

## 텍스트 마이닝을 이용한 정보보호인식 분석 및 강화 방안 모색\*

이 태 현\*\*, 윤영주\*\*\*, 김희웅\*\*\*\*

### 요약

최근 정보보호 분야에서는 사회공학, 랜섬웨어와 같은 정보보호 기술만으로는 막을 수 없는 공격이 증가하고 있으며, 이에 따라 정보보호인식의 중요성이 부각되고 있다. 또한 정보보호 업계의 수익악화가 두드러짐에 따라 정보보호 업계의 신성장동력을 탐색하고 해외시장을 개척하고자 하는 노력이 증대 되고 있다. 이에 따라 본 연구는 사람들이 생각하는 정보보호 관련 이슈들을 도출하고, 온라인에서의 정보보호 관련 이슈의 국가간 비교 분석을 통하여 한국의 정보보호인식의 개선방안을 제안하고자 한다. 이를 위해 본 연구에서는 토픽 모델링 기법을 적용하여 한국과 미국, 중국의 정보보호 관련 이슈를 확인 하고, 감성 분석을 통하여 점수를 측정해 비교 분석하였다. 본 연구의 학술적 시사점은 비정형 데이터인 트위터의 트윗을 텍스트 마이닝 기법인 토픽 모델링과 감성 분석 기법을 통해 분석하고, 도출된 이슈를 기반으로 국가간 비교 연구를 수행 하였으며 이를 바탕으로 한국의 정보보호인식 강화 방안을 탐색하였다는 점에서 의의가 있다. 또한 본 연구의 실무적 시사점은 트위터 API 를 통한 실제 데이터를 이용한 연구로 본 연구 모델을 활용하여 국내 이슈 및 해외 시장 분석에 활용 가능할 것 이라는 점에 있다.

주제어: 정보보호인식, 텍스트 마이닝, 토픽 모델링, 감성 분석, 소셜미디어, 트위터

## The Analysis of Information Security Awareness Using A Text Mining Approach

Lee, Tae-Heon, Youn, Young-Ju, Kim, Hee-Woong

### 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 be prevented because it cannot be solved by information security technology. Also, the profitability of information security industry has been decreasing for years. Therefore, 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 and to analyze them. 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 them. The study is 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 is also 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 for analyzing domestic and foreign issues using the analysis of real data of the Twitter API.

Keywords: information security awareness, text mining, topic modeling, sentiment analysis, social media, twitter

2016년 8월 26일 접수, 2016년 8월 31일 심사, 2016년 11월 30일 게재확정

\* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2015S1A3A2046711).

\*\* SK Infosec 사원(arrisian@naver.com)

\*\*\* 연세대학교 정보대학원 석사과정(youn\_win@yonsei.ac.kr)

\*\*\*\* 교신저자, 연세대학교 정보대학원 교수(kimhw@yonsei.ac.kr)

## I. 서론

국내에서의 잦은 사이버 테러와 위협으로 인해 정보보호 관련 규정이 크게 강화되고 각종 보안 의무 조치가 늘어났다. 2014년 1월에는 국내 주요 카드사의 1억 4,000만 건이 넘는 개인정보가 유출되었고, 2014년 12월 시작된 한국수력원자력 사태에서는 지난 2015년 8월 4일의 2차 해킹 자료가 언론사에 발송되었을 뿐만 아니라 12월 4일에 새로운 최신 악성 프로그램 모듈이 발견되는 등, 한국수력원자력 사태는 아직 현재 진행 중이다(아이티투데이, 2015). 이에 청와대 임종인 안보특별보좌관에 따르면, 정부에서는 청와대 국가 안보실 산하에 사이버보안 컨트롤타워 기능을 강화하기 위하여 사이버 안보 비서관을 신설하였으며, 정보보호산업 발전의 토대 마련을 위해 정보보호산업의 진흥에 관한 법률을 제정하여 2015년 12월 23일 시행하였다. 이러한 정부의 정책에도 불구하고 2015년의 경영실적에서는 정보보호 분야 관련 기업의 실적은 약세를 보이고 있다(아이뉴스, 2015).

