. Use this functionality during the data refinement process to verify data connections, linked words, and the outcomes of applied refined keywords.

③ **View Analysis Result**
View TF (Term Frequency) and N-gram analysis results.
Tip. TF and N-gram analysis results can help adjust, remove, or identify relevant keywords effectively.

Cleansing Data

Refine keywords using various features.

- <17> -

Keyword Refining

Original / Cleansing Data

Original Text

results found / Total 322 results

1	person high education effect student achievement
2	more prediction retail media
3	creatorpreneur upcoming trend creator economy s
4	good education stock
5	moc market usd north america technavo
6	hr software system complete guide
7	creative business idea
8	top virtual assistant service
9	amazing home business idea
10	top elearning trend learning lab lm
11	e learning trend prediction e learning trend pred
12	e trend edunext world e learning promise year e
13	trend management system additional future e le
14	trend forecast archive craig weits february e tren
15	category online education kaideeducation evok
16	blended learning future e learning consultant ele

Word substitution Exact match, Case sensitive

Enter word directly AI auto-refined words.

Change Remove

please enter the word to be changed.

+ Remove data containing the word to be changed. ?

Word substitution

Enter word directly AI auto-refined words.

<input type="checkbox"/>	Refining Before	Refining After	TF
<input type="checkbox"/>	e learning	elearning	236
<input type="checkbox"/>	research first	researchfirst	11
<input type="checkbox"/>	renewable energy	renewableenergy	5
<input type="checkbox"/>	artificial intelligence	artificialintelligence	5
<input type="checkbox"/>	professor openweb	professoropenweb	4
<input type="checkbox"/>	gatlab com	gatlabcom	4

Analysis Results

TF

results found / Total 1017 results

1	learning	322
2	trend	321
3	e	304
4	technology	55
5	education	54

Word substitution

Enter word directly AI auto-refined words.

<input type="checkbox"/>	Refining Before	Refining After	TF
<input type="checkbox"/>	e learning	elearning	236
<input type="checkbox"/>	research first	researchfirst	11
<input type="checkbox"/>	renewable energy	renewableenergy	5
<input type="checkbox"/>	artificial intelligence	artificialintelligence	5
<input type="checkbox"/>	professor openweb	professoropenweb	4
<input type="checkbox"/>	gatlab com	gatlabcom	4

Modification Details

User Dictionary Delete

<input type="checkbox"/>	Refining Before	Refining After
<input type="checkbox"/>	smart city	smartcity
<input type="checkbox"/>	renewable energy	renewableenergy
<input type="checkbox"/>	privacy policy	privacypolicy
<input type="checkbox"/>	pricing value	pricingvalue
<input type="checkbox"/>	co pilot	copilot
<input type="checkbox"/>	issue alert	issuealert

Apply existing version

- Direct Word Refinement**
Directly input the word to be changed and the replacement word. If replacing multiple words with one, click the [+] button to add words.
- Keyword and Data Removal**
Remove specific words from the entire dataset. Add words to remove by clicking the [+] button. To remove data containing specific keywords, select the checkbox.
- AI Automated Refinement**
Calculate the probability of words appearing consecutively based on AI, recommending up to 200 word pairs.
- Apply the changes to the modification history by pressing the button.

Data Analysis

View the analysis list and check the details of the data.

- <18> -

TEXTOM Data Analysis

③ "text mining techniques"

④ **Collection Information** Copy Download

The word count is calculated based on the original text.

Cleansing Version	Morpheme Analyzer	Noun	Adjective	Verb	Foreign Language	Number	Others	Total
1	-	1144Cases	336Cases	0Cases	0Cases	0Cases	0Cases	1480Cases

⑤ **Analyze List**

The analyzed data will be deleted after 90 days.

Category	Detailed functions	Shortcut
Word Analysis	Word Analysis	Analyze
	N-gram	Analyze
	TF-IDF	Analyze
Matrix	Named Entity Recognition	Analyze
	Matrix 1-mode	Analyze
	Matrix 2-mode	Analyze
Network Analysis	Network Attribute	Analyze
	Centrality	Analyze
	Ego Network	Analyze
	Shortest path	Analyze
	CONCOR	Analyze


① Keyword Search

Total 3 cases

Data Name	Refining Date	Version	
"e-learning trends"	24-05-29	1	Delete
"text mining techniques"	24-05-29	1	Delete
test.xlsx	24-05-29	1	Delete

② Analyze Data >>

⑥ 채팅문의

- ① Search for keywords to view data containing those words.
- ② **Analysis List**
View and delete data that has been refined. Clicking on specific data displays its information in the left area. 
Tip. Click the [] button to hide the analysis list window.
- ③ **Modify Data Name**
Click the [Edit] button to easily change the data name to your preferred title.
- ④ **Analysis Information**
View the morphological analysis results of the data.
Tip. You can select the refined version and copy the analysis information or download it as an Excel (.xlsx) file.
- ⑤ **Analysis History**
Access the analysis results by clicking the "Analyze" button.
- ⑥ Navigate to the analysis results screen for the selected data.

Data Analysis

The results of a Word Frequency (TF) analysis.

- <19> -

TEXTOM Data Analysis Console

Word Analysis | Matrix | Network Analysis | Topic Analysis | Sentiment Analysis | QAP Hypothesis test

Word Frequency | N-gram | TF-IDF | Series

Word Frequency

We calculate the frequency of a specific word mentioned in the entire document. Word frequency can be used as an indicator through importance analysis to infer the topic, attitude, or sentiment of the document.

Top 20 items | Word Selection

0 items selected / Total 6658 cases

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage (%)
eur	987	2.100	2.100
company	973	2.070	4.170
Finnish	521	1.108	5.278
mn	505	1.074	6.353
sale	503	1.070	7.423
year	446	0.949	8.371
share	432	0.919	9.291

Visualization Result

Word Cloud | Bubble Graph | Vertical Bar Graph | Horizontal Bar Graph

Please select the words to visualize from the left list.

Visualization Settings

- Analysis Algorithm**
Click the button to view the description of the analysis algorithm.
- Top Words Auto-Select**
Automatically select the top 20, 50, 75, or 100 words from all words.
- Word Selection**
Click the top checkbox to automatically select the top 100 words. To manually select words, use the checkbox next to each word.
Tip. The percentage represents the amount mentioned when the total mentions are scaled to 100.
- Download Analysis Results**
Click [Selected Download] or [Download All] to download word frequencies and percentages into an Excel (.xlsx) file.
- Visualize the selected words.**
Tip. For better visibility, it is recommended to select between 50 to 70 words.
- Displays the visualization results.**

Data Analysis

Displays the visualization results of Word Frequency (TF).

- <20> -

The screenshot shows the 'Visualization Result' section with four visualization options: Word Cloud, Bubble Graph, Vertical Bar Graph, and Horizontal Bar Graph. The 'Word Cloud' option is selected. A 'Visualization Settings' panel is open, showing options for Shape Selection (Shape Selection and Direct Input), Change to Cloud (Sample Image 1 and Sample Image 2), Shape (upload field), Select Font (Default), Select Size (Default), and Color Selection (color palette). A dashed blue arrow points from the 'Direct Input' option in the settings panel to the word cloud visualization.

- ① **Visualization Results**
Provides four types of visualization results.

Word Cloud	Bubble Chart
Vertical Bar Chart	Horizontal Bar Chart

- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.

- ③ **Shape Selection (Word Cloud)**
Enables users to select a specific shape for the word cloud visualization.
Tip. Refer to sample images and upload an image of the desired shape.

- ④ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.

- ⑤ **Download Visualization**
Download the visualization results as a PNG file.

1 Analysis Algorithms

The statistical-based language analysis model extracts N consecutive elements from a string, calculates the frequency of words appearing consecutively in a sentence. Determines the types based on how many words the sentence will be divided into.

2 Bigram(2) Trigram(3) Tetragram(4)

3 Top 20items Word Selection

13Items selected / Total 3153 cases ※ Full download is available for up to 200,000 records.

5 Selected Download Download All

Word1	Word2	Frequency (Count)	
<input type="checkbox"/>	text	mining	355
<input type="checkbox"/>	mining	technique	312
<input type="checkbox"/>	natural	language	27
<input type="checkbox"/>	language	processing	25
<input type="checkbox"/>	datum	mining	23
<input type="checkbox"/>	sentiment	analysis	21

6 Visualizing Selected Words >>

7 Please select the words to visualize from the left list.

Visualization Settings

Color Selection

Line Color Arrow Shape 1 2 3 T1 Font Size 14

- ① **Analysis Algorithm**
Click the button to view the description of the analysis algorithm.
- ② **Set N-gram Range**
Select the number of consecutive words from 2 (Bigram) to 4 (Tetragram).
- ③ **Top Words Auto-Select**
Automatically select the top 20, 50, 75, or 100 words from all words.
- ④ **N-gram Word Selection**
Click the top checkbox to automatically select the top 100 words. To manually select words, use the checkbox next to each word.
- ⑤ **Download Analysis Results**
Click [Selected Download] or [Download All] to download word frequencies and percentages into an Excel (.xlsx) file.
- ⑥ Visualize the selected words.
- ⑦ Displays the visualization results

Data Analysis

Displays the visualization results of N-gram.

- <22> -

TEXTOM Data Analysis Console

Word Analysis Matrix Network Analysis Topic Analysis Sentiment Analysis QAP Hypothesis test

Word Frequency N-gram TF-IDF Series

N-gram Analysis Algorithms

The statistical-based language analysis model extracts N consecutive elements from a string, calculates the frequency of words appearing consecutively in a sentence. Determines the types based on how many words the sentence will be divided into.

Bigram(2) Trigram(3) Tetragram(4)

Top 20items Word Selection

15Items selected / Total 3153 cases ※ Full download is available for up to 200,000 records. Selected Download Download All

	Word1	Word2	Frequency (Count)
<input type="checkbox"/>	text	mining	355
<input checked="" type="checkbox"/>	mining	technique	312
<input type="checkbox"/>	natural	language	27
<input type="checkbox"/>	language	processing	25
<input type="checkbox"/>	datum	mining	23
<input checked="" type="checkbox"/>	sentiment	analysis	21

Visualizing Selected Words >>

1 Visualization Result View Download

2 Visualization Settings

Color Selection

Line Color Arrow Shape 1 2 3 T! Font Size 14

3 **4**

① Visualization Result

The results are provided as a directional network graph.

Tip. It is important to select relevant words according to the analysis purpose, and it is advisable to exclude common words that do not contain relevant information.

Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.

② Visualization Setting

Users can modify the attributes of the visualization results as desired.

③ Zoom In

Click [View] to display an enlarged version of the selected visualization in a pop-up window.

Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.

④ Download Visualization

Download the visualization result as a PNG file.

TEXTOM Data Analysis Console

Word Analysis | Matrix | Network Analysis | Topic Analysis | Sentiment Analysis | QAP Hypothesis test

Word Frequency | N-gram | **TF-IDF** | Series

TF-IDF 1 Analysis Algorithms

We calculate the frequency of a specific word mentioned in the entire document.
Word frequency can be used as an indicator through importance analysis to infer the topic, attitude, or sentiment of the document.

2 Top 20items Word Selection

0items selected / Total 1480 cases 4 Selected Download Download All

3 <input type="checkbox"/>	Word	TF-IDF	DF	IDF
<input type="checkbox"/>	datum	165	97.000	1.149
<input type="checkbox"/>	analysis	123	70.000	1.475
<input type="checkbox"/>	information	87	36.000	2.140
<input type="checkbox"/>	research	80	31.000	2.290
<input type="checkbox"/>	sentiment	78	26.000	2.465
<input type="checkbox"/>	language	78	35.000	2.168
<input type="checkbox"/>	social	77	29.000	2.356
<input type="checkbox"/>	..	77	29.000	2.356

5 Visualizing Selected Words >>

Visualization Result View Download

Word Cloud | Bubble Graph | Vertical Bar Graph | Horizontal Bar Graph

6 Please select the words to visualize from the left list.

Visualization Settings

- ① **Selected Algorithm Description**
Click the button to view the description of the selected algorithm to analyze.
- ② **Top Words Auto-Select**
Automatically select the top 20, 50, 75, or 100 words from all words.
- ③ **TF-IDF Word Selection**
Click the top checkbox to automatically select the top 100 words. To manually select words, use the checkbox next to each word.
Tip. The higher the TF-IDF value, the more important the word is within a specific document.
- ④ **Download Analysis Results**
Click [Selected Download] or [Download All] to download word frequencies and percentages into an Excel (.xlsx) file.
- ⑤ Visualize the selected words.
- ⑥ Displays the visualization results

1 Visualization Result

Word Cloud | Bubble Graph | Vertical Bar Graph | Horizontal Bar Graph

algorithm processing sentiment model
 extraction tool datum
 feedback classification research pattern
 insight trend

2 Visualization Settings

Shape Selection | Direct Input

Default

Select Font: Default

Select Size: Default | Default | Apply

Color Selection: High | Mid | Low

3 Visualization Settings

Shape Selection | Direct Input

Change to Cloud: Sample Image 1, Sample Image 2

Shape

You can upload a template image to create visualizations in a specific shape. Refer to the sample images and upload an image (jpg, png) of the desired shape.

Please upload the image file. | Apply

Select Font: Default

Select Size: Default | Default | Apply

Color Selection: High | Mid | Low

4 View

5 Download

- ① **Visualization Results**
Provides four types of visualization results.

Word Cloud	Bubble Chart
Vertical Bar Chart	Horizontal Bar Chart

- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.

- ③ **Shape Selection (Word Cloud)**
Enables users to select a specific shape for the word cloud visualization.
Tip. Refer to sample images and upload an image of the desired shape.

- ④ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.

- ⑤ **Download Visualization**
Download the visualization result as a PNG file.

