Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2016 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 6-27-2016

A TEXT MINING APPROACH TO THE ANALYSIS OF INFORMATION SECURITY AWARENESS: KOREA, UNITED STATES, AND CHINA

Tae-Heon Lee
SK Infosec, arrsian@naver.com

Won-Kyung Sung
Yonsei University, sung_wonkyung@yonsei.ac.kr

Hee-Woong Kim
Yonsei University, kimhw@yonsei.ac.kr

A TEXT MINING APPROACH TO THE ANALYSIS OF INFORMATION SECURITY AWARENESS: KOREA, UNITED STATES, AND CHINA

Tae-Heon Lee, SK Infosec, Gyeonggi-do, Seoul, Korea, arrsian@naver.com

Won-Kyung Sung, Graduate school of Information, Yonsei University, Seoul, Korea, Sung_wonkyung@yonsei.ac.kr

Hee-Woong Kim, Graduate school of Information, Yonsei University, Seoul, Korea, kimhw@yonsei.ac.kr

Abstract

Recently in Korea, the importance of information security awareness has been receiving a growing attention. Attacks such as social engineering and ransomware are hard to prevent because it cannot be solved by information security technology. Also, the profitability of information security industry has been decreasing for years. Because of this, many companies try to find a new growth-engine and an entry to the foreign market. The main purpose of this paper is to draw out some information security issues that people of each country think and to analyze it. Finally, this study identifies issues and suggests how to improve the situation in Korea. For this, Topic Modeling analysis has been used to find information security issues of each country. Moreover, the score of sentiment analysis has been used to compare each country. The study contributes to the literature by exploring and explaining what critical issues are and how to improve the situation based on the identified issues of the Korean information security industry. Also, this study adds to the literature by demonstrating how text mining can be applied to the context of information security awareness. From a pragmatic perspective, the study has the implications for information security enterprises. This study is expected to provide a new and realistic method of analyzing domestic and foreign issues using the analyzing real data of the Twitter API.

Keywords: Information Security Awareness, Text Mining, Topic Modeling, Sentiment Analysis, Social Media, Twitter

1 INTRODUCTION

Recently in Korea, because of increased cyber terrorism and online security risk, Korea has been reinforced regulations pertaining to cyber security. For example, 2014 January, one staff led to the inadvertent disclosure of personal information in some major card corporations; likewise, in 2014 October, Korea Hydro & Nuclear Power Co. Ltd., which is the public nuclear enterprise of Korea, has been hacked into by someone, and these problems are still in progress. After undergoing several incidents like, Korea government considered two countermeasures. First, government newly created a secretary position for cyber security affairs within The national security to bolster control tower coping with cyber security. Second, in 2015 December, they enforced a new Information Protection Industry Promotion Act. However, despite these tides, information protection industry still experiences the recession. On account of continuing deterioration of the industry, the Korea protection information industry has been striving to cultivate a new market and to develop new fields.

These days, furthermore, the importance of information security awareness has been being emphasized as well as information protection technology (Lim, 2006). As mentioned above, many cases penetrate through the vulnerability of human beings, which are kinds of Social Engineering and Zero Day Attack. Due to the fact that it is impossible to prevent, it is important to enhance information security awareness. In previous research, however, the research scopes are limited to certain factors influencing awareness and action of information security (Kim, 2010; Baek, 2010; Humaidi and Balakrishnan, 2015). In addition, analysis general users felt concerning information protection awareness. However, researches to analyze haven't been studied yet. In conclusion, this study suggests analyzing the information security awareness in tweets uploaded by three countries' general users -Korea, China, and the USA- to enhance information security awareness of Korea.

