The International Emergency Management Society
2015 Annual Conference, 30th September - 2nd October 2015, Rome, Italy

A Study on Disaster Response Terminologies Mind Map by SN Analysis

Kang, Seong Kyung [1] and Yu, Hwan [2] and Lee, Young Jai [3]

[1] *Graduate Student, MIS Department, Dongguk University, Seoul, Korea*
[2] *Graduate Student, MIS Department, Dongguk University, Seoul, Korea*
[3] *Professor, MIS Department, Dongguk University, Seoul, Korea*
Email: hshs4123@naver.com, yuhany@hanmail.net, yjlee@dgu.edu

ABSTRACT:
Communication problems arise in the cooperative process of numerous organizations at disaster sites. This is due to organizations having different policies, systems, and terminologies that they utilize. However, for prompt disaster response, unified terminologies are required for smooth communication and cooperation. Therefore, this study analyzes the relationship of specific terms and outputs their order by importance with terminology closeness as the primary criterion. This is done for the purpose of standardizing disaster related terminologies based on Text Mining and Social Network (SN) Analysis, and the study visualizes the outcome by using the Mind Map. The Disaster Response Terminologies Closeness Centrality and the Mind Map expressive process are made up of Term Document Matrix (TDM) composition, normalization, Euclidean Distance, Closeness Centrality calculation, and the Mind Map expressive process. Through these techniques, the closeness of terms can be determined and the resulting data can be utilized in terminology standardization, knowledge deduction through standard terminology, disaster outbreak deduction through disaster symptom information, and disaster response scenario composition.

KEYWORDS:
Disaster Response Terminologies, Text Mining, SN Analysis, Closeness Centrality, Mind Map

1. INTRODUCTION

Every year, various disasters and safety accidents occur. These accidents can be attributed to improper safety management, and result in large damages due to insufficient after-accident response · recovery process.
Uncertainty exists in disaster site activities. It is difficult to predict which secondary accidents will occur when managing a response effort. Therefore, in order to prevent accidental impact diffusion, rescue and recovery should be promptly executed. In order to achieve this, the cooperation of numerous organizations participating in response · recovery process is required, and the disaster executive capacity is significant.
Communication should be smooth in a disaster response cooperation process. However, currently involved organizations have different policies and systems and so their disaster related terminologies are inconsistent. This problem causes hardship in organizations' communication, so terminology standardization is necessary.
Technology that could be applied to terminology standardization is Text Mining (Turban, 2012) and Social Network (SN) Analysis (Kim, Ki-Hwan, 2015). Text Mining deduces frequency of terms based on the relationship between documents and their corresponding terminology. SN analysis produces correlation between nodes (terminology) based on the results of Text Mining. Finally, a Mind Map is developed as an overall presentation tool for the results outputted by the SN Analysis. It shows the level of closeness between all involved terms and their resulting levels of importance will be utilized disaster response terminology standardization.

2. THEORETICAL BACKGROUND

2.1. TDM (Term Document Matrix)

TDM (Term Document Matrix) means 'term/accident case table' where the row indicates the case and the column indicates the term. Correlation of the term and the accident case can be shown in Term Frequency Indicator. However, a frequent term in a document doesn't necessarily indicate the importance of the term. Therefore, the frequency is determined first, and then the following formula is used to perform normalization. (Turban, 2012, Barbro, 1966)

$$\mathrm{idf}(i,j) = \begin{cases} 0 & (\text{if } wf_{ij} = 0 \\ (1+\log(wf_{ij}))\log\frac{N}{df_i} & (\text{if } wf_{ij} \geq 1 \end{cases}$$

Legend

N : Total accident case document
dfi : Document Frequency [Number of documents including the specific basic factor(i)]
i : i th basic factors (Term, Instance)
j : j th accident case (document)
wfi : term frequency [Number the specific basic factor(i) appears in the corresponding case]

