

# Association Modeling on Keyword and Abstract Data in Korean Port Research\*

Hee-Young Yoon

Department of Tax Accounting, Soongeui Women's College, South Korea

Il-Youp Kwak<sup>†</sup>

Department of Applied Statistics, Chung-Ang University, South Korea

JKT 24(5)

Received 27 February 2020

Revised 10 April 2020

Accepted 7 August 2020

## Abstract

**Purpose** – This study investigates research trends by searching for English keywords and abstracts in 1,511 Korean journal articles in the Korea Citation Index from the 2002–2019 period using the term “Port.” The study aims to lay the foundation for a more balanced development of port research.

**Design/methodology** – Using abstract and keyword data, we perform frequency analysis and word embedding (Word2vec). A t-SNE plot shows the main keywords extracted using the TextRank algorithm. To analyze which words were used in what context in our two nine-year subperiods (2002–2010 and 2010–2019), we use Scattertext and scaled F-scores.

**Findings** – First, during the 18-year study period, port research has developed through the convergence of diverse academic fields, covering 102 subject areas and 219 journals. Second, our frequency analysis of 4,431 keywords in 1,511 papers shows that the words “Port” (60 times), “Port Competitiveness” (33 times), and “Port Authority” (29 times), among others, are attractive to most researchers. Third, a word embedding analysis identifies the words highly correlated with the top eight keywords and visually shows four different subject clusters in a t-SNE plot. Fourth, we use Scattertext to compare words used in the two research sub-periods.

**Originality/value** – This study is the first to apply abstract and keyword analysis and various text mining techniques to Korean journal articles in port research and thus has important implications. Further in-depth studies should collect a greater variety of textual data and analyze and compare port studies from different countries.

**Keywords:** Port Research, Text Mining, Word Embedding

**JEL Classifications:** F19, L99

## 1. Introduction

The definition of trade was originally limited to exchange of goods between parties, but the current broader definition includes all visible forms of trade (e.g., simple commodity exchanges) as well as invisible ones, such as exchange of technology, services, and capital movements.

Trade facilitates the transfer of goods from a location where they have lower economic value to a location where their value is higher. Ports are important trade locations and a key component in the international trade infrastructure. Thus, ports play a direct role in the economic development of a country by promoting international trade.

With the rapid development of information technology in recent years, researchers have begun using new analytical methods to find meaningful information in non-structural data,

\* This work was supported by the Chung-Ang University Research Grants in 2019. This paper utilizes the Korea Citation Index (KCI) DB provided by the Korea Research Foundation.

<sup>†</sup> Corresponding author: ikwak2@cau.ac.kr

© 2020 Korea Trade Research Association. All rights reserved.

including text, video, and audio. These new methods have allowed for the analysis of a wide variety of topics. Among them, text mining has become an important representative method, and researchers have employed it across a range of fields, including secretarial science, entrepreneurship, economics, statistics, bibliographical information, and construction.

In recent decades, studies have analyzed ports of importance in the national economy using different research methods to explore various issues. These studies have contributed greatly to our knowledge on port development. Domestic academic research on ports has also evolved considerably in both quantity and quality. Despite these developments and the convergence of various disciplines in port research, few studies have comprehensively and systematically classified and organized the results of all past research efforts.

Therefore, this study aims to investigate past and recent domestic studies on ports so as to inform future research directions. For this purpose, we examine English abstracts and keywords related to port research conducted in the last 18 years (2002–2019). Analyzing these data, we identify general trends and characteristics as well as terms highly correlated with major keywords.

In addition, this study visually analyzes the relationships between the main port-related keywords and compares and analyzes research topic trends in two nine-year sub-periods: 2002–2010 and 2011–2019. By analyzing 1,511 papers on ports from the Korea Citation Index, we systematically identify an intellectual structure for port research and lay the groundwork for future research in this field.

## 2. Literature Review

### 2.1. Trends in Port Research

In the early 2000s, researchers tended to view port research as a subtopic under shipping research rather than as a standalone topic, as revealed in analyses of trends in studies on shipping (Oh Se-Young, 2000) and shipping logistics (Yoon Il-Hyun, 2004). Helmick (2008) briefly reviewed the key determinants of research priorities in shipping security, suggested directions for future research, and discussed realistic agenda development. Pallis, Vitsounis and De Langen (2010) reviewed literature on port economy, policy, and management from 1997 to 2008 to quantitatively and qualitatively determine key characteristics of the port research.

The shipping subtopic included general ports, container terminals, and vessel entries and departures. Son Yong-Jung (2011) conducted a meta-analysis of the port economics research in Korea. The researcher subdivided topics and research methods in port economics and found that specific research methods were overly biased, while other methods were worthy of recommendation. Pallis et al. (2011) classified port research from 1997 to 2008 into seven research topics and presented research questions, concepts, and methods that are widely used for each topic.

Recently, the use of Social Network Analysis (SNA) has become a trend in port research. Jang Se-Eun and Lee Su-Ho (2016) used SNA to examine international journal articles and extracted keywords related to global shipping economics to identify key research topics in specific periods. Jeon Jun-Woo, Wang, and Yeo Gi-Tae (2016) analyzed the research trends of international port competition using SNA. Each keyword was presented for the following periods: 1980–1999, 2000–2007, and 2008–2015.

Son Yong-Jung (2017) used SNA to analyze the structural characteristics of co-author networks in studies on port economics. Meanwhile, Ma Hye-Min et al. (2017) quantitatively

studied port research trends, deriving the degree, betweenness, and closeness centrality in an analysis of 2,249 keywords from 589 articles in 58 domestic journals. Witte, Wiegmanns, and Ng (2019) reviewed over 80 academic articles on inland port development from 1992 to 2017. The authors suggested that although much attention was paid to inland ports as a component of the transportation, logistics, and supply chain systems, the role of inland ports as a component of the regional system was greatly overlooked.

The researchers confirmed that early research on ports simply focused on the ports themselves; however, as time passed, the focus shifted to combinations of topics or port-related sub-topics.