2 CONCEPTUAL BACKGROUND

2.1 Advanced research related to information security awareness

Information security awareness composed of two fields, general knowledge and organizational information security awareness (Cavusoglu et al. 2009). Advanced research on information security awareness generally performed by using a survey analysis (Kim, 2010; Baek, 2010; Bulgurcu et al., 2010; Sohn, 2013; Lee, 2012; Humaidi and Balakrishnan, 2015). Also, most advanced researchers focus on factors influencing behaviors and awareness of information security (Kim, 2010; Baek, 2010; Bulgurcu et al., 2010; Sohn, 2013; Lee, 2012; Humaidi and Balakrishnan, 2015). They were just limited to some specific cases; in addition, existing research doesn't include what is user's awareness of information security, how to analyze and draw issues, and how to analyze other countries' awareness of information security. Thus, this study aims to improve the awareness of information security in Korea by analyzing Korea, China, and USA' twit data respecting information security.

2.2 Topic modeling

This study performed Topic modeling by LDA (Latent Dirichlet Allocation). it is the analyzation trying to find a subject by analyzing words in the text. (Blei and Jordan, 2003, Griffiths and Steyvers, 2004). Especially LDA has been generally used to look up a hidden subject. The algorithm of LDA is as follows.

$$\prod_{k=1}^{K} p(\beta_k \mid \eta) \prod_{d=1}^{D} p(\theta_d \mid \alpha) \left(\prod_{n=1}^{N} p(z_{d,n} \mid \theta_d) p(w_{d,n} \mid \beta_{1:K}, z_{d,n}) \right)$$

The algorithm predicts the topic proportions of the document, per-word topic assignment, and topic with the individual words and the document. In the algorithm, the hyper parameters of α and η have an equal value in relation to the entire document. The value α is a parameter which decides the value of topic proportions (θ) of each document and decides the Dirichlet distribution type of θ. The value η is a parameter which decides the Dirichlet distribution type of each word's ratio to entire topic (β). Then, depending on the value of θ, which follows the Dirichlet distribution, the per-word topic assignment (z)

within the document is determined. With the value z, that represents the per-word topic, and the value β, which is the ratio to the whole topic of each word, the word w is defined. Studies using Topic Modeling have been generally conducted in two fields. One is the analysis of research trends and the other is social media analysis. The above one is mainly to analyze Trend of each field (Mann et al., 2006; Park, 2013; Kim, 2015) or to raise an analysis efficiency (Hillard et al., 2008; Yu, 2015). The later one is mainly to analyze whole twit data (Hong and Davison., 2010; Jin, 2013) and to propose a new analysis model or to improve existing models (Wang et al., 2013; Bae, 2013; Park, 2015).

2.3 Sentiment analysis

The sentiment analysis is a technology to analyze people' Subjective attitudes and opinions in the text. It has been studied to improve an accuracy of analysis and to build a Sentiment Lexicon (Baccianella et al., 2010; Taboada et al., 2011; An, 2015). In the social media analysis, Sentiment analysis has restricted to conduct a prediction of an election result, correlations of news with stock price, and searching fashion trend. Thus, we can confirm that the existing researchers have not yet analyzed the information security issue through social media analysis though the analysis has the advantage of scoring each fields' values. Therefore, this study is going to utilize sentiment analysis to score values and to compare them.

3 RESEARCH METHODOLOGY

This research is to improve people's awareness of information security by finding the people's concerns on the information security issues and comparing information security issues that are arising in different countries. To conduct that, the Topic Modeling method is applied to identify information security issues in Korea, the U.S, and China, and the sentiment analysis technique is used to compare the scores measured on each topic. The outline of the procedure is as follows Figure 1.