2.2. Euclidean Distance

Euclidean Distance is used to calculate the distance between two nodes in an N Dimensional space. When the two nodes P and Q have the following coordinates P=(p1, p2, p3, …, pn) and Q=(q1, q2, q3, …, qn), the formula calculating the distance between the two nodes is the following. (Liwei Wang, 2005)

$$d(o,i) = \sqrt{\sum_{k=1}^{n}(p_k - q_k)^2} = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}, \qquad k = 1,2,\cdots,n$$

d(o,i) = Distance of object and instance

2.3. Social Network Analysis

Currently social network technology based on social relationships is rapidly developing, and SN analysis is also under rapid development. SN analysis is implemented using various analysis tools, and the methods are Degree Centrality, Closeness Centrality, Betweeness Centrality, Structural Hole, Sub-Network Analysis, etc. (Lee, Soo-Sang, 2012)

The development process for Disaster response Term Mind Map uses Closeness Centrality out of the available methods. Closeness Centrality shows the total centrality by measuring how close one node is to another node on average. Closeness Centrality is measurable through closeness and distance to other dots. The distance of two dots refers to the closest distance possible between any pairing.

The study uses distance to define closeness centrality by using the following formula. (Kim, Ki-Hwan, 2015)

$$C_i = \frac{\sum_{o \neq i} \frac{1}{d(o,i)}}{n-1}$$

$d(o,i) = $ Node's shortest distance, $\quad n = $ Total node, $\quad i = $ instance

2.4. Mind Map

Mind Map means 'Map of Thoughts' which visualizes thoughts into a map and is a brain development method to increase thinking skills, creativity, and memory. Memories are lost over time, and so Mind Map has the advantage of recalling thoughts organically. (http://terms.naver.com/entry.nhn?docId=67812&cid=43667&categoryId=43667)

In this study, Mind Map is an expressive technology that assists and eases visualization of disaster response terms' closeness to each other.

3. STUDY DESIGN

3.1. Problem Posing

In order to respond · recover from a disaster, numerous organizations such as Emergency Rescue Agencies (Fire department, Maritime Police), Emergency Rescue Support Agencies (Military, Police), local government, and other related organizations need to all participate. However, various terms, manuals, policies, and systems in each organization designated in disaster situations bring interruption to seamless communication.

For instance, the term 'Disaster Information' is called by various organizations as: 'Disaster Management Information', 'Situational Information', 'Disaster Situation Information', 'Site Information', 'Response Information', and 'Accident Response Information'. The position term 'Liaison Officer' is also called 'Task Liaison Officer' or 'Constant Liaison Officer'. Communication errors occur between organizations since different terms that have identical definitions are used. Therefore, disaster response terms should be standardized to facilitate consistent communication.

3.2. Study Contents and Procedure

The scope of this study contains designating and analyzing terminologies related to disaster 'response' in four phases of disaster management (prevention, preparation, response, recovery). Disaster response process is classified into five domains: Command/Control, Cooperation/Coordination, Information Management, Public Information/Communication, Resource Management referencing USA NIMS (National Incident Management System), and ISO22320 EMR (Emergency Management Requirements).

This study is divided into three parts. TDM determines which terms are repeatedly used (term frequency) in a relationship between terminology and documents (disaster accident cases). The frequency of a specific term itself does not define its significance. In order to resolve this problem, the determined frequency undergoes a normalization process.

The second step is to implement SN analysis. SN analysis determines the closeness of terminologies using Closeness Centrality Analysis. The Euclidean distance equation is used to calculate the closeness of terminologies.

The final step is to illustrate the Mind Map based on Closeness Centrality. The center of the Mind Map holds the term 'Disaster Response' and its distance to the five domains is shown following Closeness Centrality. Finally, the designated terminologies in the five domains are presented in the order of closeness centrality.

4. VERIFICATION ANALYSIS

In this study, five accident cases (Ferry Sewol Sinking Accident (C1), Gumi Hydrofluoric acid Leak Accident (C2), Sampoong Department Collapse Accident (C3), Pangyo Ventilation Collapse Accident (C4), and Daegu Subway Fire Accident (C5)) were chosen. News articles (editorials with columns included), official publications (audit reports and disaster white papers), and disaster response documents are investigated and collected for each case.