The screenshot displays the 'Data Analysis Console' interface. The 'Word Analysis' tab is active, and the 'Series' sub-tab is selected. The 'Time Series Analysis' section includes a description, a time unit selector (set to 'Date'), a trendline dropdown (set to 'None'), and an 'Apply' button. Below this is a table showing data for Google(Web) and Google(New) from 2024-01-01 to 2024-03-01. To the right, a 'Visualization Result' chart shows a line graph with a blue trend line for Google(Web) and a green line for Google(New). The chart includes 'View' and 'Download' buttons. Below the chart are 'Visualization Settings' for Google(Web) and Google(New), and a 'Legend Position' dropdown.

Time Series Analysis

We conduct time series analysis based on data collection volume.

1 Analysis Algorithms

2 Select the time unit.

3 Trendline

4 Apply

5

Date	Google(Web)	Google(New)
2024-01-01	74	1
2024-02-01	69	1
2024-03-01	9	1

6 Download

Visualization Result

View Download

Google(Web) Google(New)

80
60
40
20
0

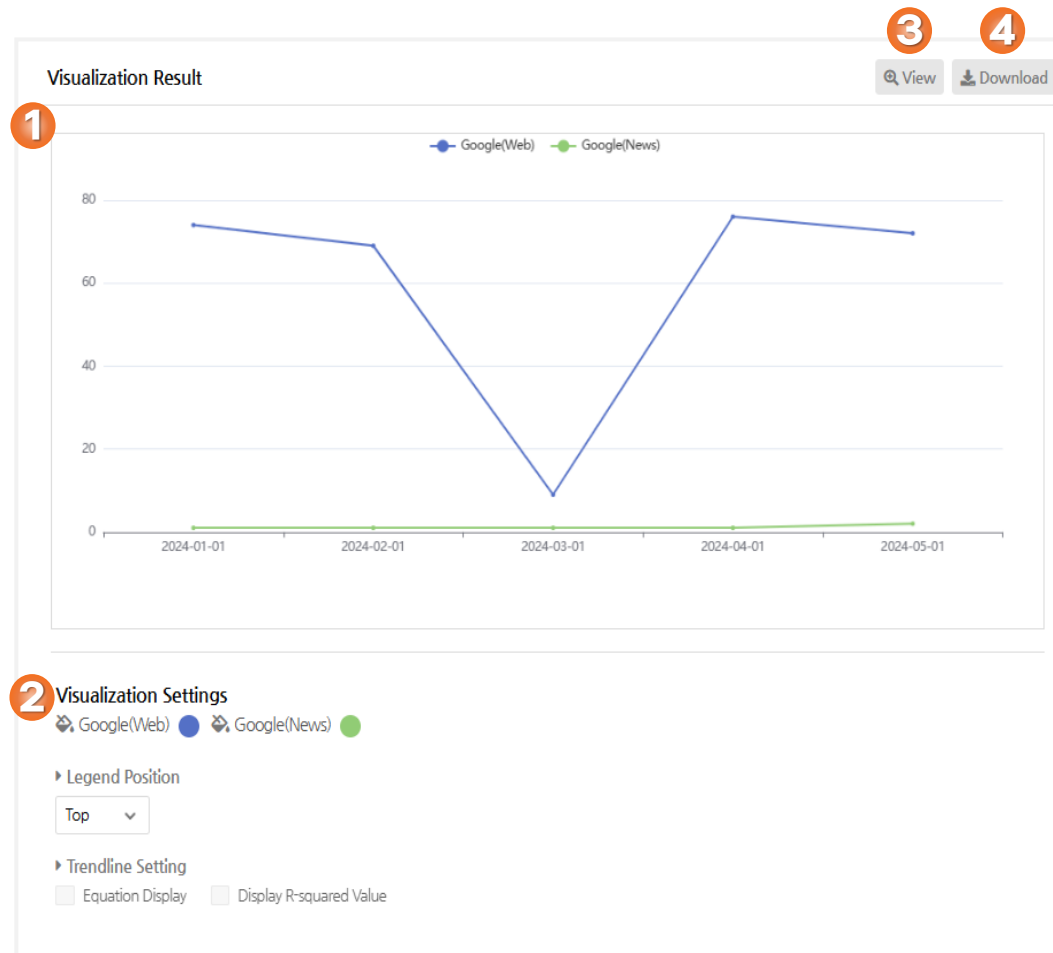
2024-01-01 2024-02-01 2024-03-01 2024-04-01 2024-05-01

Visualization Settings

Google(Web) Google(New)

Legend Position

- ① **Analysis Algorithm**
Click the button to view the description of the analysis algorithm.
- ② **Select Time Unit**
Set the time unit to day, week, month, or year based on the data collection date.
Tip. This option is available when the 'Unit Collection' feature is used during the data collection stage.
Day : Day, Week, Month, Year
Week : Week, Month, Year
Month : Month, Year
Year : Year
- ③ **Display Trend Line**
Show a linear trend line in the visualization results to easily identify data collection trends.
Tip. To display the trend line in the visualization results, ensure you select [Linear].
- ④ Apply the selected options to the time series analysis.
- ⑤ Display the time series analysis results of the data collection volume.
- ⑥ Download the analysis result as an Excel (.xlsx) file.



- ① **Visualization Results**
Displays the result of the analysis in a linear graph.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
- ④ **Download Visualization**
Download the visualization result as a PNG file.

1 Analysis Algorithms

We conduct time series analysis based on data collection volume.

2 Select the time unit.

3 Select Analysis Words

4 Trendline

5 Apply

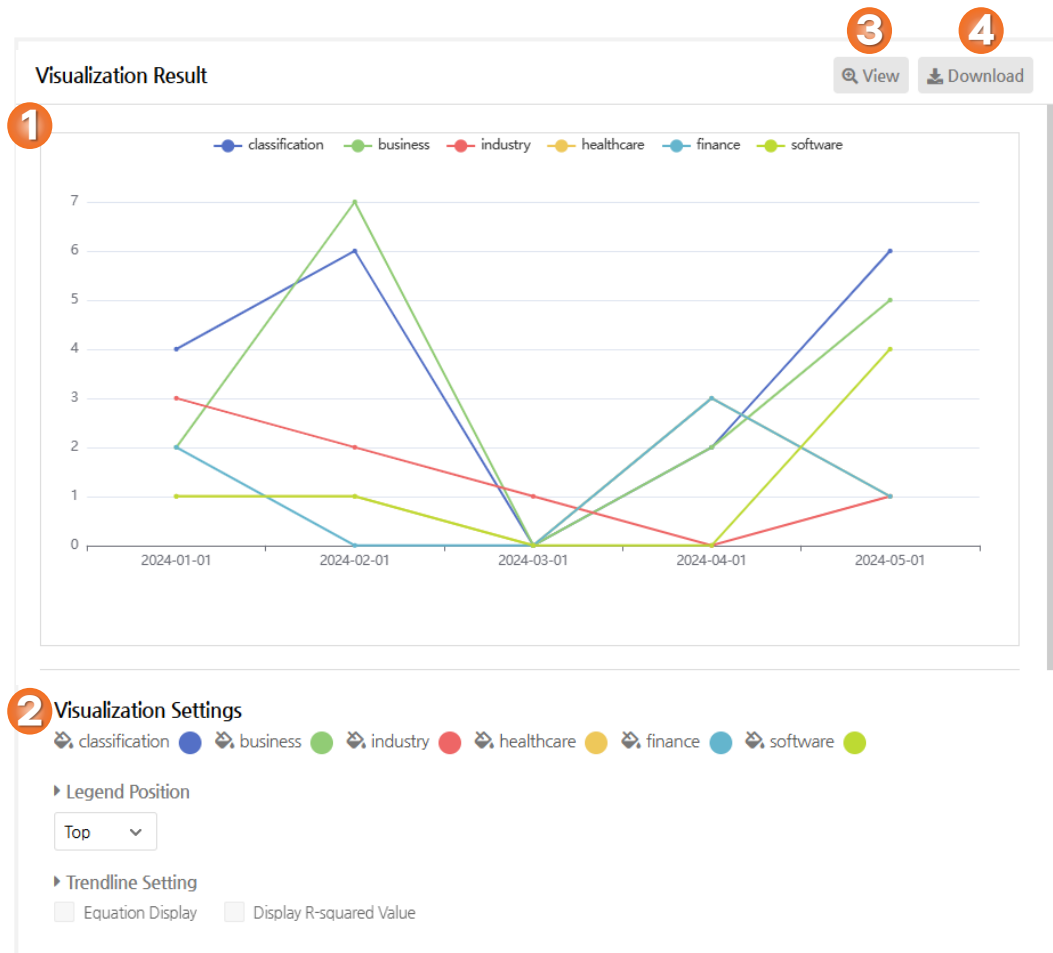
6

7 Download

Date	classification	business	industry	healthcare	finance	software
2024-01-01	4	2	3	1	2	1
2024-02-01	6	7	2	1	0	1

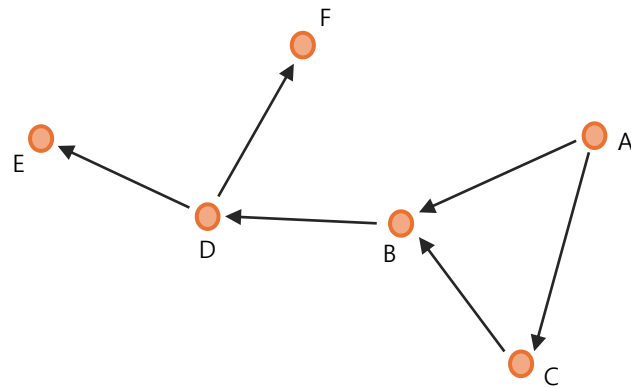
단어	빈도	백분율 (%)	누적비율 (%)
text	409	8.132829588387354	8.132829588387354
mining	384	7.635712865380792	15.768542453768145
technique	322	6.402863392324518	22.17140584609266
datum	144	2.863392324517797	25.034798170610458
analysis	84	1.6703121893020483	26.705110359912506
information	41	0.8152714257307616	27.520381785643266
language	36	0.7158480811294492	28.236229866772717
research	35	0.6959634122091867	28.932193278981906

- Selected Algorithm Description**
Click the button to view the description of the algorithm.
- Select Time Unit**
Set the time unit to day, week, month, or year based on the data collection date.
Tip. This option is available when the 'Unit Collection' feature is used during the data collection stage.
Day : Day, Week, Month, Year
Week : Week, Month, Year
Month : Month, Year
Year : Year
- Select Word for Analysis**
select 2 to 10 words by clicking the "Select Now" button.
- Display Trend Line**
Show a linear trend line in the visualization results to easily identify data collection trends.
Tip. To display the trend line in the visualization results, ensure you select [Linear].
- Apply the selected options to the time series analysis.
- Display the time series analysis result of the data collection volume.
- Download the analysis result as an Excel (.xlsx) file.



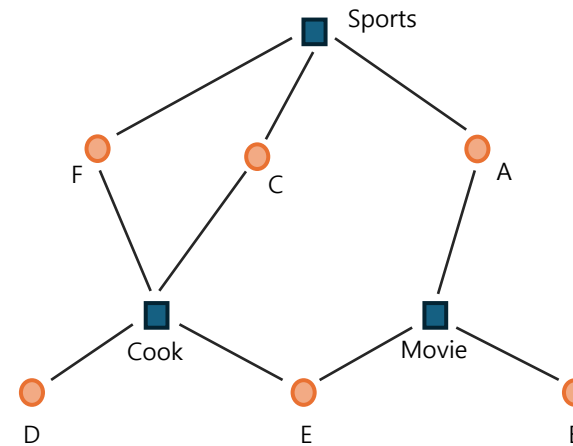
- ① **Visualization Results**
Displays the result of the analysis in a linear graph.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
- ④ **Download Visualization**
Download the visualization results as a PNG file.

① 1-Mode Matrix



	A	B	C	D	E	F
A	0	1	1	0	0	0
B	0	0	0	1	0	0
C	0	1	0	0	0	0
D	0	0	0	0	1	1
E	0	0	0	0	0	0
F	0	0	0	0	0	0

① 2-Mode Matrix



	Sprots	Movie	Cook
A	1	1	0
B	0	1	0
C	1	0	1
D	0	0	1
E	0	1	1
F	1	0	1

Matrix data is analyzed by constructing a matrix of word co-occurrence frequencies.

① 1-Mode Matrix

The most common type of network, with rows and columns representing the same words.

Tip. Suitable for analysis when the main node.

② 2-Mode Matrix

Rows and columns represent different words, allowing for analysis of relationships between different types of entities.

Tip. Suitable for analyzing relationships between different types of entities.

TEXTOM Data Analysis Console 백데이터팀 LOG OUT

Word Analysis **Matrix** Network Analysis Topic Analysis Sentiment Analysis QAP Hypothesis test

1-Mode 2-Mode

Matrix 1 Analysis Algorithms

Generate the co-occurrence matrix for the selected words.
Select between 2 and 200 words.

2 Select word Select a word Upload File

3 Top 20items Word Select

5items selected / Total 6658 cases 5 Selected Download Download All

4 <input type="checkbox"/>	Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
<input type="checkbox"/>	eur	987	2.100	2.100
<input type="checkbox"/>	company	973	2.070	4.170
<input type="checkbox"/>	Finnish	521	1.108	5.278
<input type="checkbox"/>	mn	505	1.074	6.353
<input type="checkbox"/>	sale	503	1.070	7.423
<input type="checkbox"/>

6 Apply

Matrix Results Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

7 Please select the words to visualize from the left list.

- Selected Algorithm Description**
Click the button to view the description of the algorithm.
- Set Word Selection Method**
Select 2 to 200 words for analysis. You can choose words directly from the analysis console or upload an Excel file.
- Top Words Auto-Select**
Automatically select the top 20, 50, 75, or 100 words from all words.
- Word Selection**
Click the top checkbox to automatically select the top 100 words. To manually select words, use the checkbox next to each word.
Tip. Selecting key words related to the analysis topic is crucial. Review the results after the initial selection and adjust as necessary to enhance the accuracy of the analysis.
- Download Analysis Results**
Click [Selected Download] or [Download All] to download word frequencies and percentages into an Excel (.xlsx) file.
- Apply the selected words to the analysis.
- Display the analysis results.