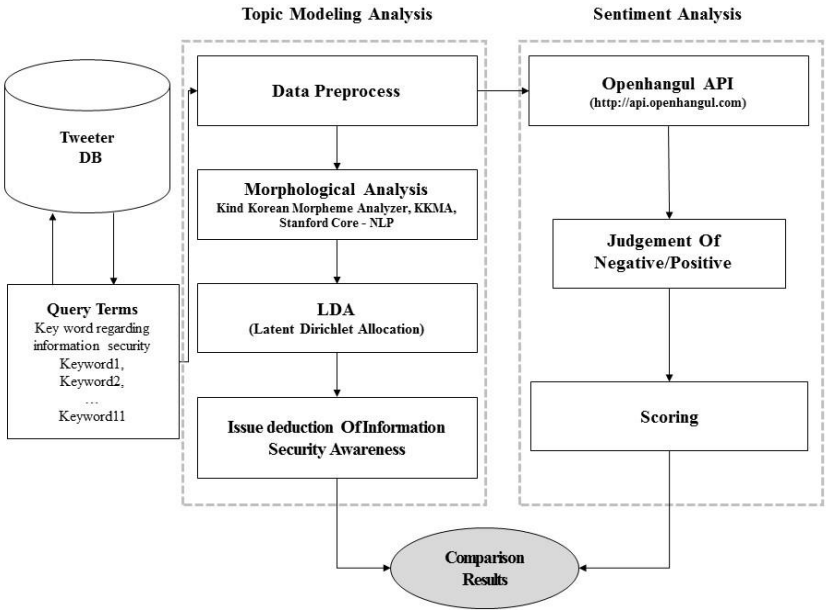


Figure 1. Research procedure

3.1 Collection of research key words

For accurate observation, keyword selection is applied before the data collection thorough interviews with 20 different IT experts such as information security operators, security consultants, and information protection experts. After, these 11 keywords are selected, these keywords are translated in English and Chinese, and checked in the google and Baidu. The final keywords are as below Table 1.

Korea	USA	China
보안	Security	安全
정보보호	Data Protection	信息保护
프라이버시	Privacy	隐私
개인정보보호	Privacy Protection	个人信息保护
개인정보	Personal Information	个人信息
정보 프라이버시	Information Privacy	信息隐私
기밀성	Confidentiality	机密性
무결성	Integrity	完整性
가용성	Availability	可用性
부인방지	Non-Repudiation	不可抵赖性
인증	Authentication	认证

Table 1. List of key words for the data collection

3.2 Data collection and data pre-process

From 2015. 10. 6 to 2015. 11. 6, 299,871 tweets in Korean, 781,822 tweets in the U.S.A and 801,011 tweets in China is collected for this research. Any of data from Twitter including the 11 selected keywords are collected by using the API, the programming language Java and R. A social media service Twitter could carry user's opinion under the limited 140 words, and it features called retweet in which user could share the other users' tweet, and this feature spread information quickly (Ryu, 2013). As a result, Twitter becomes proper social media in the market research to check users' interest and analyze the market trend (Hong and Davison., 2010; Jin, 2013; Wang et al., 2013; Bae, 2013; Park, 2015). According to research center Global Web Index, Chinese twitter users are about 35. 5 million in 2012, and it is bigger than the USA (22.9) and Korea (1.2). Therefore, this research suggests using twitter data to improve people's awareness.

3.3 Topic modeling, sentiment analysis and scoring

To search for topics throughout the loaded twitter data, this research has conducted LDA Topic Modeling; consequently, it leads to selecting 10 topics in each three countries. This research has conducted the sentiment analysis using the Openhangul API (An and Kim, 2015) and NLTK. Openhangul and NLTK are sentiment analysis for Korean and English. Because Sentiment analysis aims to determine the attitude of a writer regarding the overall contextual polarity of a document, it enables us to judge the positive/negative response of each tweet unit and calculated the ratio between the positive/negative tweet on the tweet's topic and the scoring area. Such result is converted to a score on the basis of 5 points as follows:
Score = positive rate / negative rate (%) *5

Also, we have set unified areas in order to compare the scoring result of each country. According to the international standard of the information security management system, ISO27002:2013 were set as the comparison standard, which is composed of 3 domains and 14 items. Therefore, the controlled Items of ISO27002 are used to accurately compare the domain of each country.

4 ANALYSIS AND RESULT

4.1 Topic modeling - Korea

The total 10 derived topics were composed of 7 topics of the technical protection area and 1 topics of physical protection area to which the topic was only related among the 3 countries. Topics about no. 1, 2, 5, and no.7 all mentioned about complication and complexity to use. Topic no.3 and no.4 all represented anxious in general users. Because it can be social costs, it needs to be complemented. Topics

each of no.9 and no.10 are administrative protection area and physical security area; thus, the research supposed that general users have cautions in both of them. Topic no.8 mentioned administrative protection area also. The detailed results are shown on below Table 2.