4.1. TDM (Term Document Matrix) Composition and Normalization

4.1.1 Term selection and classification

To fill a TDM, terms are selected and classified. The chosen terms are assigned to the column. (20 documents related to the five accident cases are assigned to each row)

The columns are classified into three classes: Class, Object, and Instance. Class involves Object; 'disaster response' of the four disaster management phases is set as the Class. Object is the subordinate of Class, which is set as the five domains (Command/Control, Cooperation/Coordination, Information Management, Public Information/ Communication, and Resource Management) of the 'disaster response'. Instance is the subordinate of Object; it sets terminologies used in the five domains of the disaster response. Table 4.1 shows Class, Object, and Instance set in this study. Instance assigns frequent terms found in the disaster response's five domains based on the involved documents and reports.

Table 4.1 Hierarchical Terms Classification

Category	Composition (term)				
Class	Disaster Response				
Object	Command/ Control	Cooperation/ Coordination	Information Management	Public Information/ Communication	Resource Management
Instance	responsibility	disaster site	cause	notice (Situation)	resource
	authority	field command post	casualties	public information	support (resources)
	report	initial countermeasure	accident accept	misreport	expert
	direction	relief	information system	media	rescue worker
	command	search	manual	nation people	facilities
	control	Rescue	response plan	dissatisfaction	volunteer
	warning	first aid	assessment	telecommunication	
	crosstalk	emergency measure		coverage	
	countermeasure meeting	cooperation, coordination		interview	
	countermeasure headquarter	related organization			
	prediction	medical			
	measure	recovery			
	role	transportation			

4.1.2 TDM Composition and Normalization

Once the documents and terms are selected, they are assigned to a row and column. Then the frequency of the terms mentioned in the collected document should be calculated before the input step.

Term frequency (tf) is how often a term appears in an accident case document. For example, in Gumi Hydrofluoric acid Leak Accident document #4, the term ‘command’ appears ten times and therefore the term frequency is ‘10’. Document frequency (df) is the number of documents that the term appears in out of the twenty accident case documents. For example, the term ‘command’ appears in eight of the documents, therefore the Document frequency is ‘8’. After calculating the Term frequency and Document frequency, normalization is implemented.

For example, the term ‘command’ appears 30 times in the twenty documents which results in $wf_{(i=command)} = 3$. Since $wf_i \geq 1$, the normalization formula $idf(i,j) = (1 + \log(wf_{ij}))\log\frac{N}{df_i}$ is used. N is the total number of the documents (N=20), and the number of the documents which uses the term ‘command’ is 8, so $df_{(i=command)} = 8$. Therefore, the normalization of the term ‘command’ is $idf(command, total\ document) = (1 + \log 30) * \log\frac{20}{8} = 0.9857$. (The study normalized j to the total of 20 documents, not just one specific accident case.)

Table 4.2 is part of the TDM composed in this study. This shows the calculation of terms related to ‘command/control,’ document frequency, and normalization.

Table 4.2 TDM(Command/Control Domain)

Term / Document		Disaster Response												
		Command/Control												
		responsibility	authority	report	direction	command	control	warning	crosstalk	countermeasure meeting	countermeasure headquarter	prediction	measure	role
C1	D1	1		2		1		4			2	1		1
	D2													
	D3				1							1		
	D4	2	1	4	2				3		5			1
C2	D1	1			2		1	1		1		1		1
	D2											1		
	D3											1		
	D4	4	1	8	5	10	14	13	1	1	6	13	3	10
C3	D1				1							5		
	D2					1			1	2	9	5		
	D3			2	7	11	2	2	6		5	2		
	D4					4			7	1	3	2	2	
C4	D1	8			1	1	3					1		
	D2	11		2		1	1		1	1	3	1	1	
	D3	2			2									
	D4			1			2			1	4	5		
C5	D1	2										1		
	D2													
	D3	2												
	D4	2				1	1		2	1		10	2	2
Total Term Frequency		35	2	19	21	30	24	20	21	8	37	50	8	15
Document Frequency		10	2	6	8	8	7	4	7	7	8	15	4	5
Normaliza -tion		0.7658	1.3010	1.1915	0.9241	0.9857	1.0852	1.6084	1.0588	0.8677	1.0220	0.3372	1.3302	1.3101