## 2.2. Trend Analysis Using Text Mining

Text mining involves deriving meaningful information from text data using natural language processing and statistical pattern learning. It consists of structuring, parsing, and deriving patterns from structured data and evaluating and interpreting text data. Common methods include text classification, text clustering, concept/entity extraction, emotional analysis, document summary, and entity relationship modeling. Studies in many academic fields have employed text mining to conduct analyses of academic knowledge structure, research trends, key words, and so on. Cho Su-Gon and Kim Seoung-Bum (2012) used data mining on abstracts from key international industrial engineering journals over 43 years, identifying the current status and trends in industrial engineering research. Choe Myoung-Ae et al. (2012) characterized the current status of Korean nursing research by classifying keywords and research methods (e.g., research topics and designs and data collection methods) in Korea and other East Asian countries. They then compared their analysis results with those from other countries. Bae Kyu-Yong et al. (2013) used text mining techniques to analyze the occurrence frequency of abstract terms in climate-related food papers. Importantly, text mining techniques can be useful in finding papers on similar themes before beginning a study and identifying whether a particular research topic has been addressed in recent journal issues.

Kim Gyu-Ha and Park Cheol-Yong (2015) used topic modeling to analyze the abstracts of papers published in the "Korean Journal of Data and Information Science." Topic modeling is an analytical method that is being actively researched. Park Jun-Hyung and Oh Hyo-Jung (2017) used topic modeling to analyze trends in domestic records management. They derived numerous special keywords to identify each topic's characteristics and presented an effective method for deriving topic micro keywords. Kim Se-Hyun (2018) analyzed abstracts using text mining to identify multicultural research trends in Korea. The results suggested that future multicultural studies should address the new academic discussions emerging around second-generation migrants. Lee Yong-Kyu, Yoon Soung-Woong and Lee Sang-Hoon (2018) analyzed defense research text and research topics using text network analysis, aiming to identify the overall knowledge structure. Analyzing the centrality and cohesiveness structure through controlling core stocks, they derived six knowledge structure groups in the defense sector. Jo Hye-In, Kim Jin-Woo and Lee Bong-Gyou (2019) analyzed topical research trends in the US, China, and Korea by extracting research keywords in the blockchain field using topic modeling. Park Jong-Do (2019) used topic modeling to analyze the domestic literature on multiculturalism; the author tracked core topics in multicultural research in Korea by period and identified the changes in the field over time. Lastly, Hwang Seo-I and Kim Mun-Ki (2019) used topic modeling and semantic network analysis to study associations between topics and abstracts, identify research trends in domestic artificial intelligence fields, and explore future directions. In doing so, they derived the main topics in artificial intelligence.

### 2.3. Differences from Existing Research

Trend analysis for existing port research mainly focused on research topics, research methods, and data analysis methods of previous studies. A number of studies have recently examined the academic connectivity between authors applying the SNA method. However, abstract data was not considered for analysis.

In other academic fields, text mining has been actively used to analyze research systems and trends using keywords and abstract data. However, in the port research field, a systematic research trend analysis through the text mining approach is insufficient.

This study used text mining to present the overall trends and characteristics of port research and the correlation between keywords and papers in the field was modeled using the entire English abstract data. Specifically, through the Word Embedding technique, abstract noun materials were vectorized while maintaining the semantic relationship between words. In the vectorized space, we examined noun words that showed high relevance to key keywords in the paper, and visually presented the research connection of the top 60 keywords among abstract words using the t-SNE technique. In Scattertext, the English abstract words used in the past and present were visualized by dividing them into 9-year units. Through Scattertext, we were able to examine the changes and flows of port research by determining frequently used words in the past and present.

## 3. Methods

Despite a growing body of literature on port research trends, not a single study examines the big data of paper abstracts. This section reviews research methods such as TextRank, word embedding, Scattertext, and t-SNE for analyzing text data from the abstracts of a large number of papers.

### 3.1. Data Collection

To identify studies on Korean ports, we select 1,511 papers on the Korea Citation Index website by January 2, 2020, collecting keywords, authors, journal names, subject areas, number of citations, and abstract data from these papers. From these, we select the 1,081 papers containing English abstract data and conduct further the following analyses on these selected papers.

### 3.2. Keyword Analysis

For the keyword analysis, we construct a model using Python's "numpy," "pandas," and "collections" modules that calculates the total number of articles, frequency by keyword, and number of articles per year. To improve the quality of the frequency analysis, we merge keywords with the same meaning (e.g., differing only in uppercase/lowercase, abbreviation use, or spacing).

### 3.3. Keyword Extraction Using TextRank

In addition to models that extract keywords chosen by authors, models also exist for extracting key words from document data. One of the most widely used methods in the latter category is the TextRank algorithm proposed by Mihalcea (2004). This algorithm is a graph-

based ranking model for text and extended PageRank, which is used in Google's search engine. PageRank is an algorithm proposed by Brin and Page (1998) that assigns weights to web documents with hyperlinks based on their relative importance. The ranking of websites is based on the fact that highly important websites are referenced by many other sites. The TextRank algorithm applies a similar method to text data. The formula is as follows:

$$WS(V_i) = (1 - d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_k), \quad (1)$$

where  $V_i$  is the vertex for the  $i$ th word;  $WS(V_i)$  is the TextRank value for vertex  $V_i$ ;  $w_{ij}$  is the weight on the edge between vertices  $V_i$  and  $V_j$ ;  $In(V_i)$  is the set of vertices that point to  $V_i$ ;  $Out(V_j)$  is the set of vertices to which vertex  $V_j$  points; and  $d$  is a damping factor that can be set between 0 and 1.

Using the TextRank algorithm, we extract 30 nouns of importance in port research papers and use a t-SNE plot to visualize the relationships among words along with the keywords selected by authors.