1-Mode
2-Mode

Matrix Analysis Algorithms

Generate the co-occurrence matrix for the selected words.
Select between 2 and 200 words.

▶ Select word Select a word Upload File

1 C:\fakepath\sample_matrix (1).xlsx Example File **2**

You can view the data for the analyzed words among all the words. Download **5**

4 Number	Word	Frequency _{case}
1	business	16
2	education	8
3	feedback	11
4	healthcare	6
5	finance	6
6	visualization	5

3 Apply

	A	B
1	business	
2	education	
3	feedback	
4	healthcare	
5	finance	
6	visualization	
7	software	
8	service	
9	platform	

- ① **File Upload**
Upload an Excel file (.xlsx) containing the matrix word list.
- ② **Sample File**
Click this button to download a sample format.
- ③ Apply the selected words to the analysis.
- ④ **Selected Words Review**
View the list of words applied to the analysis from the entire word set.
- ⑤ **Download Word List**
Download the list of words applied to the analysis as an Excel (.xlsx) file.

Matrix

The result of 1-Mode Matrix Analysis.

- <32> -

TEXTOM Data Analysis Console

Word Analysis **Matrix** Network Analysis Topic Analysis Sentiment Analysis QAP Hypothesis test

1-Mode 2-Mode

Matrix [Analysis Algorithms](#)

Generate the co-occurrence matrix for the selected words.
Select between 2 and 200 words.

Select word

Top 20items Word

22Items selected / Total 1479 cases

<input type="checkbox"/>	Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
<input type="checkbox"/>	text	409	8.133	8.133
<input type="checkbox"/>	mining	384	7.636	15.769
<input checked="" type="checkbox"/>	technique	322	6.403	22.171
<input checked="" type="checkbox"/>	datum	144	2.863	25.035
<input type="checkbox"/>	analysis	84	1.670	26.705
<input type="checkbox"/>	information	44	0.88	27.585

Matrix Results

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

	technique	datum	research	social	application	sentiment	processing	insight	trend
technique	0	159	37	33	40	36	27	28	30
datum	159	0	13	16	15	14	14	19	17
research	37	13	0	2	5	1	0	1	13
social	33	16	2	0	2	7	1	3	5
application	40	15	5	2	0	0	5	0	6
sentiment	36	14	1	7	0	0	2	7	4
processing	27	14	0	1	5	2	0	4	0
insight	28	19	1	3	0	7	4	0	4
trend	30	17	13	5	6	4	0	4	0
topic	23	7	0	1	1	5	4	1	1
media	21	13	0	25	1	7	0	3	4

① Matrix Results

The analysis results are provided using the following six similarity coefficients. Click on the analysis name to display the corresponding results below.

- Matrix
- Edge List
- Euclidean Coefficient
- Cosine Coefficient
- Jaccard Coefficient
- Correlation Coefficient

② Download Analysis Result

Download the analysis results as an Excel (.xlsx) file.

Tip. The option to download all analysis results at once is not available.

Matrix

The result of 1-Mode Matrix Analysis.

- <33> -

1 Matrix Results

Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

	technique	datum	research	social	application	sentiment	processing	insight	trend
technique	0	159	37	33	40	36	27	28	30
datum	159	0	13	16	15	14	14	19	17
research	37	13	0	2	5	1	0	1	13
social	33	16	2	0	2	7	1	3	5
application	40	15	5	2	0	0	5	0	6
sentiment	36	14	1	7	0	0	2	7	4
processing	27	14	0	1	5	2	0	4	0
insight	28	19	1	3	0	7	4	0	4
trend	30	17	13	5	6	4	0	4	0
topic	23	7	0	1	1	5	4	1	1
media	21	13	0	25	1	7	0	3	4

2 Matrix Results

Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

Word1	Word2	Weight(Tf)
technique	datum	159
technique	research	37
technique	social	33
technique	application	40
technique	sentiment	36
technique	processing	27
technique	insight	28
technique	trend	30
technique	topic	23
technique	media	21
technique	network	22
technique	tool	91

3 Matrix Results

Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

	technique	datum	research	social	application	sentiment	processing	insight	trend	topic
technique	0	0.9055	0.5917	0.5527	0.5917	0.7113	0.4226	0.622	0.5917	0.6464
datum	0.9055	0	0.5527	0.5917	0.7418	0.2928	0.7327	0.6666	0.5917	0.2928
research	0.5917	0.5527	0	0	0	0	0	0	0.2928	0
social	0.5527	0.5917	0	0	0	0.2928	0	0	0.2928	0
application	0.5917	0.7418	0	0	0	0	0	0	0	0
sentiment	0.7113	0.2928	0	0.2928	0	0	0	0	0.2928	0
processing	0.4226	0.7327	0	0	0	0	0	0	0.2928	0
insight	0.622	0.6666	0	0	0	0.2928	0.2928	0	0	0
trend	0.5917	0.5917	0.2928	0.2928	0	0	0	0	0	0
topic	0.6464	0.2928	0	0	0	0	0	0	0	0
media	0.4226	0.5527	0	0	0	0.2928	0	0	0	0

4 Matrix Results

Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

	technique	datum	research	social	application	sentiment	processing	insight	trend	topic
technique	0	0.2752	0.2431	0.2249	0.2277	0.1971	0.2332	0.1852	0.188	0.1545
datum	0.2752	0	0.0877	0.1078	0.0363	0.1018	0.0869	0.0888	0.0901	0.0508
research	0.2431	0.0877	0	0.0476	0.0964	0.0224	0	0.0261	0.2123	0
social	0.2249	0.1078	0.0476	0	0.0493	0.1151	0.0295	0.0803	0.0815	0.0276
application	0.2277	0.0363	0.0964	0.0493	0	0	0.1195	0	0.055	0.0279
sentiment	0.1971	0.1018	0.0224	0.1151	0	0	0.0557	0.1264	0.0769	0.1303
processing	0.2332	0.0869	0	0.0295	0.1195	0.0557	0	0.0548	0	0.1336
insight	0.1852	0.0888	0.0261	0.0803	0	0.1264	0.0548	0	0.1194	0.0303
trend	0.188	0.0901	0.2123	0.0815	0.055	0.0769	0	0.1194	0	0.0307
topic	0.1545	0.0508	0	0.0276	0.0279	0.1303	0.1336	0.0303	0.0307	0
media	0.1851	0.115	0	0.5622	0.0316	0.1474	0	0.1028	0.1044	0

5 Matrix Results

Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

	technique	datum	research	social	application	sentiment	processing	insight	trend	topic
technique	0	0.1995	0.0868	0.0788	0.079	0.0734	0.0689	0.0603	0.0605	0.0494
datum	0.1995	0	0.0558	0.0677	0.0227	0.0681	0.047	0.0529	0.0532	0.0301
research	0.0868	0.0558	0	0.0294	0.0597	0.0149	0	0.0163	0.1333	0
social	0.0788	0.0677	0.0294	0	0.0307	0.0769	0.0169	0.0508	0.0517	0.0181
application	0.079	0.0227	0.0597	0.0307	0	0	0.0689	0	0.035	0.0185
sentiment	0.0734	0.0681	0.0149	0.0769	0	0	0.0344	0.0862	0.0526	0.0925
processing	0.0689	0.047	0	0.0169	0.0689	0.0344	0	0.0384	0	0.0833
insight	0.0603	0.0529	0.0163	0.0508	0	0.0862	0.0384	0	0.0784	0.0208
trend	0.0605	0.0532	0.1333	0.0517	0.035	0.0526	0	0.0784	0	0.0212
topic	0.0494	0.0301	0	0.0181	0.0185	0.0925	0.0833	0.0208	0.0212	0
media	0.0524	0.0606	0	0.3333	0.0188	0.0943	0	0.0638	0.0652	0

6 Matrix Results

Download

Matrix Edge List Euclidean Coefficient Cosine Coefficient Jaccard Coefficient Correlation Coefficient

	technique	datum	research	social	application	sentiment	processing	insight	trend	topic
technique	0	0.0682	0.0057	0.0435	0.1497	0.0621	0.0113	0.0224	0.0799	0.0028
datum	0.0682	0	0.0313	0.0401	0.0056	0.0004	0.0244	0.0921	0.0336	0.0191
research	0.0057	0.0313	0	0.0375	0.0491	0.0603	0.092	0.0508	0.3045	0.0694
social	0.0435	0.0401	0.0375	0	0.0547	0.0572	0.053	0.0212	0.0205	0.0338
application	0.1497	0.0056	0.0491	0.0547	0	0.0777	0.0853	0.0732	0.0537	0.0625
sentiment	0.0621	0.0004	0.0603	0.0572	0.0777	0	0.0172	0.1402	0.0188	0.0865
processing	0.0113	0.0244	0.092	0.053	0.0853	0.0172	0	0.0356	0.0731	0.0802
insight	0.0224	0.0921	0.0508	0.0212	0.0732	0.1402	0.0356	0	0.0727	0.0224
trend	0.0799	0.0336	0.3045	0.0205	0.0537	0.0188	0.0731	0.0727	0	0.0552
topic	0.0028	0.0191	0.0694	0.0338	0.0625	0.0865	0.0802	0.0224	0.0552	0
media	0.035	0.0641	0.0754	0.0951	0.0321	0.1692	0.0702	0.0566	0.093	0.0529

- Matrix**
Displays the co-occurrence frequency of selected words within the entire document.
- Edge List**
Shows pairs of selected words (word1, word2) and their frequency (Weight) in a list format.
Tip: The option to download all analysis results at once is not available.
- Euclidean Coefficient**
Measures and provides the similarity between selected words using Euclidean distance.
Tip: The smaller the value between two vectors, the more similar they are.
- Cosine Coefficient**
Calculates and provides the cosine similarity between two words vectors.
Tip: The coefficient ranges from 0 to 1, with values closer to 1 indicating higher similarity between the vectors.
- Jaccard Coefficient**
Shows the Jaccard similarity between words.
Tip: The coefficient ranges from 0 to 1, with higher values indicating greater similarity.
- Correlation Coefficient**
Indicates the correlation between words.
Tip: Higher values indicate a stronger relationship between the two variables.

Matrix

Perform 2-Mode Matrix Analysis.

- <34> -

1-Mode | 2-Mode

Matrix 1 Analysis Algorithms

Generate the co-occurrence matrix for the selected words.
Select between 2 and 200 words.

2 Select word

Please click the "Apply" button after selecting both the row and column words to view the matrix results.

3 Column word Row word

4 Top 20Items Word

5 0Items selected / Total 1479 cases 6 Selected Download

<input type="checkbox"/>	Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
<input type="checkbox"/>	text	409	8.133	8.133
<input type="checkbox"/>	mining	384	7.636	15.769
<input type="checkbox"/>	technique	322	6.403	22.171
<input type="checkbox"/>	datum	144	2.863	25.035
<input type="checkbox"/>	analysis	84	1.670	26.705
<input type="checkbox"/>	information	41	0.815	27.520
<input type="checkbox"/>	language	36	0.716	28.236
<input type="checkbox"/>	research	35	0.696	28.932
<input type="checkbox"/>	social	33	0.656	29.588
<input type="checkbox"/>	application	32	0.636	30.225
<input type="checkbox"/>	sentiment	32	0.636	30.861
<input type="checkbox"/>	nlp	28	0.557	31.418

7

8

Column word | Row word

Top 20Items Word

0Items selected / Total 1479 cases Selected Download

<input type="checkbox"/>	Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
<input type="checkbox"/>	text	409	8.133	8.133
<input type="checkbox"/>	mining	384	7.636	15.769
<input type="checkbox"/>	technique	322	6.403	22.171
<input type="checkbox"/>	datum	144	2.863	25.035
<input type="checkbox"/>	analysis	84	1.670	26.705
<input type="checkbox"/>	information	41	0.815	27.520
<input type="checkbox"/>	language	36	0.716	28.236
<input type="checkbox"/>	research	35	0.696	28.932
<input type="checkbox"/>	social	33	0.656	29.588
<input type="checkbox"/>	application	32	0.636	30.225
<input type="checkbox"/>	sentiment	32	0.636	30.861
<input type="checkbox"/>	nlp	28	0.557	31.418

- Selected Algorithm Description**
Click the button to view the description of the algorithm.
- Set Word Selection Method**
Select 2 to 200 words for analysis. You can choose words directly from the analysis console or upload an Excel file.
- Select the tab to switch between column words and row words.
- Top Words Auto-Select**
Automatically select the top 20, 50, 75, or 100 words from all words.
- Word Selection**
Click the top checkbox to automatically select the top 100 words. To manually select words, use the checkbox next to each word.
Tip. Selecting key words related to the analysis topic is crucial. Review the results after the initial selection and adjust as necessary to enhance the accuracy of the analysis.
- Download Analysis Results**
Click [Selected Download] or [Download All] to download word frequencies and percentages into an Excel (.xlsx) file.
- After selecting the column and row words, click the [Apply] button to save.
- Display the analysis results.

TEXTOM Data Analysis Console

Word Analysis **Matrix** Network Analysis Topic Analysis Sentiment Analysis QAP Hypothesis test

1-Mode 2-Mode

Matrix

Generate the co-occurrence matrix for the selected words.
Select between 2 and 200 words.

Select word

Please click the "Apply" button after selecting both the row and column words to view the matrix results.