No	Topic	Keyword
Technical protection area		
1	Mobile payment security	Security, Credit card, Installation, Upgrade, Micropayments, Law, error, Civil Complaint
2	App store authentication	Security, Myself, Success, ID, Storage, App store, Failure, Protection
3	Phishing scam	Security, Incident, Fraud, Report, Site, Revision, Login, Solution
4	Pharming scam	Security, Real-time, Link, Discovery, Incident, Company, Memo, Anxiety
5	Mobile certification system for payment	Security, Apple, Samsung pay, Service, Incident, Authentication, Approval, App certification
6	Mobile app security	Security, Market, Active x, Reflection, log, Link, User, Encryption
7	Inconvenience of identity authentication	Personal information, iPhone, Name illegally, I-Pin, Nonsense, Inconvenience, Log in, Character certification
Administrative protection area		
8	Goal of information security	Personal information, Level, Direction, Trust, Objective, Respect, Cost, Inconvenience
9	Hospital information leakage	Personal information, Objective, Hospital, Research center, Report, Distribution, Myself, List
Physical protection area		
10	Inter-office security	Security, Inter-office, Security risk, Security search, Issue, System, Backup, Change

Table 2. Topic modeling results in Korea

4.2 Topic modeling – United States of America

The total 10 derived topics were composed of 2 topics of the technical protection area and 8 topics of the administrative protection area, which was the largest among the 3 countries. The topics no. 1 and 2 and no. 3 deals with the scope of cyber monitoring in US and personal data transfer. Both topics are based on the issue of invalidating the Safe Harbor agreement which was originally made by the Court of Justice of the European Union and America and currently banned the cross-border data transfer of European countries. The topic of no. 9 was derived from the issue that the U.S. House of Representatives had enforced the legislation of the law that charges the penalty to U.S. car producers in case of not submitting the information protection policy. In the case of the topic no. 8, the main point is to compel the individual agreement of online accounts, which is an issue about the privacy agreement related to the social media, that was proposed in 23 states and the change in 2016 is under the discussion. The detailed results are shown on below Table 3.

No	Topic	Keyword
Technical protection area		
1	Online certification system	Security, Government, Non-repudiation, Oath, State, Order, Secrecy, System
2	Mobile fingerprint authentication system	Security, Authentication, Android, System, Fingerprint, Key, Apple, Password
Administrative protection area		
3	Amendment of data protection law	Security, Privacy, Law, Digital, Information, Id, Secrecy, Act
4	Use of medical information research	Security, Confidentiality, Protection, Datum, Health, Law, Patient, Doctor
5	Amendment of data protection standard	Security, Integrity, Question, Commission, Protection, Journalistic, Nation, Officer
6	Cross border transfer of personal data	Security, Protection, Data, EU, Court, Privacy, Law, Rule
7	Amendment of digital personal data protection act	Security, Privacy, people, Information, Privacy, Time, Person, Policy

8	Agreement of personal data use in social media	Security, Media, Confidentiality, Integrity, Nation, Secrecy, Blog, Plan
9	Compulsory establishment of data protection policy	Security, Penalty, Privacy, Competition, Element, Director, Web, Client
10	Cyber monitoring	Security, Integrity, Right, Value, Money, Trust, Character, Person

Table 3. Topic modeling results in the US

4.3 Topic modeling – China

Topics were composed of 5 topics of the technical protection area and 5 topics of the administrative protection area. The topics are usually related to web, DB, and site censorship. Topic no. 1, 3 and no. 4 commonly composed of security, the web, data, and the internet. Those topics and no. 2 is mainly from the twits respecting enacting cyber security law. Many topics in China derived from the enactment of cybersecurity. The detailed results are shown on below Table 4.