### 3.4. Word Embedding

The word embedding method vectorizes semantic relationships among words by machine learning algorithms. Specifically, each word is vectorized to a  $k$ -dimensional vector,  $\mathbf{x} = (x_1, x_2, \dots, x_k)$ , in the  $k$ -dimensional space,  $\mathcal{R}^k$ , to preserve the semantic relationship by the machine learning algorithm. Thus, when the vectors corresponding to "king," "queen," "man," and "woman" are  $\mathbf{x}_{king}$ ,  $\mathbf{x}_{queen}$ ,  $\mathbf{x}_{man}$ , and  $\mathbf{x}_{woman}$ , the semantic relationship of  $\mathbf{x}_{king} - \mathbf{x}_{man} + \mathbf{x}_{woman} \approx \mathbf{x}_{queen}$  is preserved. One of the most widely used word embedding methods is Word2vec, proposed by Mikolov et al. (2013).

#### 3.4.1. Preprocessing

Before performing word embedding, "cleaning" is necessary to break the text down into a format that the computer can easily understand. We remove punctuation and include frequently used bigrams employing the "Phrases" function in the "gensim.models" module.

#### 3.4.2. Word2vec Model

In Word2vec, the CBOW and skip-gram models have been proposed. CBOW models can determine, for a given word, which word best fits the context when the word position of the sentence containing the given word is blank. Meanwhile, the skip-gram model performs word embedding by considering which words appear before and after a given word.

In this study, we train words as 100-dimensional vectors in an  $\mathcal{R}^{100}$  space using a CBOW model through the "Word2Vec" function of Python's "gensim" module. Furthermore, using Word2Vec's "most similar" function, we examine cosine similarity for the nouns that are most relevant to the main keywords. We calculate the cosine similarity as  $\cos\theta = \frac{\mathbf{x}_1 \cdot \mathbf{x}_2}{|\mathbf{x}_1| |\mathbf{x}_2|}$ .  $\theta$  represents the angle between the two vectors,  $\mathbf{x}_1$  and  $\mathbf{x}_2$ . The value of the cosine similarity lies between -1 and 1 and is close to 1 when the two vectors have a positive correlation, close to 0 when the correlation is small, and close to -1 when the correlation is negative.

### 3.5. t-SNE

To visualize multidimensional data in the  $\mathcal{R}^{100}$  space, we need to carry out dimension reduction to a low-dimensional space. In addition to conventional linear dimension re-

duction techniques such as principal component analysis, researchers have used nonlinear dimension reduction techniques such as t-SNE in recent years. t-SNE, proposed by Maaten and Hinton (2008), tends to reveal clusters in data better than other methods, and thus has been used in many studies.

### 3.6. Scattertext

The Scattertext tool proposed by Kessler (2017) finds distinguishing terms and presents them in an interactive scatter plot with non-overlapping term labels. Scattertext is powerful for identifying terms that represent two contrasting concepts. We visualize the abstract terms from the research in the first subperiod (2002–2010) and second subperiod (2011–2019) in Fig. 5. Each term is represented as two numerical values (x, y) in an x/y plot if the x-axis represents the relative frequency of the 2011–2019 research while the y-axis represents that of the 2002–2010 research. Thus, a smaller distance from the top left implies a higher association with the 2002–2010 research, and a smaller distance from the lower right implies a higher association with the 2011–2019 research. In addition, Kessler (2017) proposed the scaled *F*-score to determine the important terms representing two contrasting concepts.

## 4. Results

### 4.1. Sample and Characteristics

As shown in Fig. 1, 1,151 valid samples are selected from a total of 1,155 papers over the 2002–2019 period (search date: January 2, 2020). The number of papers increases every year, from 36 publications in 2002 to 83 publications in 2011. Since 2011, the number of papers published every year has remained largely similar from 2012 to 2019.

Fig. 1. Number of Papers by Year

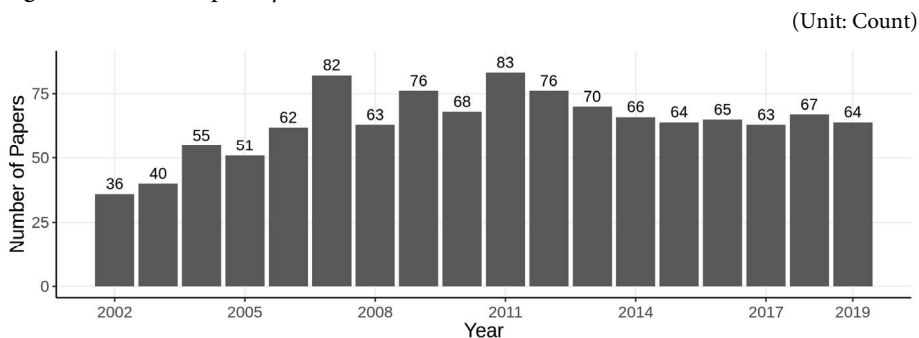
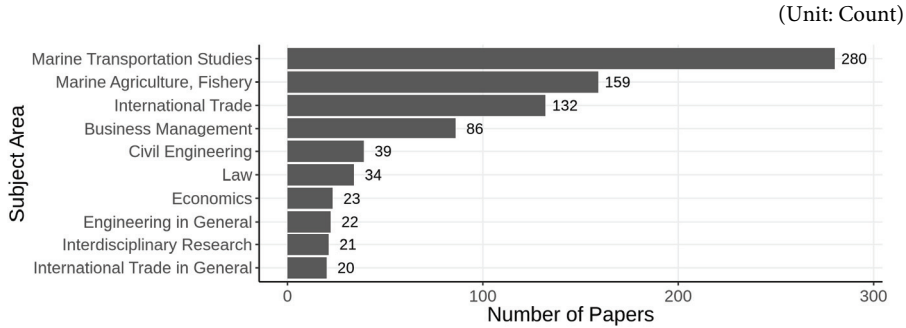


Fig. 2 shows the number of papers by subject area. We find that “Marine Transportation Studies” was the most studied area with 280 papers (24.33%), followed by “Marine Agriculture, Fishery” (159 papers; 13.81%) and “International Trade” (132 papers; 11.47%). Although all studies on ports are not represented in this figure due to space limitation, we find that 102 fields including “Fishery Science,” “Public Administration,” “Environmental Engineering,” and “Industrial Engineering” are relevant from an academic perspective.