Column word Row word

Top 20Items Word

10Items selected / Total 1479 cases

<input type="checkbox"/>	Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
<input type="checkbox"/>	text	409	8.133	8.133
<input type="checkbox"/>	mining	384	7.636	15.769
<input checked="" type="checkbox"/>	technique	322	6.403	22.171
<input type="checkbox"/>	datum	144	2.863	25.035
<input type="checkbox"/>	analysis	84	1.670	26.705
<input type="checkbox"/>	information	41	0.815	27.520

Matrix Results

	customer	public	university	news	researcher	healthcare	finance	software
technique	22	9	10	8	7	7	6	7
research	0	1	3	0	4	0	0	1
nlp	1	0	0	0	2	0	0	0
processing	0	0	0	2	0	0	5	0
insight	5	1	0	5	0	0	0	1
trend	2	1	1	0	0	1	0	0
topic	5	0	0	0	0	0	0	0
network	0	1	0	1	0	1	0	4
tool	2	0	1	0	0	0	0	1
classification	4	0	0	0	0	1	0	0
unstructured	4	0	0	1	0	2	0	0
knowledge	0	0	1	1	0	2	0	0

① Matrix Results

The analysis results are provided using the following six similarity coefficients. Click on the analysis name to display the corresponding results below.

- Matrix
- Edge List
- Euclidean Coefficient
- Cosine Coefficient
- Jaccard Coefficient
- Correlation Coefficient

② Download Analysis Result

Download the analysis results as an Excel (.xlsx) file.

Tip. The option to download all analysis results at once is not available.

Matrix

The result of 2-Mode Matrix Analysis.

- <36> -

1 Matrix Results Download

Matrix	customer	public	university	news	researcher	healthcare	finance	software
technique	22	9	10	8	7	7	6	7
research	0	1	3	0	4	0	0	1
rip	1	0	0	0	2	0	0	0
processing	0	0	0	2	0	0	5	0
insight	5	1	0	5	0	0	0	1
trend	2	1	1	0	0	1	0	0
topic	5	0	0	0	0	0	0	0
network	0	1	0	1	0	1	0	4
tool	2	0	1	0	0	0	0	1
classification	4	0	0	0	0	1	0	0
unstructured	4	0	0	1	0	2	0	0
knowledge	0	0	1	1	0	2	0	0

2 Matrix Results Download

Word1	Word2	Weight
customer	technique	22
customer	research	0
customer	rip	1
customer	processing	0
customer	insight	5
customer	trend	2
customer	topic	5
customer	network	0
customer	tool	2
customer	classification	4
customer	unstructured	4
customer	knowledge	0

3 Matrix Results Download

Matrix	customer	public	university	news	researcher	healthcare	finance	software
technique	0.5	0	0	0	0	0	0.2928	0.2928
research	0	0	0	0	0	0	0	0
rip	0	0	0	0	0	0	0	0
processing	0	0	0	0	0	0	0.2928	0
insight	0.2928	0	0	0	0	0	0	0
trend	0	0	0	0	0	0	0	0
topic	0.5	0	0	0	0	0	0	0
network	0	0	0	0	0	0	0	0
tool	0	0	0	0	0	0	0	0
classification	0	0	0	0	0	0	0	0
unstructured	0	0	0	0	0	0	0	0
knowledge	0	0	0	0	0	0	0	0

4 Matrix Results Download

Matrix	customer	public	university	news	researcher	healthcare	finance	software
technique	0.0942	0.1195	0.1285	0.075	0.0998	0.0824	0.0382	0.0864
research	0	0.0293	0.1143	0	0.1268	0	0	0.0293
rip	0.0198	0	0	0	0.0449	0	0	0
processing	0	0	0	0.0394	0	0	0.1052	0
insight	0.0628	0.0442	0	0.0715	0	0	0	0.0442
trend	0.0212	0.0449	0.0435	0	0	0.0485	0	0
topic	0.0647	0	0	0	0	0	0	0
network	0	0.0449	0	0.0362	0	0.0485	0	0.0449
tool	0.052	0	0.0539	0	0	0	0	0.055
classification	0.0249	0	0	0	0	0.0545	0	0
unstructured	0.0764	0	0	0.0434	0	0.0557	0	0
knowledge	0	0	0.0806	0.0505	0	0.1296	0	0

5 Matrix Results Download

Matrix	customer	public	university	news	researcher	healthcare	finance	software
technique	0	0	0	0	0	0	0	0
research	0	0	0	0	0	0	0	0
rip	0	0	0	0	0	0	0	0
processing	0	0	0	0	0	0	0	0
insight	0	0	0	0	0	0	0	0
trend	0	0	0	0	0	0	0	0
topic	0	0	0	0	0	0	0	0
network	0	0	0	0	0	0	0	0
tool	0	0	0	0	0	0	0	0
classification	0	0	0	0.029	0	0	0	0
unstructured	0	0	0	0	0.0316	0	0	0
knowledge	0	0	0	0	0	0.1296	0	0

6 Matrix Results Download

Matrix	customer	public	university	news	researcher	healthcare	finance	software
technique	0.0751	0.0894	0.0223	0.0178	0.0223	0.0452	0.0159	0.0518
research	0.0625	0.0137	0.1374	0.0384	0.1374	0.0444	0.0343	0.0368
rip	0.0169	0.0411	0.0411	0.0328	0.0873	0.038	0.0293	0.0232
processing	0.0581	0.0447	0.0447	0.1045	0.0447	0.0413	0.2812	0.0274
insight	0.1535	0.0395	0.0395	0.1899	0.0395	0.0365	0.0282	0.0242
trend	0.0295	0.0314	0.0314	0.0305	0.0382	0.0397	0.0272	0.0234
topic	0.036	0.0337	0.0337	0.0269	0.0337	0.0312	0.024	0.0207
network	0.0424	0.0353	0.0326	0.0261	0.0326	0.0431	0.0333	0.2719
tool	0.0502	0.0344	0.0491	0.0274	0.0344	0.0317	0.0246	0.021
classification	0.1388	0.0329	0.0329	0.0262	0.0329	0.0557	0.0234	0.0201
unstructured	0.144	0.0316	0.0316	0.0505	0.0316	0.1463	0.0225	0.0194
knowledge	0.0482	0.0371	0.0583	0.0296	0.0371	0.1714	0.0284	0.0227

- ① **Matrix**
Displays the co-occurrence frequency of selected words within the entire document.
- ② **Edge List**
Shows pairs of selected words (word1, word2) and their frequency (Weight) in a list format.
Tip. The option to download all analysis results at once is not available.
- ③ **Euclidean Coefficient**
Measures and provides the similarity between selected words using Euclidean distance.
Tip. The smaller the value between two vectors, the more similar they are.
- ④ **Cosine Coefficient**
Calculates and provides the cosine similarity between two words vectors.
Tip. The coefficient ranges from 0 to 1, with values closer to 1 indicating higher similarity between the vectors.
- ⑤ **Jaccard Coefficient**
Shows the Jaccard similarity between words.
Tip: The coefficient ranges from 0 to 1, with higher values indicating greater similarity.
- ⑥ **Correlation Coefficient**
Indicates the correlation between words.
Tip: Higher values indicate a stronger relationship between the two variables.

TEXTOM Data Analysis Console

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network Shortest path CONCOR Clustering

1 Analysis Algorithms

Network Attribute

Analyzing structural descriptive statistics for the entire network, measuring the interconnectedness of nodes within the network using relevant indices.
※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

2 Analysis Word

3 Select Matrix Analysis Word.

Network Attribute Analysis Result

- Selected Algorithm Description**
Click the button to view the description of the algorithm.
- Select word**
Click [Select Directly] to navigate to the 1-Mode Matrix word selection page and select words for analysis.
Tip. A 1-Mode Matrix analysis must be conducted first.
- Display the analysis results.**

TEXTOM Data Analysis Console 백터이터팀 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network Shortest path CONCOR Clustering

Network Attribute 2 [Analysis Algorithms](#)

Analyzing structural descriptive statistics for the entire network, measuring the interconnectedness of nodes within the network using relevant indices.
※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

Analysis Word Download

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
technique	322	6.403	22.171
datum	144	2.863	25.035
research	35	0.696	28.932
social	33	0.656	29.588
application	32	0.636	30.225
sentiment	32	0.636	30.861
processing	26	0.517	33.565
insight	26	0.517	34.082
trend	25	0.497	34.579

1 << Reselect

Network Attribute Analysis Result 4 Download

Network Measures	Value
Nodes	22
Total Edges	151
Diameter	2
Degree Centralization	0.38095
Closeness Centralization	0.5251
Betweenness Centralization	0.00017
Connected Components	1
Overall Reciprocity	0

① **Reselect Words**
Click [Reselect] to navigate to the 1-Mode Matrix word selection page and reselect words for analysis.

② **Download Analysis Words**
Download the selected words as an Excel (.xlsx) file.

③ **Network Attribute Analysis Results**
Calculate eight types of structural property metrics for the 1-Mode Matrix.

Nodes
Total Edges
Diameter
Degree Centralization
Closeness Centralization
Betweenness Centralization
Connected Components
Overall Reciprocity

④ **Download Analysis Result**
Download the analysis result as an Excel (.xlsx) file.

The screenshot displays the TEXTOM Data Analysis Console interface. The top navigation bar includes the TEXTOM logo, a home icon, and the text 'Data Analysis Console'. On the right side of the navigation bar, there is a user profile icon labeled '박대익터팀' and a 'LOG OUT' button. Below the navigation bar, there are several tabs: 'Word Analysis', 'Matrix', 'Network Analysis' (which is currently selected), 'Topic Analysis', 'Sentiment Analysis', and 'QAP Hypothesis test'. Under the 'Network Analysis' tab, there are sub-tabs: 'Network Attribute', 'Centrality' (selected), 'Ego Network', 'Shortest path', 'CONCOR', and 'Clustering'. The main content area is divided into two panels. The left panel, titled 'Centrality', contains a description: 'Centrality measures quantify the relative importance of specific nodes in a network. Through centrality analysis, we can identify nodes that are more influential (meaningful) from a structural perspective.' Below this description is a note: 'Network analysis is generated based on the [Matrix] 1-Mode selected word.' At the bottom of this panel, there is a section labeled 'Analysis Word' with a 'Select Directly' button. The right panel, titled 'CONCOR analysis result', is currently empty and contains a large orange circle with the number '3' and a list icon, with the text 'Select Matrix Analysis Word.' below it. Three numbered callouts are overlaid on the screenshot: '1' points to the 'Analysis Algorithms' link in the Centrality description; '2' points to the 'Analysis Word' section; and '3' points to the 'Select Matrix Analysis Word.' button in the CONCOR analysis result panel.

- ① **Selected Algorithm Description**
Click the button to view the description of the algorithm.
- ② **Select word**
Click [Select Directly] to navigate to the 1-Mode Matrix word selection page and select words for analysis.
Tip. A 1-Mode Matrix analysis must be conducted first.
- ③ **Display the analysis results.**

Network Analysis

The result of Centrality Analysis.

- <40> -

TEXTOM Data Analysis Console 백테이터링 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute **Centrality** Ego Network Shortest path CONCOR Clustering

Centrality Analysis Algorithms

Centrality measures quantify the relative importance of specific nodes in a network. Through centrality analysis, we can identify nodes that are more influential (meaningful) from a structural perspective.
※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

Analysis Word Download

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
technique	322	6.403	22.171
datum	144	2.863	25.035
research	35	0.696	28.932
social	33	0.656	29.588
application	32	0.636	30.225
sentiment	32	0.636	30.861
processing	26	0.517	33.565
insight	26	0.517	34.082

1 << Reselect

CONCOR analysis result Download

Word	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality	Pagerank	Clustering Coefficient
technique	1	1	0.055	0.295	0.247	0.619
datum	1	1	0.055	0.295	0.154	0.619
research	0.619	0.724	0.011	0.203	0.039	0.756
social	0.857	0.875	0.028	0.267	0.05	0.693
application	0.714	0.778	0.026	0.219	0.04	0.638
sentiment	0.857	0.875	0.03	0.264	0.048	0.673
processing	0.571	0.7	0.01	0.186	0.031	0.758
insight	0.667	0.75	0.012	0.217	0.039	0.758
trend	0.667	0.75	0.014	0.216	0.042	0.736
topic	0.762	0.808	0.02	0.243	0.031	0.708
media	0.571	0.7	0.006	0.194	0.036	0.803
network	0.524	0.677	0.01	0.17	0.026	0.745
tool	0.714	0.778	0.021	0.224	0.027	0.686
classification	0.619	0.724	0.013	0.201	0.024	0.744

3 4 Download

① **Reselect Words**
Click [Reselect] to navigate to the 1-Mode Matrix word selection page and reselect words for analysis.

② **Download Analysis Words**
Download the selected words as an Excel (.xlsx) file.

③ **Centrality Analysis Result**
Displays five types of centrality metrics and clustering coefficients. The centrality metrics are as follows.

Degree Centrality
Closeness Centrality
Betweenness Centrality
Eigenvector Centrality
PageRank

④ **Download Analysis Result**
Download the analysis result as an Excel (.xlsx) file.

TEXTOM Data Analysis Console 백데이터팀 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality **Ego Network** Shortest path CONCOR Clustering

Ego Network Analysis Algorithms

The network targeting ego (Ego) nodes from the entire network is extracted, and the connection relationships with other nodes (alter) connected to the ego nodes are analyzed.
※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

2 ▶ Ego Node Enter the Ego Node word.

3 ▶ Analysis Word Select Directly

Ego Network Analysis Results

Ego Network Properties Ego Network Centrality Visualization Result

4 Please select analysis options.

① **Selected Algorithm Description**
Click the button to view the description of the algorithm.

② **Enter Ego Node Word**
Input the word that will be the focus of the analysis. The network will be extracted based on this central word.

③ **Select word**
Select the word to analyze in relation to the ego node (word).
Tip. A 1-Mode Matrix analysis must be conducted first.

④ Display the analysis results.