No	Topic	Key-word
Technical protection area		
1	Web server protection	Security, Web, Safety, Support, Option, Person, Data, Internet
2	DB data protection	Security, DB, Work, Data, Safety, Firewall, Time, Domain
3	Web application protection	Security, Web, Site, Shop, App, Id, Work, Information
4	Web traffic	Security, Traffic, Stop, Internet, Information, Injury, Web, Data
5	Cyberattack defense	Security, Attack, Location, Speed, Password, Type, Time, Defense
Administrative protection area		
6	Concern about information leakage in mobile	Security, Data, Mobile, Disclosure, Person, Information, Thing, Anxiety
7	Acts about information protection incidents	Security, Original, Law, Company, Accident, Internet, Product, Act
8	Penalty of information and communications industry	Security, Fine, Government, Data, Method, Communication, Penalty, Law
9	Site censor	Security, Law, Support, Scan, System, Check, Site, Datum
10	Information protection incident	Security, Safety, Fake, Information, Law, Prevention, Problem, Crime

Table 4. Topic modeling results in China

4.4 Topic modeling – Comparisons between countries

In comparison with topics among countries, Korea has the most many topics concerning technical protection area, and the USA derived the topics regarding administrative protection area. China derived respectively 5 topics, respecting administrative protection area and technical protection area. Detailed is as follows. The topic ‘1. Online certification system’ in the USA scored 3.7, whereas Korea's related Topics ‘2. App Store identity authentication’ and ‘7. Inconvenience of identity authentication’ scored 2.0 and 1.7. Thus, Korea needs to complement the technologies and to reform inconvenient services. Like as ‘6. Concern about information leakage in mobile’ scored 2.4 in China, ‘3. Phishing scam’ and ‘4. Pharming scam’ marked the worst score 1.5 and 1.4 points. It indicated that China and especially Korea should strive to resolve them in various fields of activity. Therefore, it suggests improving convenience of general users. The detailed results are shown on below Table 5.

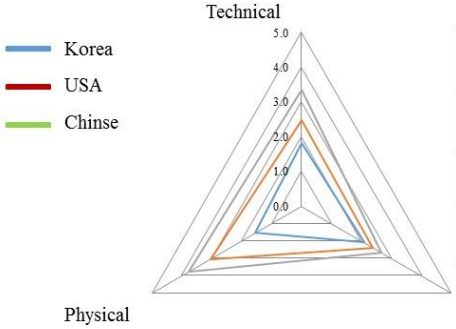
Topic	Score	Country	Topic	Score	Country
Technical Protection area					
Mobile payment security	1.7	Korea	Online certification system	3.7	USA
App store authentication	2	Korea	Mobile fingerprint authentication system	3.4	USA

Phishing scam	1.5	Korea	Web Server Protection	3.0	China
Pharming scam	1.4	Korea	DB data protection	3.2	China
Mobile certification system for payment	1.7	Korea	Web application protection	2.6	China
Mobile app security	1.5	Korea	Web traffic	2.3	China
Inconvenience of identity authentication	1.7	Korea	Cyberattack defense	2.2	China
Administrative protection area					
Goal of information security	3.1	Korea	Compulsory establishment of data protection policy	3.1	USA
Hospital information leakage	1.4	Korea	Cyber monitoring	3.7	USA
Amendment data protection law	2.6	USA	Concern about information leakage in mobile	2.4	China
Use of medical information research	3.4	USA	Acts about information protection incidents	2.4	China
Amendment of data protection standard	2.6	USA	Penalty of information and communications industry	2.4	China
Cross border transfer of personal data	2.5	USA	Site censor	2.6	China
Amendment of digital personal data protection act	2.5	USA	Information protection incident	2.4	China
Physical protection area					
Inter-office security	1.5	Korea			

Table 5. Comparison of topic modeling results

4.5 Sentiment analysis

This research has conducted the sentiment analysis in accordance with ISO270: 2014's technical area, physical area, and administrative area. The score ranges from 0 to 5. A score closer 0 means twits are mainly negative and a score closer 5 means twits are mainly positive. The result is that, in the all three areas, China received the highest score. USA and Korea followed it. Below Figure 2 is the scores of each country. For details, this research measured scores by ISO27002:2013 in terms of technical, administrative, physical fields. In the topics, '9. Access control' and '10. Encryption', Korea and USA is lower than China. It indicated that there are some issues that message encryption Google and Apple applied made USA government hard to deal with terrorism and a different viewpoint that USA prohibited Chinese digital protectionism through TPP. The detailed result is as follows Table 6.