* C : Accident case, D : Document (Name of the document in reference)

4.2. Euclidian Distance and SN Analysis (Closeness Centrality)

The Euclidian Distance is calculated based on the TDM composition and normalization, and then the SN analysis is performed. The study uses closeness centrality of the SN analysis and the distances between Object and Instance, and Class and Object are determined.

4.2.1 Euclidian Distance

The Euclidian Distance formula $\sqrt{\sum_{k=1}^{n}(p_k-q_k)^2} = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2+\cdots+(p_n-q_n)^2}$ is used to find the distance between terms. This study finds the Euclidian Distance in 2 dimension, so n=2. The value substituted in the Euclidian Distance formula is the term Normalization value (X axis) and the document frequency (Y axis).

The first step is to find all Euclidian distances of the five Objects and Instances in each domain. The higher leveled term in the comparison becomes the ‘standard’. The normalization value and document frequency are both 0. For example, to find the distance between command/control and command, the object

'command/control' has a higher level than the instance 'command'. Therefore, the object 'command/control' becomes the standard 0. The calculation is $d(\text{command/control}, \text{command}) = \sqrt{(0-0.9857)^2 + (0-8)^2} = 8.0605$.
Each object's Euclidian distance is the sum of the relevant domain Instance Euclidian distances. The Euclidian distances of the Objects and Instances are shown in Table 4.3.

4.2.2 SN Analysis (Closeness Centrality)

Instance Closeness Centrality is calculated by using the formula $C_i = \frac{\sum_{o \neq i} \frac{1}{d(o,i)}}{n-1}$. d (o,i) is the closest distance of the object and instance, and therefore the Euclidian distance is substituted. N is the total number of Nodes, which becomes the number of all terms in each domain.

This study calculated the closeness centrality for each instance, so the formula $C_i = \frac{\frac{1}{d(o,i)}}{n-1}$ was used. For example, the closeness centrality of the term 'command' is $C_{command} = \frac{\frac{1}{8.0605}}{13-1} = 0.0103$. N varies depending on which Instance value corresponds to which object. Closeness Centrality of the Object is $C_o = \frac{\sum_{o \neq i} \frac{1}{d(o,i)}}{5-1}$

Table 4.3 Euclidean Distance and Closeness Centrality calculation

Object	Command/ Control	Total U.D : 92.6271 C.C : 0.00269899	Cooperation/ Coordination	Total U.D : 58.9100 C.C : 0.00424376	Information Management	Total U.D : 58.6774 C.C : 0.00426058
Instance	responsibility	U.D : 10.0293	disaster site	U.D : 13.0122	cause	U.D : 12.0134
		C.C : 0.0083		C.C : 0.0064		C.C : 0.0139
	authority	U.D : 2.3859	field command post	U.D : 5.1296	casualties	U.D : 14.0069
		C.C : 0.0349		C.C : 0.0162		C.C : 0.0119
	report	U.D : 6.1172	initial countermeasure	U.D : 9.0339	accident accept	U.D : 7.1006
		C.C : 0.0136		C.C : 0.0092		C.C : 0.0235
	direction	U.D : 8.0532	relief	U.D : 4.2372	information system	U.D : 7.0810
		C.C : 0.0103		C.C : 0.0197		C.C : 0.0235
	command	U.D : 8.0605	search	U.D : 4.2469	manual	U.D : 11.0266
		C.C : 0.0103		C.C : 0.0196		C.C : 0.0151
	control	U.D : 7.0836	rescue	U.D : 12.0194	response plan	U.D : 4.2268
		C.C : 0.0118		C.C : 0.0069		C.C : 0.0394
	warning	U.D : 4.3112	first aid	U.D : 3.3850	assessment	U.D : 3.4546
		C.C : 0.0193		C.C : 0.0246		C.C : 0.0482
	crosstalk	U.D : 7.0796	emergency measure	U.D : 3.3386		
		C.C : 0.0118		C.C : 0.0250		
	countermeasure meeting	U.D : 7.0536	cooperation, coordination	U.D : 11.0207		
		C.C : 0.0118		C.C : 0.0135		
	countermeasure headquarter	U.D : 8.0650	related organization	U.D : 6.1687		
		C.C : 0.0103		C.C : 0.0135		
	prediction	U.D : 15.0038	medical	U.D : 5.17702		
		C.C : 0.0198		C.C : 0.0161		
	measure	U.D : 4.2154	recovery	U.D : 3.5181		
		C.C : 0.0198		C.C : 0.0237		
	role	U.D : 5.1688	transportation	U.D : 4.2268		
		C.C : 0.0161		C.C : 0.0197		