**Fig. 2.** Number of Articles by Subject Area



Regarding the distribution of port-related papers by journal, we see from Fig. 3 that the “Journal of Korea Port Economic Association” has the highest number of papers at 197 (17.12%). Next are the “Journal of Korean Navigation and Port Research” and the “Journal of Shipping & Logistics” with 116 (10.08%) and 105 (9.12%) papers, respectively. The full sample contains 219 journals.

**Fig. 3.** Number of Articles by Journal

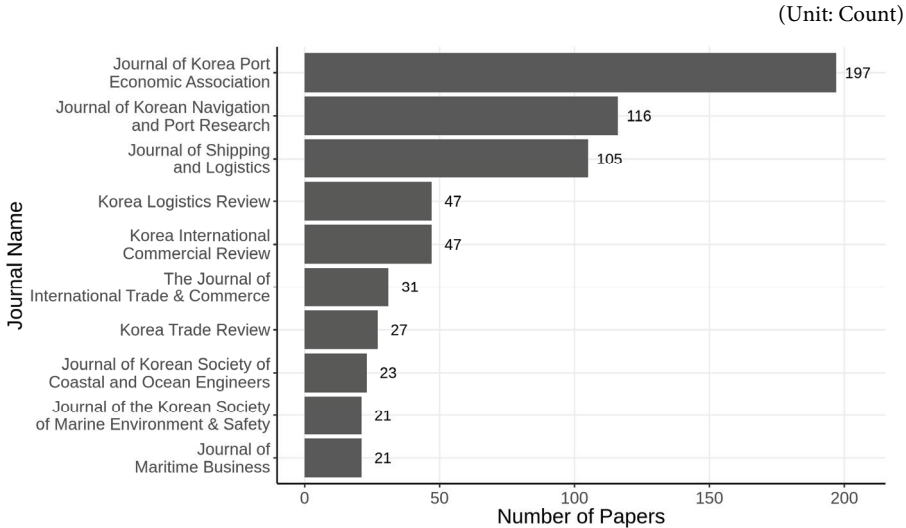


Table 1 lists the papers in the sample with the highest citation count. The most cited paper is “The Study of the Effect of Container Terminal Service Quality on Customer Satisfaction and Loyalty: The Moderating Effect of the Company Size and Port Location,” with 47 citations. The numbers and subject areas of highly cited papers reveal the study areas that researchers are most interested in. Specifically, studies citing the most frequently cited paper tend to focus on service quality and customer satisfaction related to ports and various other fields. Most studies citing the second-most frequently cited paper are concerned with the determinants of port competitiveness. Additional studies in other research areas focus on Busan New Port. The ranking of the cited papers suggests active research in this area.

**Table 1.** Highest Cited Papers

Rank	Cited by	Title	Author	Year
1	47	The Study of the Effect of Container Terminal Service Quality on Customer Satisfaction and Loyalty: The Moderating Effect of the Company Size and Port Location	Ahn Kwang-Ho, Kim Min-Sung and Kim Byung-Il	2008
2	42	A Study on the Determinants Analysis of the Port Competitiveness and Development Strategies of Busan New Port	Shin Ge-Seon	2007
2	42	Extraction of Port Selection Factors for Increasing Shippers' Attraction of Small and Medium Ports	Yeo Ki-Tae, Lee Hong-Girl and Oh Se-Woong	2004
4	36	A Study on Alternatives for Developing the Workers' Education Program in the Port Logistics Field using an NCS Model	Park Du-Jin and Kim Hyeon-Deok	2012
5	35	An Analysis of the Causes and Solutions of Non-co-operations and Conflicts Evolved from the Use of the Commons between Local Governments	Kang Yun-Ho	2005
6	34	The Determinants of the Efficiency of Korean Ports - Using Panel Analysis and Heteroscedastic Tobit Model	Mo Soo-Won	2008
6	34	An Analysis of the Industrial Structure and Regional Economic Effects of Ulsan Regional Port-Logistic Industry	Kim Sang-Choon and Choi Bong-Ho	2008
8	33	An Empirical Study on the Port Logistics Service Quality and Customer Satisfaction - With a Focus on the Shipping Companies of Busan and Gwangyang Ports-	Song Chae-Hun and Song Sun-Yok	2004
8	33	A Comparative Analysis of Terminal Efficiency at Northeast Asia and America Container Ports	Ha Myung-Shin	2009
8	33	An Evaluation of the Competitiveness of Chinese Container Ports	Yeo Ki-Tae	2002

#### 4.2. Keyword Analysis

Table 2 shows the frequency of 4,940 keywords appearing in the abstracts of the 1,151 analyzed papers. In selecting keywords, researchers tend to select the most important terms in a paper. To determine the most actively researched areas in the port literature, we list the top 30 keywords with their frequencies.

The keyword "Port" shows the highest frequency at 60. "Port Competitiveness" (33 times) and "Port Authority (29 times)" have also attracted considerable attention from researchers.



Among the many ports in Korea, authors seem to be most interested in “Busan Port” (25 times), “Incheon Port” (16 times), and “Gwangyang port” (14 times). Regarding research analysis methods, authors tend to use “DEA” (63 times), “AHP” (58 times), “Factor Analysis” (19 times), and “Social Network Analysis” (16 times).

**Table 2.** Keyword Frequency

Rank	Keyword (Frequency)	Rank	Keyword (Frequency)
1	DEA (63)	16	Incheon Port (16)
2	Port (60)	17	China (16)
3	AHP (58)	18	RFID (16)
4	Port Competitiveness (33)	19	Port State Control (16)
5	Port Authority (29)	20	Social Network Analysis (16)
6	Container Port (29)	21	Competitiveness (15)
7	Port Logistics (26)	22	Port Industry (14)
8	Port Hinterland (26)	23	Logistics (14)
9	Efficiency (25)	24	Gwangyang Port (14)
10	Busan Port (25)	25	Port Management (13)
11	Container Terminal (24)	26	Port Competition (13)
12	Customer Satisfaction (20)	27	Green Port (12)
13	Factor Analysis (19)	28	Hinterland (12)
14	Service Quality (17)	29	Busan New Port (11)
15	Hub Port (16)	30	Port Development (10)

**Note:** Abbreviations are used for Data Envelopment Analysis (DEA), Analytic Hierarchy Process (AHP), and Radio Frequency Identification (RFID).