TEXTOM Data Analysis Console 백대이터팀 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality **Ego Network** Shortest path CONCOR Clustering

Ego Network Analysis Algorithms

The network targeting ego (Ego) nodes from the entire network is extracted, and the connection relationships with other nodes (alter) connected to the ego nodes are analyzed.
 ※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

Ego Node

1 Analysis Word **4** Download

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
tool	18	0.358	37.9
classification	18	0.358	38.258
business	16	0.318	39.909
algorithm	14	0.278	42.792
pattern	13	0.259	43.05
textual	11	0.219	45.476

2 << Reselect **3** ✓ Apply

Ego Network Analysis Results

Ego Network Properties Ego Network Centrality Visualization Result Download

Ego Node	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
business	15	94	210	0.44761	116	0.00014

- Review analysis word list.
- Reselect Words**
Click [Reselect] to navigate to the 1-Mode Matrix word selection page and reselect
 Tip. A 1-Mode Matrix analysis must be conducted first.
- Apply the words to be analyzed alongside the ego node to the analysis.
- Download Analysis Words**
Download the selected words as an Excel (.xlsx) file.

Ego Network Analysis Results

1 Ego Network Properties Ego Network Centrality Visualization Result

Download

Ego Node	Size	Ties	Pairs	Ego-Density	Broker	Ego Between
business	15	94	210	0.44761	116	0.00014

2 Ego Network Properties Ego Network Centrality Visualization Result

3 Download

Ego Node	Alters	Degree	Closeness	Betweenness	Eigenvector	Pagerank	Clustering Coefficient
business	datum	1	1	0.032	0.303	0.157	0.752
	research	0.733	0.789	0.015	0.231	0.065	0.782
	social	0.933	0.938	0.021	0.289	0.056	0.791
	application	0.667	0.75	0.009	0.215	0.044	0.822
	sentiment	0.867	0.882	0.017	0.273	0.057	0.808
	insight	0.733	0.789	0.009	0.237	0.037	0.855
	topic	0.867	0.882	0.024	0.267	0.034	0.756
	media	0.733	0.789	0.011	0.234	0.04	0.818
	tool	0.733	0.789	0.012	0.235	0.044	0.818
	classification	0.733	0.789	0.012	0.234	0.034	0.818
	business	1	1	0.032	0.303	0.054	0.752
	textual	0.733	0.789	0.01	0.235	0.04	0.836

- ① **Ego Network Properties**
Provides six types of ego network metrics as analysis results.

Size
Ties
Pairs
Ego-Density
Broker
Ego Betweenness

- ② **Ego Network Centrality**
Displays five types of centrality metrics and the clustering coefficient as analysis results. The centrality metrics are as follows.

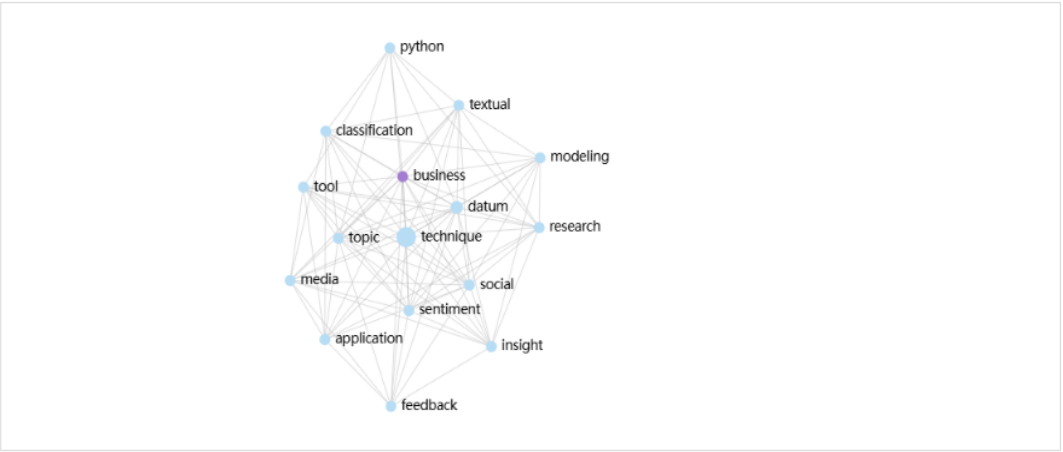
Degree Centrality
Closeness Centrality
Betweenness Centrality
Eigenvector Centrality
PageRank

- ③ **Download Analysis Result**
Download the analysis results as an Excel (.xlsx) file.

Ego Network Analysis Results

1 Ego Network Properties Ego Network Centrality Visualization Result 3 4

View Download



2 Visualization Settings

Color Selection

Ego Node Alter Node Connection Line

Text Settings

Color Font Size 14

Node Settings

Size Frequency Criteria

- ① **Visualization Result**
The visualization result of ego network is provided as an undirected network graph.
Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
Tip. When set based on word frequency, frequently occurring words are displayed as larger nodes. When set based on betweenness centrality, words that play a more important role in the network are displayed as larger nodes.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.
- ④ **Download Visualization**
Download the visualization result as a PNG file.

TEXTOM Data Analysis Console

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network **Shortest path** CONCOR Clustering

1 Analysis Algorithms

Shortest path

Destination word
Network analysis is generated based on the [Matrix] 1-Mode selected word.

2 ▶ Starting word Please enter the starting word.

3 ▶ Destination word Please enter the destination word.

▶ Analysis Word

Shortest path analysis result

4 Please select analysis options.

- 1 Selected Algorithm Description**
Click the button to view the description of the algorithm.
- 2 Enter the initial start(source) word and the destination word to check the path within the network.**
Tip. Be sure to enter words that exist within the network.
- 3 Select word**
Click [Select Directly] to navigate to the 1-Mode Matrix word selection page and select words for analysis.
Tip. A 1-Mode Matrix analysis must be conducted first.
- 4 Display the analysis results.**

Network Analysis

The results of Shortest Path Analysis.

- <46> -

TEXTOM Data Analysis Console

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network **Shortest path** CONCOR Clustering

Shortest path Analysis Algorithms

Destination word
※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

Starting word: topic
Destination word: media

1 Analysis Word **4** Download

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
processing	26	0.517	33.565
insight	26	0.517	34.082
trend	25	0.497	34.579
topic	22	0.437	35.494
network	21	0.418	36.767
media	21	0.418	36.349
tool	18	0.358	37.9

2 << Reselect **3** Apply

Shortest path analysis result

5 Starting word **6** Download

Starting word	Destination word	Shortest path length	Path
topic	media	2	topic,technique,media topic,datum,media topic,social,media topic,application,media topic,sentiment,media topic,insight,media topic,trend,media topic,tool,media topic,classification,media topic,business,media topic,textual,media topic,feedback,media

- Review analysis word list.
- Reselect Words**
Click [Reselect] to navigate to the 1-Mode Matrix word selection page and reselect.
Tip. A 1-Mode Matrix analysis must be conducted first.
- Apply the selected words to the analysis.
- Download Analysis Words**
Download the selected words as an Excel (.xlsx) file.
- Shortest Path Analysis Results**
The result display the shortest path length and detailed paths.
- Download Analysis Result**
Download the analysis result as an Excel (.xlsx) file.

Shortest path analysis result

Shortest path Visualization Result

1

3 4
View Download

2 Visualization Settings

▸ Color Selection

Ego Node Alter Node Connection Line

▸ Text Settings

Color Font Size 14

▸ Node Settings

Size Frequency Criteria

- ① **Visualization Result**
The results of analysis are presented in a network chart.
Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
Tip. When set based on word frequency, frequently occurring words are displayed as larger nodes. When set based on betweenness centrality, words that play a more important role in the network are displayed as larger nodes.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.
- ④ **Download Visualization**
Download the visualization result as a PNG file.

TEXTOM Data Analysis Console

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network Shortest path **CONCOR** Clustering

CONCOR 1 Analysis Algorithms

It is an analysis technique that clusters words based on their structural equivalence by analyzing the relationships between words that co-occur simultaneously. ※
Network analysis is generated based on the [Matrix] 1-Mode selected word.
※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

2 Cluster Number 4

3 Analysis Word Select Directly

CONCOR analysis result

CONCOR Visualization Result

4 Please select analysis options.

① **Selected Algorithm Description**
Click the button to view the description of the algorithm.

② Select Number Of Cluster.

③ **Select word**
Click [Select Directly] to navigate to the 1-Mode Matrix word selection page and select words for analysis.
Tip. A 1-Mode Matrix analysis must be conducted first.

④ Display the analysis results.

TEXTOM Data Analysis Console 빅데이터팀 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network Shortest path **CONCOR** Clustering

CONCOR Analysis Algorithms

It is an analysis technique that clusters words based on their structural equivalence by analyzing the relationships between words that co-occur simultaneously. ※ Network analysis is generated based on the [Matrix] 1-Mode selected word. ※ Network analysis is generated based on the [Matrix] 1-Mode selected word.

Cluster Number 4

Analysis Word **4** Download

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
technique	322	6.403	22.171
datum	144	2.863	25.035
research	35	0.696	28.932
social	33	0.656	29.588
application	32	0.636	30.225
sentiment	32	0.636	30.861
processing	26	0.517	33.565

2 << Reselect **3** ✓ Apply

CONCOR analysis result

CONCOR Visualization Result **6** 다운로드

Cluster	Word
5 1	network
1	tool
1	classification
1	business
1	algorithm
1	pattern
2	textual
2	python
2	feedback
2	university
2	modeling
3	sentiment

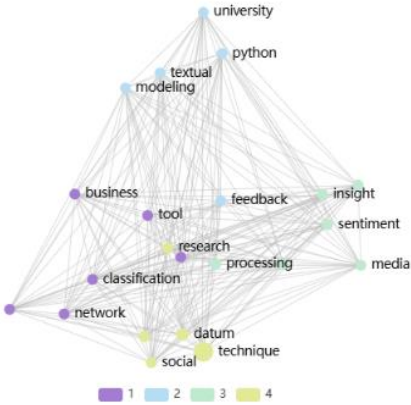
- Review analysis word list.
- Reselect Words**
Click [Reselect] to navigate to the 1-Mode Matrix word selection page and reselect.
Tip. A 1-Mode Matrix analysis must be conducted first.
- Apply the selected words to the analysis.
- Download Analysis Words**
Download the selected words as an Excel (.xlsx) file.
- Shortest Path Analysis Results**
The results display the words and the groups to which they belong.
- Download Analysis Result**
Download the analysis result as an Excel (.xlsx) file.

CONCOR analysis result

CONCOR Visualization Result

View Download

1



2 Visualization Settings

Color Selection

Cluster1 Cluster2 Cluster3 Cluster4 Connection Line

Text Settings

Color Font Size 14

Node Settings

Size Frequency Criteria

- ① **Visualization Result**
The results of analysis are presented in a network chart.
Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
Tip. When set based on word frequency, frequently occurring words are displayed as larger nodes. When set based on betweenness centrality, words that play a more important role in the network are displayed as larger nodes.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.
- ④ **Download Visualization**
Download the visualization result as a PNG file.

TEXTOM Data Analysis Console

백데이터팀 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network Shortest path CONCOR **Clustering**

1 Analysis Algorithms

This is a hierarchical clustering method for identifying small-scale groups within a network. It calculates the patterns of relationships that each node forms with other nodes, classifying nodes with similar relationship patterns into the same cluster.
Network analysis is generated based on the [Matrix] 1-Mode selected word.

2 Analysis Model: Louvain

3 Analysis Word: Select Directly

4 Please select analysis options.

Clustering analysis result

Clustering Visualization Result

① **Selected Algorithm Description**
Click the button to view the description of the algorithm.

② **Select Analysis Model**
Clustering analysis offers the following four algorithms:

- Louvain
- Leiden
- Girvan Newman (GN)
- Clauset-Newman-Moore

Tip. For quick analysis: Louvain
 For high accuracy: Leiden
 For simple network structures: GN
 For hierarchical structures: Clauset-Newman-Moore

③ **Select word**
Click [Select Directly] to navigate to the 1-Mode Matrix word selection page and select words for analysis.
 Tip. A 1-Mode Matrix analysis must be conducted first.

④ Display the analysis results.

TEXTOM Data Analysis Console 백데이터팀 LOG OUT

Word Analysis Matrix **Network Analysis** Topic Analysis Sentiment Analysis QAP Hypothesis test

Network Attribute Centrality Ego Network Shortest path CONCOR **Clustering**

Clustering Analysis Algorithms

This is a hierarchical clustering method for identifying small-scale groups within a network. It calculates the patterns of relationships that each node forms with other nodes, classifying nodes with similar relationship patterns into the same cluster.
* Network analysis is generated based on the [Matrix] 1-Mode selected word.