Object	Public Information/ Communication	Total U.D : 57.3388	Resource Management	Total U.D : 49.8239
		C.C : 0.004360053		C.C : 0.00501767
Instance	notice (Situation)	U.D : 4.2469	resource	U.D : 10.0412
		C.C : 0.0294		C.C : 0.0199
	public information	U.D : 5.1543	support (resources)	U.D : 10.0202
		C.C : 0.0243		C.C : 0.0200
	misreport	U.D : 2.6762	expert	U.D : 12.0141
		C.C : 0.0467		C.C : 0.0166
	media	U.D : 5.1543	rescue worker	U.D : 11.0281
		C.C : 0.0243		C.C : 0.0181
	nation people	U.D : 13.0098	facilities	U.D : 13.0113
		C.C : 0.0096		C.C : 0.0154
	dissatisfaction	U.D : 4.1726	volunteer	U.D : 2.5625
		C.C : 0.0300		C.C : 0.0780
	telecommunication	U.D : 3.0648		
		C.C : 0.0408		
	coverage	U.D : 9.0346		
		C.C : 0.0138		
	interview	U.D : 3.3105		
		C.C : 0.0378		

4.3. Mind Map

The Mind Map is illustrated after each term's closeness centrality is calculated. The center (Start point) of the Mind Map is the Class 'Disaster Response'. The Mind map shows term closeness, which allows comparison under two aspects [refer to Figure 1]

First, compare the Class and Object. The comparison of 'Disaster Response' and five objects is possible. The closest domain to the disaster response out of Command/Control, Cooperation/Coordination, Information Management, Public Information/ Communication, and Resource Management is 'Command/Control,' which has the smallest numeric value. Lower numeric values signify closer distances between terms. Therefore, the order of closeness (from closet to furthest) to disaster response would be the following: Command/Control, Cooperation/Coordination, Information Management, Public Information/ Communication, and Resource Management. This study substitutes closeness centrality into α to easily compare the five objects while making the Mind Map. (α = Closeness Centrality x 10^4)

Distance comparison of the Object and Instance is also eligible. The closest term to 'command/control' is the term 'responsibility,' which has the lowest Closeness Centrality value. By arranging the leftover Instances in order, terms that are closer to the 'command/control' domain can be determined.