	Sentiment analysis score		
	Korea	USA	China
Technical protection area	1.8	2.5	3.4
Administrative protection area	2.1	2.4	2.7
Physical security area	1.6	3.1	3.8

Figure 2. Comparison of sentiment analysis between countries

Controlled items (ISO27002:2013)	Korea	USA	China
Technical protection area			
Access Control	1.6	1.1	3.3
Cryptography	15	0.9	4.5
Operation Security	2.0	3.8	3.2
Communication security	1.7	3.7	3.0

System acquisition	2.3	2.9	3.0
Administrative protection area			
Information Security	2.5	2.1	3.0
Organization of Information Security	2.3	3.6	2.9
Human Resource Security	2.5	3.5	2.9
Asset Management	2.4	2.0	2.2
Supplier relationships	-	-	-
Information security incident management	1.5	1.3	2.5
Compliance	-	1.6	-
Physical and environmental security	-	2.6	2.6
Physical security area			
Physical and complementary environment	1.6	3.1	3.8

Table 6. Comparison of sentiment analysis results

5 DICUSSION AND CONCLUSION

5.1 Discussion and Implications

This research analyzes unstructured data uploaded by general users through text mining. There are some academic implications. First, this research has employed text mining techniques instead of the survey method. Secondly, while the existing research on information protection awareness are limited on the correlation between the information protection awareness and the affecting factors, this research is conducted on the general user's opinions itself. Third, this research has conducted a comparative study between the nations based on the derived information protection issue. Therefore, it can be applied to future comparative studies. This research also has a practical significance as it has utilized real data collected in Twitter. We expect that it can be a research model for domestic and foreign market analysis. Finally, the methods of this research will enable the comprehensive analysis, from every field's aspect.

5.2 Limitations and Future direction of the research

This research has some limitations. First one is data limitation. We faced with the difficulties during the data collection process in China. Hence, we used Twitter database instead of Weibo being most popular in China; moreover, as we can't analyze data written in Chinese, we translated it into English using the Google translator. This might cause the discrepancy although such discrepancy can be minimized as the detailed unit of analysis of text mining technique is vocabulary. Secondly, it is difficult to directly compare the score of each nation. Score revision might be required depending on each nation's information protection condition. Third, this research analyzes the data collected only during 32 days.

5.3 Conclusion

To analyze information security awareness of each country, this study employed data mining process with extracted twit data on the basis of 11 keywords formulated by expert interview. As a result, each of the countries calculated 10 topics, each topic's 8 keywords, and sentiment score in three security fields, which are derived from ISO 27002. In the result, the research confirmed that Korea's main topics are derived by technical security, which is about phishing, farming, mobile app security, Loop Pay intranet hacking, and information leakage in the hospital, and they scored about 1.5 points. It means that Social Engineering Hacking- phishing, farming- and information leakage of mobile was most communicated in public and was more concerned than other countries; thus, the finding indicated Korean information security industry needs to consider a countermeasure to dispel worries in public. This research, in conclusion, demonstrated that public' issue in information security awareness and each information security field's affective score derived from data mining with unstructured data is also valid as well as existing work method- like scrap articles, and they are utilized as exploratory analysis in practical fields by comparison with other countries' result as well as drilling down into the detail.

References

An, J.K. and Kim, H.W. (2015). Building a Korean Sentiment Lexicon Using Collective Intelligence. Journal of Intelligence and Information Systems, 21(2), 49-67.

Baccianella, S., Esuli, A. and Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. LREC(10), 2200-2204.