The keyword “Port” shows the highest frequency at 60. “Port Competitiveness” (33 times) and “Port Authority (29 times)” have also attracted considerable attention from researchers. Among the many ports in Korea, authors seem to be most interested in “Busan Port” (25 times), “Incheon Port” (16 times), and “Gwangyang port” (14 times). Regarding research analysis methods, authors tend to use “DEA” (63 times), “AHP” (58 times), “Factor Analysis” (19 times), and “Social Network Analysis” (16 times).

#### 4.2.1. Word Embedding

We vectorize the nouns used in the abstracts of 1,081 articles in 100-dimensional space to preserve the semantic relationship between words in word embedding. We consider abstracts of papers because the abstract summarizes the most important contents of the paper.

Table 3 lists the nouns ranked first to fifth based on cosine similarity, in which eight selected keywords are found to be highly relevant in the embedded space. Through this analysis, we can confirm the concepts and relevancies of major keywords, such as “Port Competitiveness,” “Port Authority,” and “Port Logistics.” “Incheon Port” and “Busan Port” differ sharply in terms of highly related words. “Incheon Port” has a high number of relevant ports such as “Pyeongtaek” and “Dangjin,” while for “Busan Port,” the relevant port is “Gwangyang Port,” and terms such as “Port Competition” and “Hub Port” are highly related. In the case of “China,” in terms of country names or geographical terms, “Japan,” “Asia,” and “Europe” are highly correlated.

**Table 3.** Lists of Five Relevant Nouns for Eight Selected Keywords

<b>Port Competitiveness (Rank 4)</b>	<b>Port Authority (Rank 5)</b>	<b>Port Logistics (Rank 7)</b>	<b>Container Terminal (Rank 11)</b>
Strategic (0.82)	Central government (0.86)	SCM (0.699)	Port operation (0.762)
Integration (0.82)	Port management (0.85)	Distribution (0.693)	SCM (0.748)
Coexistence (0.80)	Legal (0.79)	Port hinterland (0.684)	Hinterland (0.738)
Alternative (0.79)	Local government (0.79)	Security (0.638)	Manpower (0.730)
Sustainability (0.78)	Scheme (0.78)	Cluster (0.637)	Automation (0.714)
<b>Incheon Port (Rank 16)</b>	<b>Busan Port (Rank 10)</b>	<b>Customer Satisfaction (Rank 12)</b>	<b>China (Rank 17)</b>
Pyeongtaek (0.91)	Advantage (0.78)	Service quality (0.86)	Japan (0.91)
Dangjin (0.88)	Gwangyang port (0.77)	Customer loyalty (0.86)	Asia (0.88)
Dalian (0.88)	Port competition (0.75)	Relationship (0.85)	Europe (0.88)
Gwangyang port (0.88)	Hub ports (0.73)	Positive effect (0.85)	FTA (0.85)
Gunsan (0.87)	Gateway (0.73)	Reputation (0.82)	America (0.85)

**Note:** Cosine similarity scores are included in parentheses.

#### 4.2.2. *t*-SNE Visualization

Fig. 4 depicts a graph with the embedded nouns in abstracts, which are determined using word embedding in 100-dimensional space and displayed in a two-dimensional space through *t*-SNE, a non-linear dimensional reduction method. We vectorize the nouns used in the abstracts of 1,511 articles in 100-dimensional space to preserve the semantic relationship between words through Word2vec. All words from abstracts are indicated as gray points. Fig. 4 shows the 60 keywords with the highest frequencies and the 30 major nouns extracted using TextRank. These 30 nouns are commonplace nouns such as “Result,” “Korea,” “Ship,” and “Policy.” However, the 60 keywords selected by authors tend to be used in professional contexts, such as “Port Network,” “Customer Satisfaction,” and “International Logistics.” Nouns extracted using TextRank are shown in square brackets; “Port” and “Logistics” are included as both TextRank-selected nouns and highly frequent keywords. A total of 88 words are represented in the figure, which also depicts the semantic relations between the selected words.

We can classify port research keywords into four groups. For convenience of explanation, we refer to them as groups A, B, C, and D. Group A comprises analytical methods commonly used in port research, including “Analytic Hierarchy Process (AHP),” “Data Envelopment Analysis (DEA),” “SNA,” “Factor Analysis,” and “Simulation.” The words “Method,” “Study,” “Research,” and “Data” selected by TextRank are included in group A as well. These words are used together to describe analytical methods.

Group B is based on the themes of port development and operation from the government’s perspective. We find that “Port Authority,” “Port Governance,” “Port Policy,” “Port Management,” and “Port City” are highly related. In addition, “Port Competitiveness,” “Port Cluster,” and “Co-opetition” appear in studies together.

Group C can be subdivided into two categories: marketing and container terminals. In terms of marketing, “Port Marketing,” “Customer Satisfaction,” “Service Quality,” and “Loyalty” are closely related. In terms of container terminals, “Port Operation,” “Port Competitiveness,” and “Port Facilities” are found to be highly correlated.

Domestic ports such as “Gwangyang Port,” “Incheon Port,” and “Busan Port” are concentrated in group D. The names of domestic ports and themes of “Hub Port,” “Competition,” and “Concentration” seem to appear in papers together.