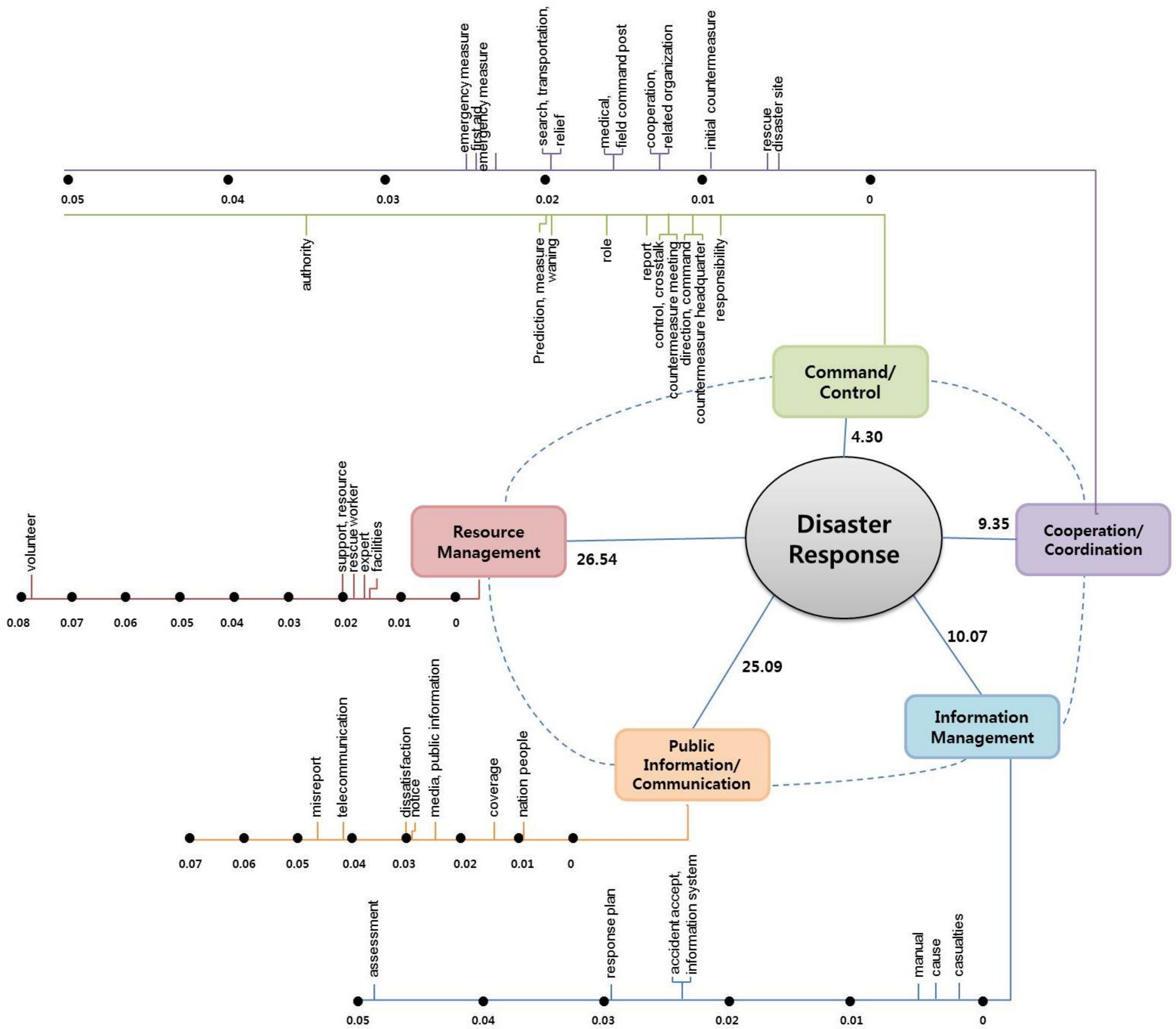

Figure 1 Disaster response term Closeness

5. CONCLUSION

This study has analyzed terms related to the disaster response of disaster management's four phases. Complex disasters and safety accidents are drastically increasing, but communication problems continue to persist between various organizations. Therefore, a terminology analysis was performed for the purpose of characterizing the root cause of communication problems.

By calculating and analyzing terminology frequency in various accident case documents that contain response procedures, this study determined which terms are closely related to disaster response. The levels of correlation between terms are determined by their closeness to each other. This means core terms can be identified in disaster response Command/Control, Cooperation/Coordination, Information Management, Public Information/ Communication, and Resource Management.