이와 같이 보안 시장의 수익 악화가 계속됨에 따라 국내 정보보호 업계의 새로운 먹거리를 찾고, 새로운 시장을 개척하기 위한 노력이 증대되고 있다. 또한, 오늘날의 정보보호에서는 정보보호를 위한 기술만이 아닌 사용자의 정보보호인식의 중요성이 부각되고 있다(임재호, 2006). 한국수력원자력 사태의 사례와 같이 시스템이 아닌 사람의 취약점을 공략하는 사회 공학적(Social Engineering) 해킹이나 제로데이 공격(Zero Day Attack) 등은 기술적으로 원천봉쇄하는 것이 사실상 불가능하며, 이에 따라 정보보호인식을 강화하기 위한 노력이 매우 중요한 시점이라고 할 수 있다.

그러나 정보보호인식 관련 기존 연구에서는 대부분이 정보보호인식과 행위에 영향을 미치는 요인에 대한 연구에 국한되어 있으며(김영곤, 2010; 백민정 외, 2010; 손승희, 2013; Humaidi, et al., 2015),

사용자가 느끼는 정보보호인식에 대한 분석, 이에 대한 이슈 분석과 특징을 도출해 내는 연구가 이루어지지 않았다는 한계가 있다. 따라서 본 연구에서는 사용자의 정보보호인식에 관련된 한국, 미국, 중국의 트위터 데이터를 상대적으로 분석한 후 주요 이슈와 특징을 도출하여 한국의 정보보호인식에 대한 강화 방안을 제시하고자 한다.

## II. 문헌 연구

### 1. 정보보호인식 관련 선행 연구

정보보호인식은 정보보호에 대한 일반적인 지식과 조직 정보보호에 대한 인식으로 정의된다(Cavusoglu, 2009). 정보보호인식에 대한 기존 선행 연구에서는 주로 설문 분석 방법을 기반으로 하는 연구와(김영곤, 2010; 백민정 외, 2010; 손승희, 2013; Humaidi, et al., 2015), 특정 기업이나 대상을 선정한 사례 조사 형태의 연구가 진행되었다.

이와 같이 관련 기존 연구에서는 사용자가 느끼는 정보보호인식에 대한 분석 및 특징을 도출해 내는 연구가 이루어지지 않았으며, 국외의 사용자에 대한 분석과 분석 방법론이 제안되지 않았다는 한계가 있다. 따라서 본 연구에서는 사용자의 정보보호인식에 관련된 한국, 미국, 중국의 트위터 데이터를 분석하여 주요 이슈를 분석하고 특징을 도출하여 한국의 정보보호인식에 대한 강화 방안을 제시하고자 한다.

### 2. 토픽 모델링 방법 및 선행 연구

토픽 모델링은 문서의 주제를 도출하기 위해 텍스트 내 단어를 분석하는 방법론이다(Bauer, et al., 2015; Griffiths, et al., 2004). 특히 Blei, et al.(2003)가 제안한 LDA(Latent Dirichlet Allocation) 알고리즘은 단어가 독립적이지 않다는 가정에 따라 문서의 숨겨진 주제를 찾는 알고리즘으로, 토픽 모델링 연구에

서 표준 도구로 사용되어 왔다. 본 연구에서 분석을

$$\prod_{i=1}^k P(\beta_i | \eta) \prod_{d=1}^D P(\theta_d | \alpha) \left( \prod_{n=1}^N P(z_{d,n} | \theta_d) P(w_{d,n} | \beta_{i^k} z_{d,n}) \right) \quad (1)$$

위해 사용한 MALLET의 LDA 알고리즘은 수식(1)과 같다.