Fig. 4. Semantic Correlations among Words Visualized via t-SNE

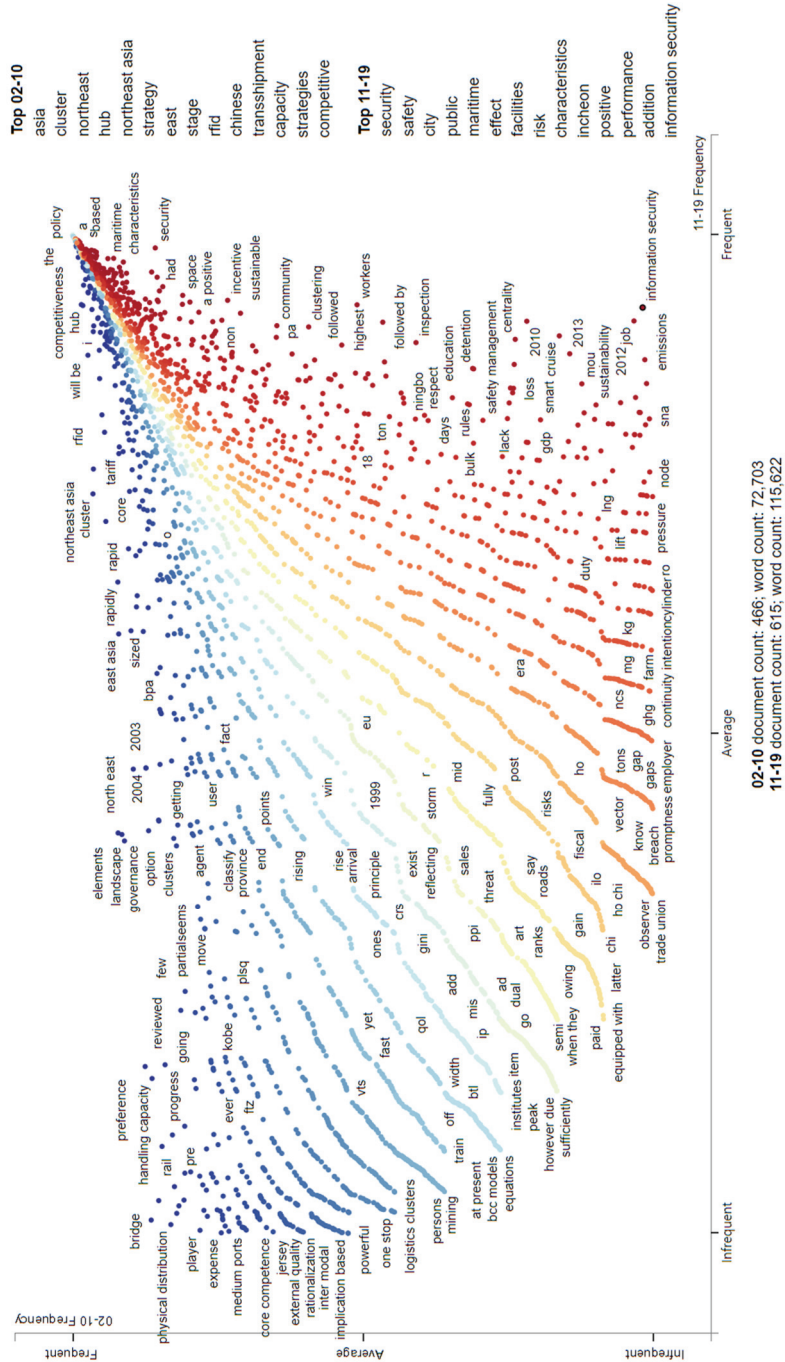


### 4.3. Scattertext

This study collects data for a total of 18 years, 2002–2019. The sample period is divided into two subperiods of nine years each (2002–2010 and 2011–2019) to analyze changes in keywords of port research using Scattertext and a scaled F-score.

In Fig. 5, we use the Scattertext method to visualize the words in the abstract by dividing them into the earlier period (2002–2010) and the later period (2011–2019).<sup>1</sup> The *x*-axis represents the 2011–2019 period and the *y*-axis represents the 2002–2010 period; words with higher values on the axes indicate that they are used often. Therefore, stop words—words that are used frequently in all periods, such as “the” and “a”—are located in the upper right corner. Similarly, words appearing at the top left are more relevant in the 2002–2010 period while those appearing at the bottom right are more relevant in the 2011–2019 period. The Scattertext analysis identifies which words were used in the earlier and later periods to determine changes in research topics and topics of similar importance over the entire 18 years. The graph shows that “Information Security,” “SNA,” and “MOU” are used more often in the later period (2011–2019), whereas “Bridge,” “Rail,” “Player,” and “Medium Port” are used more often in the earlier period (2002–2010).

Fig. 5. Distinguishing Terms for the 2002–2010 and 2011–2019 Periods Using Scattertext



#### 4.3.1. Frequent words in the 2002-2010 Period

Scattertext calculates the scaled *F*-score value based on the screening of words used more meaningfully in 2002–2010 than in 2011–2019. Table 4 shows the words more frequently used in the 2002–2010 period than in 2011–2019. The 10 major words (left) are selected in order of their scaled *F*-score, while the authors' selections (right) are the 10 words that authors consider most important.

**Table 4.** Words More Frequently Used in 2002-2010 ('02-'10) than in 2011-2019 ('11-'19) Selected by Scaled *F*-score (left) and Authors' Choice of Additional Informative Terms (right)

(Unit: Frequency Count)

Scaled F Score	'02-'10	'11-'19	Authors' pick	'02-'10	'11-'19
Asia	137	62	Governance	36	17
Cluster	76	17	Port cluster	30	7
Northeast	95	37	e-business	19	0
Hub	120	61	Hub port	63	13
Northeast Asia	95	37	Port industry	50	33
Strategy	111	82	Handling capacity	20	4
RFID	56	25	Physical distribution	16	0
Chinese	74	48	International logistics	28	8
Transshipment	58	31	Port authority	75	64
Capacity	51	23	Port competitiveness	44	38

Next, we discuss the contexts of the more frequently used words in the 2002–2010 subperiod. “Asia” is used in various ways, such as “Northeast Asia,” “Southeast Asia,” “East Asia,” “Asia,” and “North Asia.” Though many papers use “Asia” in the 2002–2010 subperiod, the number declines in the 2011–2019 subperiod. The term “Cluster” is used in the contexts of “Port Cluster” and “Port Industry cluster.” In addition, the second “Port Cluster” among the authors' picks seem to have semantic connectivity. “Strategy” is used in the context of various aspects of strategic research. In studies during 2002–2010, “Strategy” appears in “Port Marketing Strategy,” “Hub Port Strategy,” “Port Development Strategy,” and “Global Strategy”; in studies during 2011–2019, “Strategy” appears less frequently than in 2002–2010, but was still used in various ways, such as “Port Cluster Strategy,” “Marketing Strategy,” and “Port Operational Strategy.” The word “Chinese” is used in the contexts of “Chinese Port,” “Chinese Company,” “Chinese Port Industry,” “Chinese Economy,” and “Chinese Port Competitiveness” in both the 2002–2010 and 2011–2019 subperiods. “Transshipment” appeared in papers focusing on “Hubs,” “Cargo,” and “Containers.”