The terminology analysis procedure can be utilized through various means. First, the standardization of terminology is eligible based on their importance. The usage of different terms in disaster sites causes communication problems and leads to the prevention of a timely response effort. This delay of response causes secondary accidents and additional casualties. Therefore, standardization of terminology through analysis is necessary.

The terminology closeness analysis results are utilized in researching and accumulating data for disaster management. By applying past accident case analysis, mitigation alternation and outbreak possibility for future Disaster sign data can be deduced. Various accident response scenarios can be prepared based on the disaster sign data. Disaster response scenarios consisting of core terms should be used for the education and training of disaster officials. These officials would provide feedback by detecting additional problems. This will lead to proper preparation for all future accidents.

REFERENCES

Barbro B., Teija L. and Kaisa S. (1996). Neural Networks and Genetic Algorithms for Bankruptcy Predictions *Proceedings of The third World Congress on Expert Systems,* **Volume:11 Issue:4,** 123-130

Turban, E., Sharda, R., and Delen, D. (2010). Decision Support and Business Intelligence Systems (9th Edition), Pearson Education

Liwei Wang, Yan Zhangy, Jufu Fengy. (2005). On the Euclidean Distance of Images, Peking University

Lee, Soo-Sang. (2012). Network analysis methods (1st Edition), Nonhyoung

Kim, Ki-Hwan. (2015). Analyzing domestic research author's and reviewer network using social network analysis, Dongguk University

http://terms.naver.com/entry.nhn?docId=67812&cid=43667&categoryId=43667

Accident Case Document

Ferry Sewol Sinking Accident	D1	Jindo ferry Sank··· Ministry of Oceans and Fisheries, Marine accident appointed 'serious' alarm (sbs, 2014.4.16)
	D2	Ferry sinking accidents··· Why casualties got to be so big?(Yeonhap News, 2014.4.16)
	D3	Initial stage chaos·wore life jackets too late······ accident response underdeveloped (Gookminilbo, 2014.4.17)
	D4	Government response issues and improvement plans of Sewol ferry accident (Korea Institute of Public Administration, 2014)
Gumi Hydrofluoric acid Leak Accident	D1	Poor initial response brought bigger damage (Chosun Ilbo, 2012.10.05.)
	D2	Gumi Hydrofluoric Acid Leak, shame to say 'chemistry advanced nation' (Joseh Ilbo, 2012.10.17.)
	D3	Gumi Hydrofluoric Acid Leak should be handled as the second Phenol situation (Kyunghyang News, 2012.10.04
	D4	Response plan for harmful chemical substance leak (2013, Gimje Fire Department Research)
Sampoong Department Collapse Accident	D1	Sampoong Department Collapse Accident (Chosun ilbo, 1995.6.30)
	D2	Sampoong Department Collapse (Choongang Ilbo, 1995.7.1)
	D3	Department Collapse Accident, fragile conduct structure (Hankyoreh, 1995.7.4)
	D4	Sampoong Department Collapse (Defense White paper, 2012.6.25)
Pangyo Ventilation Collapse Accident	D1	Facility safety policy reinforcement for building code (Daily Korea, 2014.10.19)
	D2	Pangyo Ventilation Collapse Accident Core analysis(Weekly Hyundai, 2014.10.27)
	D3	Pangyo Ventilation Collapse Accident Reconstruction (Seoul Nat. Univ. Construction environment Institute, 2014. 11. 11)
	D4	Security measures promotion plan (Kyungki Development Institute, 2014)
Daegu Subway Fire Accident	D1	What a tragedy (Kookmin Ilbo, 2003.02.18)
	D2	Domestic subway Conflagration Defenseless (Yeonhap News , 2003.02.18)
	D3	Self-portrait through subway tragedy (Moonhwa Ilbo, 2003.02.19)
	D4	Problems and Improvements of disaster management focused on the Daegu Subway Tragedy (Kyungnam Development Institute, 2013)