Bae, J.H., Son, J.E. and Song, M. (2013) Analysis of Twitter for 2012 South Korea Presidential Election by Text Mining Techniques. Korea Intelligent Information Systems Society, 19(3), 141-156.

Baek, M.J. and Son, S.H. (2010). A Study on the Effect of Information Ethics on the Information Security Awareness and Behavior in Organization. KSI KISS, 28(4), 119-145

Blei, D., Ng, A. and Jordan, M. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3, 993-1022.

Bulgurcu, B., Cavusoglu, H. and Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS quarterly, 34 (3), 523-548.

Cavusoglu, H., and Son, JY. and Benbasat, I. (2009). Information security control resources in organizations: A multidimensional view and their key drivers. working paper, Sauder School of Business, University of British Columbia.eMarketer.

Choi, P.A. and Lee, M.H. (2014). The Effects of Industrial Security Activities of Local Institution on the Technology Outflow Prevention. National Association of Korean Local Government, 16(1), 119-141.

Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. Proceedings of the National Academy of Sciences, 101(suppl 1), 5228-5235.

Hillard, D., Purpura, S. and Wilkerson, J. (2008). Computer-assisted topic classification for mixed-methods social science research. Journal of Information Technology & Politics, 4(4), 31-46.

Hong, L. and Davison, B. D. (2010). Empirical study of Topic Modeling in twitter. In Proceedings of the first workshop on social media analytics. ACM, pp. 80-88

Humaidi, N. and Balakrishnan, V. (2015). Leadership styles and information security compliance behavior: The mediator effect of information security awareness. International Journal of Information and Education Technology, 5(4), 311-318.

Jin, S.A., Co C.E., Jeong, Y.K. and Song, M. (2013). Topic-Network based Topic Shift Detection on Twitter. Korea Society for Information Management, 30(1), 285-302.

Kim, S.H., Song, S.J. and Song, M. (2015). Investigation of Topic Trends in Computer and Information Science by Text Mining Techniques. KCI, 32(1), 135-152.

Kim, S.Y., Song, S.J. and Song, M. (2015). Investigation of Topic Trends in Computer and Information Science by Text Mining Techniques. Korea Society for Information Management, 32(1), 135-152.

Kim, Y.G. (2010). A Study on the Influence of the Components Related to Informaton Security Governance on the Perception and Behavior of Employees. KONI, 14(6), 935-950.

Lee, U.K., Choi, J.E. and Lee, H.G. (2012). The Role of Control Transparency and Outcome Feedback on Security Protection in Online Banking. Information Systems Review, 14(3), 75-97.

Lim, C.H. (2006). Effective way to improve information security. KIISC, 16(2), 30-36.

Mann, G.S., David, M. and Andrew, MC. (2006). Bibliometric impact measures leveraging topic analysis. Digital Libraries. JCDL'06. Proceedings of the 6th ACM/IEEE-CS Joint Conference on. IEEE. pp. 65-74.

Park, J.D. (2015). A Study on Mapping Users' Topic Interest for Question Routing for Community-based Q&A Service. Korea Society for Information Management. 32(3), 397-412.

Park, J.H. and Song, M. (2013). A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling. Journal of the Korean Society for information Management, 30(1), 7-32.

Ryu, W.J., Ha, J.W., Md. Alam, H. and Lee, S.K. (2013). Extracting Trends from Twitter using a Topic Modeling Technique," Korea Information Science Society, pp. 191-193.

Sohn, S.H. (2013). A Study on the Effect of Information Security Awareness and Behavior of Job Autonomy and Burden and Mobility under Smart Work Environment. KSI KISS, 31(4), 17-39.
Taboada,M., Brooke, J., Tofiloski, M., Voll, K. and Stede, M. (2011). Lexicon-based methods for sentiment analysis. Computational linguistics, 37(2), 267-307.
Yu, S.Y. (2015). Combining Ego-centric Network Analysis and Dynamic Citation Network Analysis to Topic Modeling for Characterizing Research Trends. Korea Society for Information Management, 31(4), 153-169.