▶ Analysis Model Louvain

1 ▶ Analysis Word **4** Download

Word	Frequency (Count)	Percentage (%)	Cumulative Percentage(%)
technique	322	6.403	22.171
datum	144	2.863	25.035
research	35	0.696	28.932
social	33	0.656	29.588
application	32	0.636	30.225
sentiment	32	0.636	30.861

2 << Reselect **3** ✓ Apply

Clustering analysis result

Clustering Visualization Result **6** Download

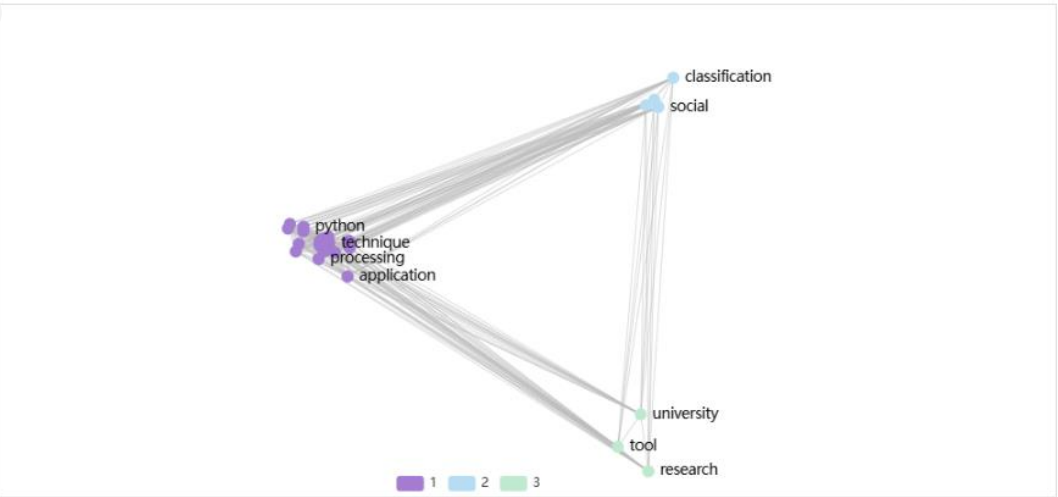
Cluster	Word
5 1	technique
1	datum
1	application
1	sentiment
1	processing
1	insight
1	trend
1	topic
1	media
1	network
1	business
1	algorithm

- Review analysis word list.
- Reselect Words**
Click [Reselect] to navigate to the 1-Mode Matrix word selection page and reselect.
Tip. A 1-Mode Matrix analysis must be conducted first.
- Apply the selected words to the analysis.
- Download Analysis Words**
Download the selected words as an Excel (.xlsx) file.
- Clustering Analysis Results**
The results display the words and the clusters to which they belong.
- Download Analysis Result**
Download the analysis result as an Excel (.xlsx) file.

Clustering analysis result

Clustering Visualization Result

1



2 Visualization Settings

▶ Color Selection

Cluster1 Cluster2 Cluster3 Connection Line

▶ Text Settings

Color Font Size 14

▶ Node Settings

Size Frequency Criteria

3 4

View Download

① Visualization Result

The results of analysis are presented in a network chart.

Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.

② Visualization Setting

Users can modify the attributes of the visualization results as desired.

Tip. When set based on word frequency, frequently occurring words are displayed as larger nodes. When set based on betweenness centrality, words that play a more important role in the network are displayed as larger nodes.

③ Zoom In

Click [View] to display an enlarged version of the selected visualization in a pop-up window.

Tip. You can zoom in and out of the visualization by scrolling the mouse wheel up and down.

④ Download Visualization

Download the visualization result as a PNG file.

Topic Analysis

Check Optimal Topic Performance before Conducting LDA Analysis.

- <54> -

TEXTOM Data Analysis Console

Word Analysis Matrix Network Analysis

Optimal Topic Performance LDA Topic Modeling

Optimal Topic Performance 1 [Analysis Algorithms](#)

We calculate Perplexity and Coherence scores to determine the optimal number of topics.
 ※Perplexity: A measure of model accuracy, where lower scores indicate higher accuracy.
 ※Coherence: A measure of semantic consistency within topics, where higher scores indicate higher coherence.

2 Parameter Selection Default User-defined

Alpha 0.5

Beta 0.01

Iterations 50

Number of Topics 10

Number of Words per Topic 30

Random Value Use Not Use

3 ✓ Apply

Parameter Selection Default User-defined

Alpha 0.5

Beta 0.01

Iterations 50

Number of Topics 10

Number of Words per Topic 30

Random Value Use Not Use

✓ Apply

Alpha	A hyperparameter that adjusts the distribution of topics in documents
Beta	A hyperparameter that adjusts the distribution of words in topics
Iterations	The number of repetitions during model training
Number of Topics	The number of groups you want to create
Number of Words per Topic	The number of words to be included in each group
Random Seed	Proceed with random assignments (however, it is difficult to reproduce or verify the results of the experiment)

1 Selected Algorithm Description
Click the button to view the description of the algorithm.

2 LDA Analysis Parameters
Before performing LDA analysis, you can set the parameters to their default values or choose custom settings to achieve optimal topic performance.
 Tip. It is recommended to use the default values. After reviewing the topic analysis results, adjust the parameters through custom settings if necessary.

3 Apply the selected options to proceed with the analysis.

Topic Analysis

Display the Optimal Topic Results.

- <55> -

Optimal Topic Performance

We calculate Perplexity and Coherence scores to determine the optimal number of topics.
 ※ Perplexity: A measure of model accuracy, where lower scores indicate higher accuracy.
 ※ Coherence: A measure of semantic consistency within topics, where higher scores indicate higher coherence.

Parameter Selection: Default (selected) / User-defined

Alpha: 0.5
 Beta: 0.01
 Iterations: 50
 Number of Topics: 10
 Number of Words per Topic: 30
 Random Value: Use (selected) / Not Use

Optimal Topic Results

Performance Calculation Results (1) | Visualization Result (2) | Download

Ct.TPpn	Perplexity Score	Coherence Score
2	-7.750030718951168	-5.43802073366585
3	-8.280960263714329	-5.945381255671292
4	-8.81498748152829	-6.158824315179703
5	-9.308894746353092	-6.350951218461269
6	-9.762512090275306	-6.270810401181756
7	-10.200287237161751	-6.443355753172101
8	-10.63743854028157	-6.24344134667472
9	-11.063876726460522	-6.308784512643645
10	-11.46492159386468	-6.254293310636997

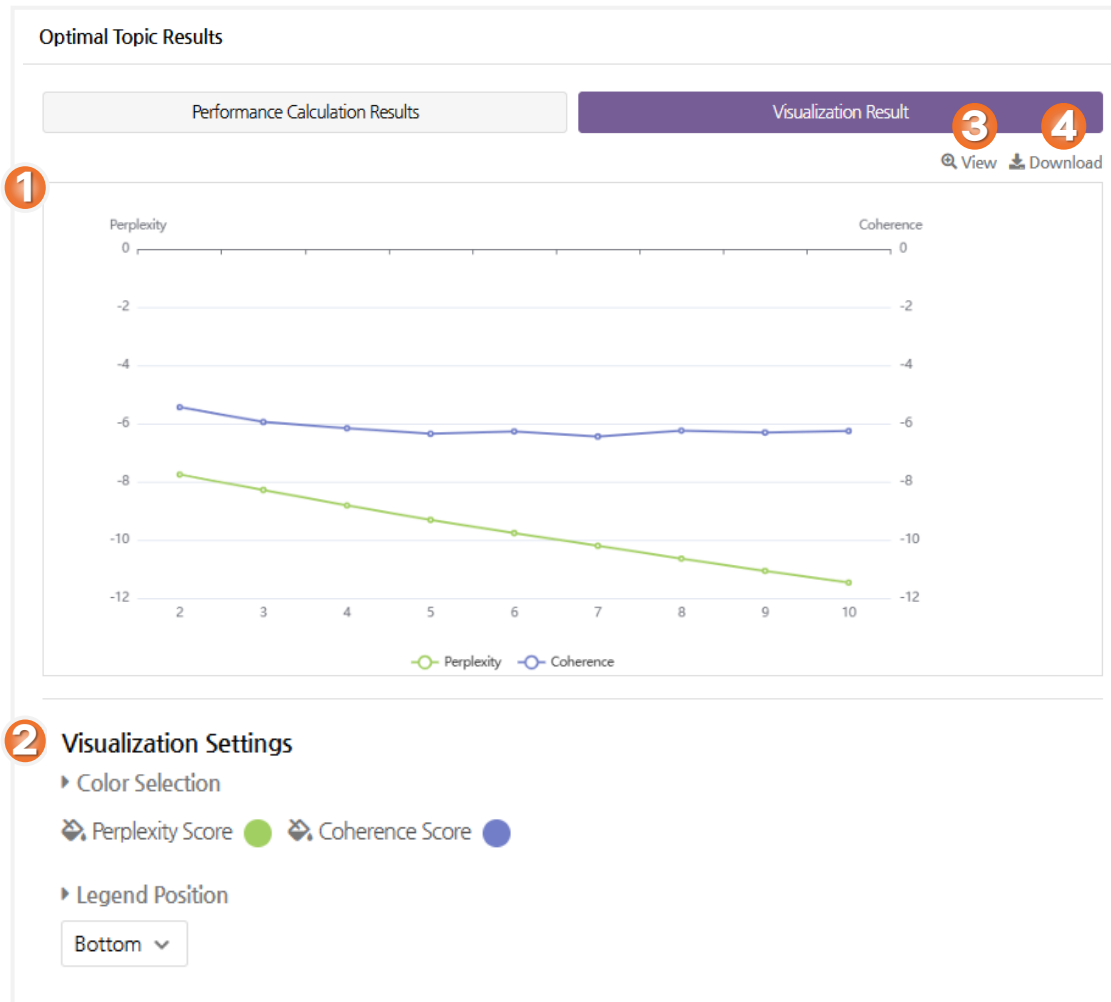
Apply

① **Topic Performance Analysis Results**
 The following metrics are calculated as topic performance results. Generally, a lower Perplexity value and a higher Coherence value indicate the optimal number of topics.

- **Perplexity Score**
 A metric for the predictive performance of the topic model; lower values indicate higher predictive accuracy.
- **Coherence Score**
 A metric for evaluating the semantic consistency of words within a topic; higher values indicate more coherent topics.

Tip. After determining the optimal number of topics based on the topic performance analysis results, reset the LDA parameters to this number of topics and proceed with the analysis.

② **Download Analysis Result**
 Download the analysis result as an Excel (.xlsx) file.



- ① **Visualization Result**
The results of analysis are presented in a line chart.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
- ④ **Download Visualization**
Download the visualization result as a PNG file.

Topic Analysis

The results of LDA Topic Modeling.

- <57> -

TEXTOM Data Analysis Console

Word Analysis Matrix Network Analysis **Topic Analysis** Sentiment Analysis QAP Hypothesis test

Optimal Topic Performance LDA Topic Modeling

LDA Topic Modeling Analysis Algorithms

Topic analysis groups keywords with similar meanings into topics, allowing for the exploration of the latent semantic structure of the entire dataset.

Parameter Selection Default User-defined

Alpha 0.5

Beta 0.01

Iterations

Number of Topics

Number of Words per Topic

Random Value

LDA Topic Modeling Results

1 Topic Words Topic Documents LDAvis

Topic Classification	Words Included in the Topic	Cl, TMwp
1	text	0.08
1	mining	0.078
1	technique	0.059
	datum	0.049
	social	0.015
	media	0.012
	analysis	0.011
	science	0.008
	trend	0.008
	learning	0.007
	review	0.007
	analytic	0.007

2 Topic Documents

Cmm, TMdo	Topic Classification	Probability of Topic Inclusion
text mining natural language processing nlp broad field text mining technique computer	1	0.394518
introduction r datum mining datum datum mining unstructured text text mining technique valuable information textual source such social media	1	0.68466854
similar pattern potential benefit diachronic text mining technique large dynamic text corpus drug	9	0.67909384
impact social media information public dual carbon policy context novel methodological framework elm text mining technique approach	5	0.30820972
comparative analysis web tool low text mining technique istat survey information communication technology enterprise European conference quality official	10	0.64433706
much understandable patient information leaflet study text mining technique nlp artificial intelligence tool drug effectiveness problem	7	0.62764347
emotional intelligence attention unsupervised learning comparison semi text mining technique irony Greek political tweet	1	0.72588754
artificial intelligence voting advice application automatic content analysis legislative document text mining technique 48th hawaii international conference system science	8	0.44991785

3 Download

- ① **Topic Word Analysis Results**
Calculate the extracted topics and the topic probability(%) of words from the topic word analysis.

Tip. Word topic probability indicates the likelihood of a word being assigned to a topic.

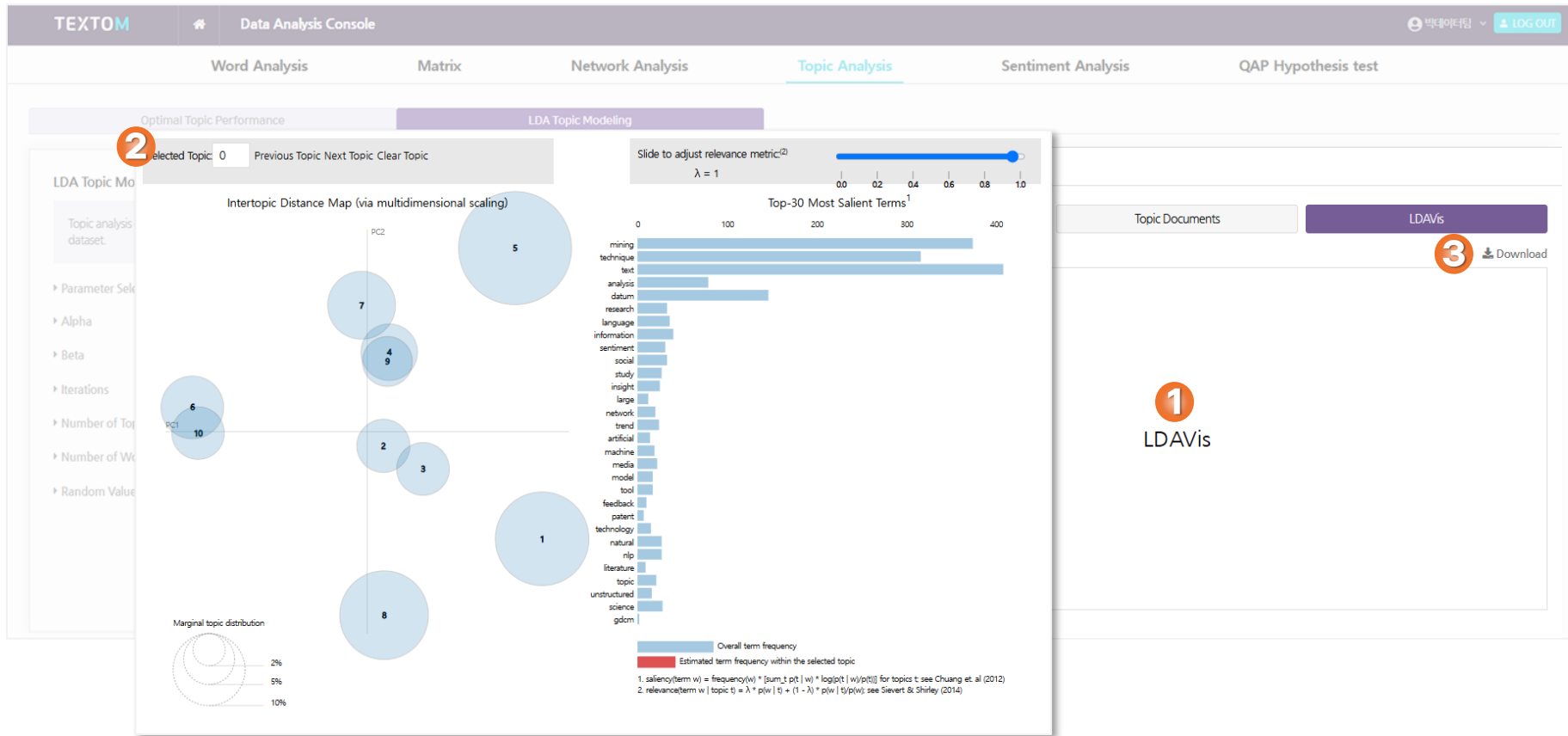
- ② **Topic Document Analysis Results**
Calculate the original texts assigned to the extracted topics and their topic probability(%) from the topic document analysis.