#### 4.3.2. Frequent words in the 2011-2019 Period

Table 5 shows the words that appear more frequently in the 2011–2019 subperiod than in the 2002–2010 subperiod. As in Table 4, in this table too, the words on the left are selected based on the scaled *F*-score, and those on the right indicate authors' choice of important words; the frequency of words in the 2002–2010 and 2011–2019 papers are listed. The contexts of the words used more often in 2011–2019 are as follows. “Security” is used in “Container Security” and “Port Security” in 2002–2010, while in 2011–2019, more security-related topics appear, such as “Supply Chain Security” and “Security Policy,” in addition to the previous two themes, “Container Security” and “Port Security.” “Maritime Affairs,”

“Maritime Transport,” “Maritime Authorities,” and “Maritime Accidents” appear in both periods, and “Maritime Security,” “Maritime Terrorism,” “Maritime Trade,” and “Maritime Safety” are mainly manifested in 2011–2019. Thus, “Maritime” appears more often in the 2011–2019 subperiod. For “Information Security,” the 10th term selected based on scaled *F*-score, the main research topics include “Information Security Management System,” “Information Security Policies,” and “Information Security Organization.” “Job” first appears in the 2011–2019 subperiod, and considerable related research was conducted over this period. The relevant contexts are “Job Stress,” “Job Embeddedness,” and “Job Satisfaction.” Furthermore, the recent increase in demand in the cruise industry seems to have increased the number of studies using the term “Cruise.” More research on cruises was conducted in 2011–2019 than in 2002–2010, and the main themes are “Cruise Service,” “Cruise Ship,” and “Cruise Port.”

**Table 5.** Words More Frequently Used in 2011–2019 ('11-'19) than in 2002–2010 ('02-'10) Selected by Scaled *F*-score (left) and Authors' Choice of Additional Informative Terms (right).

Scaled <i>F</i> Score	(Unit: Frequency Count)				
	'02-'10	'11-'19	Authors' pick	'02-'10	'11-'19
Security	19	231	Job	0	50
Safety	40	155	Workers	5	61
City	33	135	Cruise	1	43
Public	20	105	Incentive	10	64
maritime	47	154	port security	1	40
Facilities	72	205	Centrality	2	42
Risk	19	93	Port city	3	43
Positive	30	103	Sustainable	9	55
Performance	64	166	Pyeongtaek	11	50
Information security	0	59	Transportation	41	110

## 5. Discussion

In trend analyses of the domestic port research, Oh Se-Young (2000) and Yoon Il-Hyun (2004) derived results based on research topics, researcher types, and research methods in specific academic journals and detailed subject areas related to ports. In this study, we extend the scope of this analysis by using big data on abstracts. Our study makes several unique and important contributions. First, it is the first to analyze English abstracts in a large number of Korean journal articles over an extended period. Second, it is the first to visualize correlations between words by applying Word2vec on data from abstracts. Third, we comparatively visualize words important in past and more recent research to assess changes in port research trends over time. Nonetheless, the lack of similar research on ports remains a significant concern. We hope future studies will analyze port research trends using other methodologies.

The keywords we identify reflect current issues being faced by the Korean port economy as well as the concerns of scholars who are actively trying to address them. Port research is generally based on careful consideration of governments, businesses, and other entities of high relevance to ports, and has the potential to impact these entities. In other words, port research highlights current issues and, in turn, can impact future decision-making.

In sum, this study is concerned with the trend of Korean port research. In the future, more



studies that conduct comparative analyses on similar topics over different periods and in varying settings are needed. Comparing the findings of Korean analyses with those from overseas research would be informative, particularly for decision makers in governments and industries.

## 6. Conclusion

To understand port research trends, this study uses text mining to analyze the English keywords and English abstracts of 1,511 Korean journals conducted during 2002–2019. First, the main finding is that port research has evolved through the convergence of various academic fields, covering 102 subject areas and 219 journals. The most researched subject area in terms of number of papers is “Marine Transportation” with 280 papers (24.33%), followed by “Agriculture and Oceanography” with 159 papers (13.81%) and “Trade” with 132 papers (11.47%). In terms of journals, the “Korean Journal of Port Economics” has the highest number of papers with 197 (17.12%), followed by the Korean Journal of Korean Port Research with 116 (10.08%) and Shipping & Logistics Research with 199 (6.61%).

Second, the results of the frequency analysis of 4,940 keywords in 1,511 papers shows that “Port” (60 times), “Port Competitiveness” (33 times), and “Port Authority” (29 times) are the main focus of most researchers. Word embedding analysis reveals five words with high correlations with major keywords, and a t-SNE plot presents a visualization and analysis of the port research keyword groups.

Lastly, a Scattertext analysis divides the 18-year analysis period into two sub-periods (2002–2010 and 2011–2019) to compare words used in earlier versus more recent studies. In particular, we examine which words are used more in the two periods and the contexts they are used in.