- ③ **Download Analysis Result**
Download the analysis result as an Excel (.xlsx) file.

Topic Analysis

The visualization results of LDA Topic Modeling.

- <58> -



① Click this text to open the visualization in a pop-up.

② **Visualization Results**
Use clicks and cursor movement to view the word distribution for each topic and the similarities between topics.

- **Circle Size** : Represents the word distribution of the topic.
- **Circle Distance** : Indicates the similarity between topics.
- λ : Adjusts the distinctiveness between topics. Closer to 1 shows words that frequently appear in one topic, while closer to 0 shows words that do not frequently appear in other topics.

Tip. The visualization shows the top 30 keywords for each topic, so it is recommended to carefully refine words (e.g., single characters, numbers, etc.).

② **Download Analysis Result**
Download the analysis result as an HTML file.

TEXTOM Data Analysis Console 빅데이터팀 LOG OUT

Word Analysis Matrix Network Analysis Topic Analysis **Sentiment Analysis** QAP Hypothesis test

Document classification Sentiment Word

Document Classification Sentiment Analysis 1 Analysis Algorithms

Utilizing CoreNLP, detect the inherent emotions in sentences and classify the sentiment of documents as positive, neutral, or negative.

Document Classification Sentiment Analysis Results

Overall Document Sentiment Visualization Result 3 다운로드

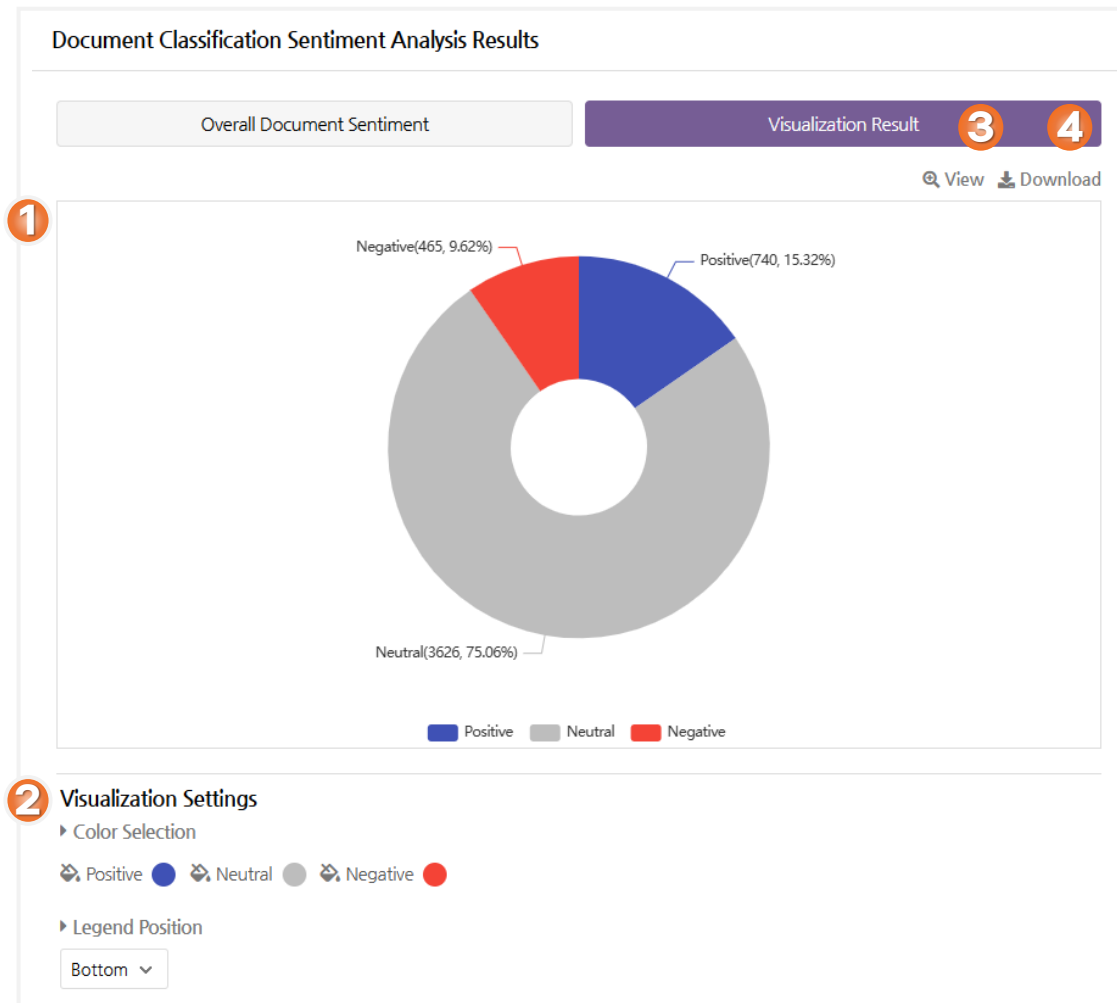
Category	Frequency (Count)	Ratio (%)
Total	4831	100
Positive	740	15.32
Neutral	3626	75.06
Negative	465	9.63

4 Total Positive Neutral Negative

5 Original Text 6 Download

Original Text	Sentiment Classification
gran company plan production russia company	Neutral
technopolis stage area less square meter order company computer technology telecommunications statement	Neutral
international electronic industry company elcoteq ten employee tallinn facility contrary early layoff company rank office worker daily postimee	Neutral

- Selected Algorithm Description**
Click the button to view the description of the algorithm.
- Summary of Overall Classification Result**
Generate a summary of the frequency and percentage(%) of total/positive/neutral/negative documents from the comprehensive sentiment analysis results.
- Download the comprehensive sentiment analysis results as an Excel (.xlsx) file.**
- Sentiment Classification Tab**
Click the Sentiment tab to view the classification results of original texts by sentiment.
- View Sentiment Classification Results**
Check the classification results of original texts by sentiment.
- Download the Sentiment Classification result as an Excel (.xlsx) file.**



- ① **Visualization Result**
The results of analysis are presented in a pie chart.
- ② **Visualization Setting**
Users can modify the attributes of the visualization results as desired.
- ③ **Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
- ④ **Download Visualization**
Download the visualization result as a PNG file.

TEXTOM Data Analysis Console 백데이터팀 LOG OUT

Word Analysis Matrix Network Analysis Topic Analysis **Sentiment Analysis** QAP Hypothesis test

Document classification **Sentiment Word**

Sentiment lexicon analysis 1 Analysis Algorithms

CoreNLP to identify sentiment vocabulary in the text and calculate scores based on the frequency and intensity of occurrence.

Sentiment Vocabulary Analysis Results

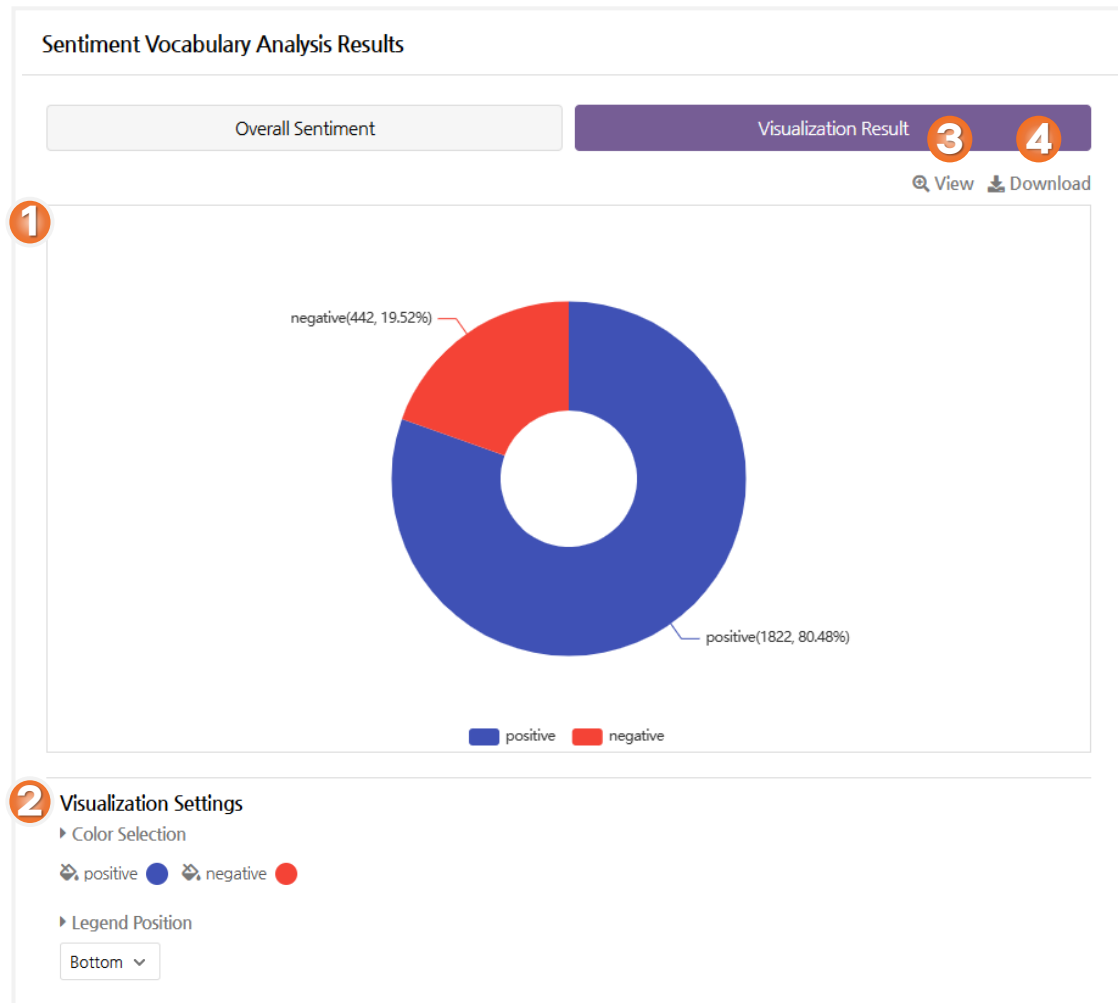
Overall Sentiment Visualization Result 3 Download

Emotion Classification	Frequency (Count)	Sentiment Intensity Ratio (%)	Frequency Ratio (%)
Total	2264	100	100
positive	1822	151.23	80.48
negative	442	-51.23	19.52

4 Total Positive Negative 6 Download

Emotion Classification	Sentiment Vocabulary	Sentiment Intensity	Frequency (Count)	Sentiment Intensity Ratio (%)	Frequency Ratio (%)
positive	share	1	432	432	19.08
positive	solution	1	172	172	7.60
negative	loss	-3	160	-480	7.07
positive	agreement	1	124	124	5.48

- Selected Algorithm Description
Click the button to view the description of the algorithm.
- Summary of Overall Sentiment Results
View the frequency and percentage results.
- Download the comprehensive sentiment analysis results as an Excel (.xlsx) file.
- Detailed Sentiment Tab
Click the tab for each detailed sentiment to view the sentiment vocabulary.
- View Sentiment Classification Results
View the vocabulary, sentiment intensity, frequency, and percentage for each detailed sentiment.
- Download the detailed sentiment vocabulary analysis result as an Excel (.xlsx) file.



- Visualization Result**
The results of analysis are presented in a pie chart.
- Visualization Setting**
Users can modify the attributes of the visualization results as desired.
- Zoom In**
Click [View] to display an enlarged version of the selected visualization in a pop-up window.
- Download Visualization**
Download the visualization result as a PNG file.

TEXTOM Data Analysis Console 백데이터팀 LOG OUT

Word Analysis Matrix Network Analysis Topic Analysis Sentiment Analysis QAP Hypothesis test

QAP Correlation analysis QAP Regression Analysis

1 Analysis Algorithms

QAP Correlation analysis

Correlation Analysis of Networks Based on Correlation Coefficients for Two or More Matrices Composed of the Same Words.
 ※ Pearson Correlation Coefficient: An indicator of the linear relationship between two variables, ranging from -1 to 1.
 - Closer to -1 indicates a negative correlation, while closer to 1 indicates a positive correlation.
 ※ Jaccard Coefficient: A measure indicating the proportion of common nodes between two sets, ranging from 0 to 1.
 - Closer to 1 implies that the two sets are similar.
 ※ Euclidean Coefficient: A measure indicating the distance between two nodes in the network, with closer values to 0 indicating closer proximity between two points.