This study has important implications as it is the first to conduct port research keyword analysis applying text mining on Korean journal abstracts. Our results may guide future researchers in the selection of topics to ensure further development of the field. However, our work may not represent the entire breadth of the existing port research as we only analyze papers that include the word “Port” in the title. For a more representative sample of papers, future studies should include papers published in all port-themed journals, even if the title does not include the word “Port.” Moreover, as this study only analyzes articles published in Korean journals, the scope of this study is limited. To overcome this limitation, it would be interesting to compare and analyze port studies from other countries.

## References

- Bae, Kyu-Yong, Ju-Hyun Park, Jeong-Seon Kim and Yung-Seop Lee (2013), “Analysis of the Abstract of Research Article in Food related to Climate Change using a Text-Mining Algorithm”, *Journal of the Korean Data and Information Science Society*, 24(6), 1429-1437.
- Brin, S. and L. Page (1998), “The Anatomy of a Large-scale Hypertextual Web Search Engine”, *Computer Networks and ISDN Systems*, 30, 107-117.
- Cho, Su-Gon and Seoung-Bum Kim (2012), “Finding Meaningful Pattern of Key Words in IIE Transaction Using Text Mining”, *Journal of the Korean Institute of Industrial Engineers*, 38(1), 67-73.
- Choe, Myoung-Ae, Kyung-Sook Bang, Nam-Cho Kim, Shin-Jeong Kim, Yong-Soon Kim, Hwa-Soon Kim et al. (2012), “Nursing Research Trends Analysis Using 2011 East Asian Forum of Nursing Scholars (EAFONS) Abstract”, *The Journal of Korean Academic Society of Nursing Education*, 18(2), 332-342.

- Helmick, J. S. (2008) "Port and Maritime Security: A Research Perspective", *Journal of Transportation Security* 1, 15-28.
- Hwang, Seo-I and Mun-Ki Kim (2019), "An Analysis of Artificial Intelligence(A.I.) Related Studies' Trends in Korea Focused on Topic Modeling and Semantic Network Analysis", *The of Digital Contents Society*, 20(9), 1847-1855.
- Jeon, Jun-Woo, Y. Wang and Gi-Tae Yeo (2016), "SNA Approach for Analyzing the Research Trend of International Port Competition", *The Asian Journal of Shipping and Logistics*, 32(3), 165-172.
- Jhang, Se-Eun and Su-Ho Lee (2016), "A Study of Themes and Trends in Research of Global Maritime Economics through Keyword Network Analysis", *Journal of Korea Port Economic Association*, 32(1), 79-95.
- Jo, Hye-In, Jin-Woo Kim and Bong-Kyu Lee (2019), "A Study on Research Trends of Blockchain using LDA Topic Modeling: Focusing on United States, China, and South Korea", *Journal of Digital Contents Society*, 20(7), 1453-1460.
- Kessler, J. S. (2017, July 30-August 4) "Scattertext: A Browser-Based Tool for Visualizing how Corpora Differ", *Proceedings of ACL 2017, System Demonstrations*, Vancouver, Canada.
- Kim, Gyu-Ha and Cheol-Yong Park (2015), "Analysis of English Abstracts in Journal of the Korean Data & Information Science Society using Topic Models and Social Network Analysis", *Journal of the Korean Data and Information Science Society*, 26(1), 151-159.
- Kim, Se-Hyun (2018), "A Study of Korea's Multicultural Research Trends using Understand Data Analysis", *Korea Journal of Population Studies*, 41(1), 1-27.
- Lee, Youn-Kyu, Seung-Woong Yoon and Sang-Hoon Lee (2018), "Analyzing Knowledge Structure of Defense Area using Keyword Network Analysis", *Journal of the Korea Society of Computer and Information*, 23(10), 173-180.
- Ma, Hye-Min, Sun-Goo Yoon, Hyung-Ho Kim and Gi-Tae Yeo (2017), "A Study on Research Trend in Field of Port through Keyword Networks", *Korea International Commerce Review*, 32(2), 109-128.
- Maaten, L. and G. Hinton (2008) "Visualizing data using t-SNE", *Journal of Machine Learning Research*, 9, 2579-2605.
- Mihalcea, R. and P. Tarau (2004, July 25-26) "TextRank: Bringing Order into Text", *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, Barcelona, Spain.
- Mikolov, T., I. Sutskever, K. Chen, G. Corrado and J. Dean (2013, December 5-10), "Distributed Representations of Words and Phrases and their Compositionality", *Conference on Neural Information Processing Systems 26 (NIPS)*, Nevada, USA.
- Oh, Se-Young (2000), "An Analysis on Shipping Research Papers in Korea: 1984-2000", *Journal of Shipping and Logistics*, 31, 1-16.
- Pallis A. A., T. K. Vitsounis and P. W. De Langen (2010), "Port Economics, Policy and Management: Review of an Emerging Research Field", *Transport Reviews*, 30(1), 115-161.
- Pallis, A. A., T. K. Vitsounis, P. W. De Langen and T. E. Notteboom (2011), "Port Economics, Policy and Management: Content Classification and Survey", *Transport Reviews*, 31(4), 445-471.
- Park, Joun-Do (2019), "A Study on Issue Tracking on Multi-Cultural Studies using Topic Modeling", *Journal of the Korean Society for Library and Information Science*, 53(3), 273-289.
- Park, Jun-Hyeong and Hyo-Jung Oh (2017), "Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea: Focused on LDA and HDP", *Journal of Korean Library and Information Science Society*, 48(4), 235-258.
- Son, Yong-Jung (2011), "A Meta-Analysis on Research of Korean Port Economics: Focusing on Research Themes and Methods", *Journal of Korea Port Economic Association*, 27(4), 237-251.
- Son, Yong-Jung (2017), "The Centrality of Joint Research in Port Economics using Social Network Analysis", *The Journal of Korean Island*, 29(1), 95-110.
- Witte, P., B. Wiegman and A. K. Y. Ng (2019), "A Critical Review on the Evolution and Development of Inland Port Research", *Journal of Transport Geography*, 74, 53-61.
- Yoon, Il-Hyun (2004), "An Analysis of the Article issued in Korea Journal of Shipping and Logistics: from vol.31 to vol.40", *Journal of Shipping and Logistics*. 42, 1-19.