2 Matrix

Please upload the matrix file.

Please upload the matrix file. +

Please upload two or more 1-Mode matrix files to verify correlations.
 ※ All matrices must be composed of the same words.

3 Parameter Selection

Pearson Correlation Coefficient Jaccard Coefficient Euclidean Distance

Please select a similarity (dissimilarity) metric for measuring correlation.

Permutation

4 Apply

QAP Correlation Analysis Results

QAP Correlations QAP P-Values

5

Please upload the file for analysis.

- Selected Algorithm Description**
Click the button to view the description of the algorithm.
- Upload the Matrix File**
Upload the matrix for QAP analysis. To check correlations, please upload at least two matrices.
- Select Parameters**
Choose [Parameter Selection] and [Permutation] from the parameters.
Tip. Permutation involves randomly rearranging the nodes of the matrix and repeating this process multiple times. It is recommended to perform this at least 1,000 times to evaluate statistical significance.
- Proceed with the analysis using the uploaded files and selected conditions.
- Display the analysis results.

QAP Correlation Analysis Results

1 QAP Correlations

QAP P-Values

2 Download

	data_01	data_02
data_01	1	0.2582237169469755
data_02	0.2582237169469755	1

Correlations

Values range between -1.0 and 1.0. A value close to 0 indicates no correlation between the two variables, while a higher value indicates greater similarity between the two networks.

P-Value (Significance Level)

Indicates the probability that the correlation coefficient appears by chance. The lower the value, the more statistically significant the correlation coefficient is considered. Typically, a P-Value below 0.05 is considered statistically significant.

3 QAP P-Values

4 Download

	data_01	data_02
data_01	-	1.7843941232785097e-15
data_02	1.7843941232785097e-15	-

① **QAP Correlation Analysis Results**
Check the QAP correlation analysis results.
Tip. Correlation coefficients only indicate linear relationships. If the relationship between two variables is non-linear, the correlation coefficient may appear low.

② **Download Correlation Analysis Results**
Download the analysis results as an Excel (.xlsx) file.

③ **QAP P-Value Analysis Results**
Choose [Parameter Selection] and [Permutation] from the parameters.
Tip. A high correlation coefficient and a low P-Value indicate a strong and statistically significant relationship between the two networks.

④ **Download P-Value Analysis Results**
Download the analysis results as an Excel (.xlsx) file.

To understand the similarities and differences between the Korean and overseas markets' preferences for Korean dramas, we conducted a QAP correlation analysis.

With a correlation coefficient of 0.2582, there is a weak positive correlation between the two networks. The extremely small P-Value of $2.927564572264821e-17$ indicates that the likelihood of this correlation occurring by chance is nearly zero.

TEXTOM Data Analysis Console 백데이터팀 LOG OUT

Word Analysis Matrix Network Analysis Topic Analysis Sentiment Analysis QAP Hypothesis test

QAP Correlation analysis QAP Regression Analysis

QAP Regression Analysis 1 Analysis Algorithms

This is a multiple regression analysis that evaluates the influence of independent variable matrices on the dependent variable matrix. It analyzes the causal relationships between two or more matrices composed of the same words.

2 Upload the dependent variable matrix
Please upload the file containing the dependent variable matrix.

Upload the independent variable matrix
Please upload the file containing the independent variable matrix.

Please upload the 1-Mode matrix file for causal relationships validation.
* Upload one dependent variable matrix and one or more independent variable matrices.
* All matrices should be composed of the same words.

3 Permutation

4 Apply

QAP Regression Analysis Results

QAP Parameters QAP Regression Coefficients

5 Please upload the file for analysis.

- ① **Selected Algorithm Description**
Click the button to view the description of the algorithm.
- ② **Upload the Matrix File**
Upload the matrices for QAP analysis. Please upload one dependent variable matrix and one or more independent variable matrices to check causal relationships.
Tip: The matrices must be 1-Mode matrices, and all matrices should be composed of the same words.
- ③ **Select Parameters**
Choose [Parameter Selection] and [Permutation] from the parameters.
Tip: Permutation involves randomly rearranging the nodes of the matrix and repeating this process multiple times. It is recommended to perform this at least 1,000 times to evaluate statistical significance.
- ④ Proceed with the analysis using the uploaded files and selected conditions.
- ⑤ Display the analysis results.

QAP Regression Analysis Results

QAP Parameters		QAP Regression Coefficients	
Category	Value		
Number of Words (N)	23		
R-Square	0.06167		
P-Value	0		

QAP Parameters

Number of Words (N)	Number of variables
R-Square	Indicates the coefficient of determination; the closer it is to 1, the better the model explains the data
P-Value	Indicates the statistical significance of the model; lower values mean that the independent variables of the model have a significant effect on the dependent variable

QAP Regression Coefficients

Unstandardized Coefficients	The effect of independent variables on the dependent variable
Standardized Coefficients	Comparison of the impact between independent variables
P-Value	Statistical significance of the model
Standard Error	Variability and reliability of the regression coefficients

QAP Parameters		QAP Regression Coefficients			
Independent Variable	Dependent Variable				
	data_01				
data_02		Unstandardized Coefficient	Standardized Coefficient	P-Value	Standard Error
		2.042576	0.248334	0	0.354899

- ① **QAP Parameters Analysis Results**
Check the results of the input values analyzed by QAP regression analysis.
- ② **Download QAP Parameters Analysis Results**
Download the analysis results as an Excel (.xlsx) file.
- ③ **QAP Regression Coefficients Results**
Check the results of the QAP regression coefficients analysis.
- ④ **Download QAP Regression Coefficients Analysis Results**
Download the analysis results as an Excel (.xlsx) file.

The screenshot displays the TEXTOM Customizing interface. On the left is a dark sidebar with navigation options: Dashboard, Collection, Data Collection, Collection List, Cleansing, Data preprocessing, Refinement List, Analysis, and Customizing. The main area is titled 'Customizing' and features a top navigation bar with tabs for Word Cloud, Bar Chart, Ego Network, Pie Chart, Line Chart, N-gram Network, 1-way Word Tree, Tree Map, Matrix, and Network. The 'Word Cloud' tab is selected. Below the tabs is a 'Visualization Result' section containing a word cloud with terms like 'Samsung Electronics', 'Busan World Expo', 'Davos', 'Busan', 'Origins', 'Expo', 'Support', 'President', 'Busan City', and 'World Exposition'. A 'Download' button is next to the visualization. To the right is an 'Upload Visualization File' section with a file input field and an 'Apply' button. Below that is a 'Visualization Settings' section with options for Select Shape (Direct Input), Select Font (Default), Select Size (Default), Select Color (High, Mid, Low), and Select Keyword. A '채팅문의' (Chatting Inquiry) button is at the bottom right.

- ① **Customizing Type**
Select the type of visualization the user wants to customize.

Word Cloud	N-gram Network
Bar Chart	1-way Word Tree
Ego Network	Tree Map
Pie Chart	Matrix
Line Chart	Network

- ② **Upload Visualization File**
Upload the data file suitable for the chosen type and click the [Apply] button.
Tip. Use [Download Example File] to check the data formatting method.

- ③ **Visualization Setting**
Users can modify the attributes of the visualization results as desired.

- ④ **Download Visualization**
Download the visualization result as a PNG file.

TEXTOM

Dashboard

Collection

Data Collection

Collection List

Cleansing

Data preprocessing

Refinement List

Analysis

Customizing

My Page

백데이터팀 ENG LOG OUT

1 Membership Information Payment History User Dictionary

ID	theimc_big	Name	빅데이터팀	Purpose of subscription	Internal
E-Mail	azurebeluga@theimc.co.kr	Cell Phone	+82 1054460996	Organization	더아이엠씨
Registration Date	2021-12-14	Capacity	102.16MB		

2 Modify information Withdrawal

3

Membership Information 4 Payment History User Dictionary

2024-05-23 ~ 2024-05-30 1 week 3 months 1 year check

30Cases

Number	Payment method	amount	item	Payment date	End datev	User ID
--------	----------------	--------	------	--------------	-----------	---------

Add Capacity

Information Page Site

채팅문의

- 1 View Member Information.
- 2 **Edit Information**
Edit your member information after entering your password.
- 3 **Membership Cancellation**
Cancel your membership by entering your ID and password.
- 4 **Payment History**
Set a period to view your payment history and related information.

TEXTOM

- Dashboard
- Collection
- Data Collection
- Collection List
- Cleansing
- Data preprocessing
- Refinement List
- Analysis
- Customizing

Add Capacity

Information Page Site

User Dictionary

Membership Information | Payment History | **User Dictionary**
백데이터팀 ENG LOG OUT

2 Search Words Result 10 / 10

3 C:\fakepath\excel_example (1).xlsx Apply to all

Download Example File

Column A: Words to change / Column B: Corrected words

4 Word to Change → Edit Word 7 Change

5 10Cases Total Use Unuse Move 6 Download Delete Selected items

<input type="checkbox"/>	Number	Word to Change	Edit Word	Registration date	User Status
<input type="checkbox"/>	10	cancer_survivor_guidelines	cancer-survival	2024-05-30 13:51:26	<input checked="" type="checkbox"/>
<input type="checkbox"/>	9	cancer_survivors	cancer-survival	2024-05-30 13:51:26	<input checked="" type="checkbox"/>
<input type="checkbox"/>	8	carbohydrate_intake	carbohydrate	2024-05-30 13:51:26	<input checked="" type="checkbox"/>
<input type="checkbox"/>	7	carotenoid	carotenoid	2024-05-30 13:51:26	<input checked="" type="checkbox"/>
<input type="checkbox"/>	6	casein	casein	2024-05-30 13:51:26	<input checked="" type="checkbox"/>
<input type="checkbox"/>	5	cd107a	CD107a	2024-05-30 13:51:26	<input checked="" type="checkbox"/>
<input type="checkbox"/>	4	cd1d	CD1D	2024-05-30 13:51:26	<input checked="" type="checkbox"/>

If there are multiple lists of replacement words for the same word being used, the most recently registered replacement word will be applied.

<< < 1 > >>

1 Select group

Group name: TEST Edit

+ Create new group Delete dictionary

채팅문의

- ① **Group Selection and Management**
Select a group to manage the user dictionary. Click the [Edit] button to change the group name.
- ② **Search Words**
Search for words to view the list of modified words that include the searched word.
Tip: Use this to check for duplicate words.
- ③ **File Upload**
Upload a user dictionary file that you have created and click the [Apply All] button. Use the [Download Sample File] to check the formatting method.
- ④ **Enter Words**
Input the words to be changed and the modified words.
- ⑤ **Select List View and Usage Options**
Choose the method for viewing the word list and whether to use the words.
- ⑥ **Download User Dictionary**
Download the user dictionary as an Excel (.xlsx) file.
- ⑦ **Delete Modified Words**
Select the checkboxes to delete modified words.

Add Capacity

Add the data capacity needed for analysis.

- <70> -

TEXTOM

Dashboard

Collection

Data Collection

Collection List

Cleansing

Data preprocessing

Refinement List

Analysis

Customizing

Add Capacity

Information Page Site

채팅문의

로그아웃

ENG

빅데이터팀

Add Capacity

General

Select data capacity/usage period

Selected data capacity and amount 1 MB 3Months \$ 4.9

1 Data Capacity - 1 + MB

Please select the additional data capacity (MB).
Data capacity can only be purchased in whole numbers. (Decimal points are not applied.)

Period of use 3Months 6Months 9Months 12Months

Please select the duration for using the data.

2 Amount of payment

Product amount \$ 4.9

Product amount 5% discount over 450 USD - \$ 0

Final payment amount \$ 4.9

PayPal

Purchase instructions

- 5% discount is applied for purchases of 450 USD / 10% discount is applied for purchases of 1,000 USD
- The 'collection unit' is available for use with purchases of 10MB or more.
- The estimated data per document is approximately 3KB for full collection and about 0.4KB for summarized collection. On average, 1,000 summarized blog, cafe, and news posts amount to 300 to 500KB. (1MB = 1024KB)
- A minimum of 30 MB and a maximum of 400 MB are used for one paper, and this amount of data can vary greatly depending on the collection keywords, period, and channel that the researcher sets to suit the research topic.
- The purchased data can be used until the end of the usage period selected at the time of purchase, and any remaining data will expire when the usage period ends.
- Please note that it is not possible to extend the usage period or recover expired data, so make sure to carefully decide on the capacity and usage period.
- Please contact us separately for payment of more than 2 million won and establishment of your own texture. textom.global@gmail.com or chatbot

Data usage guide

- The system deducts data from the remaining data owned by the individual for the amount of data used for analysis.
- Data usage is not deducted at the [Data Collection].
- However, in order to analyze the collected data, when going through the [refining/morphological analysis] stage and moving on to the [text mining] of the data analysis team, the amount of the collected data is deducted. It's possible.

Purchase Data Refund Policy

- After purchasing the data, refunds are possible within one month.
- However, if any portion of the purchased data has been used, even 1KB, refunds will not be provided.
- You can proceed with data collection for free. Please check the collected capacity and make a purchase accordingly.

Additional information

- According to PayPal policy, payments cannot be made to Korean accounts. Therefore, Korean users have a different payment system.
- You will be charged the amount converted from USD to KRW.

- ① **Select Data Capacity And Usage Period**
Choose the additional data capacity and usage period.
TIP. Capacity can be purchased in 1MB increments, and the usage period can be selected in 3-month, 6-month, 9-month, or 12-month intervals.

- ② **Estimated Payment Amount**
The estimated payment amount for the selected purchase options is calculated.
TIP. For payments over 1,500 USD and setting up your own TEXTOM, please contact us.