수식의 알고리즘은 개별 단어와 문서를 통해 문헌의 주제 비율(Topic Proportions,  $\theta$ ), 단어별 주제

(Per-word Topic Assignment,  $Z$ ), 그리고 주제(Topics,  $\beta$ )를 예측한다. 알고리즘에서  $\alpha$ 와  $\eta$ 는 하이퍼 파라미터(Hyper parameter)로 전체 문헌에 대해 같은 값을 갖는다.  $\alpha$ 는 각각의 문서가 어떠한 주제 비율로 구성될지를 나타내는  $\theta$ 값을 결정하는 파라미터이며,  $\theta$ 의 Dirichlet 분포 형태를 결정한다.  $\eta$ 는 각 단어가 어떠한 주제들의 비율로 구성될지를 나타내는  $\beta$ 의 Dirichlet 분포 형태를 결정하는 파라미터이며, Dirichlet 분포를 따르는  $\theta$  값에 따라 문헌 내에 존재

〈표 1〉 토픽 모델링 관련 선행 연구

저자	연구 주제	연구 분야	연구 결과
Mann, et al.(2006)	Bibliometric Impact Measures Leveraging TMann, et al.(2006)opic Analysis	연구 동향	컴퓨터공학 분야의 논문을 대상으로 토픽 모델링을 적용하고 논문의 영향력 측정에 유용함을 증명
Hillard, et al.(2008)	Computer-Assisted Topic Classification for Mixed-Methods Social Science Research		사회과학분야에서 알고리즘을 결합한 방법을 활용하여 연구 동향 분석 비용을 낮춤
박자현 외 (2013)	토픽 모델링을 활용한 국내 문헌정보학 연구동향 분석		문헌정보학 학술지별 주요 연구주제를 밝히고 공통점 차이점을 비교 분석
김수연 외 (2015)	텍스트 마이닝 기법을 이용한 컴퓨터공학 및 정보학 분야 연구동향 조사		DBLP 학술회의 데이터를 기반으로 컴퓨터공학과 정보학 분야의 20가지 토픽을 확인
유소영 (2015)	자아 중심 네트워크 분석과 동적 인용 네트워크를 활용한 토픽 모델링 기반 연구동향 분석에 관한 연구		데이터셋의 주제적 응집성이 높은 경우, 출현 단어의 특성인 혼잡도를 고려하는 것이 동적 인용 네트워크를 고려하는 것보다 더 적절함을 확인
Hong, et al.(2010)	Empirical Study of Topic Modeling in Twitter	소셜 미디어 분석	트위터에서 토픽 모델링 분석 기법의 품질과 효과를 사용자 분석 결과와 비교 확인
Wang, et al.(2013)	Real time event detection in twitter		토픽 모델링 기반의 트위터의 이슈 탐지를 위한 모델을 제안
배정환 외 (2013)	텍스트 마이닝을 이용한 2012년 한국대선 관련 트위터 분석		트위터 트렌드 마이닝 시스템을 개발하고 대선에 대한 사례 연구를 통해 검증
진설아 외 (2013)	트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구		트위터가 언론 뉴스에 즉각적으로 반응하며 부정적 이슈를 빠르게 확산시키는 것을 확인
박중도 (2015)	커뮤니티 기반 Q&A서비스에서의 질의 할당을 위한 이용자의 관심 토픽 분석에 관한 연구		커뮤니티 기반 질의응답 토픽 분석을 통해 적합한 응답자 식별에 활용
차운정 외 (2015)	소셜미디어 토픽모델링을 통한 스마트폰 마케팅 전략 수립 지원		트위터에서 토픽 모델링을 통해 스마트폰의 마케팅 전략 및 R&D 고려사항을 도출
조승연 외 (2015)	고객 온라인 구매후기를 활용한 추천시스템 개발 및 적용	추천 시스템 개발	토픽 모델링과 커널 학습 기반의 융합적 추천시스템을 개발

하는 단어들의 주제인  $z$ 가 결정된다. 각 단어의 주제를 나타내는  $z$ 값과 각 단어의 주제 전체에 대한 비율인  $\beta$  값에 의해 단어  $w$ 가 정해지게 된다(박자현 외, 2013; Blei, et al., 2003).

토픽 모델링 분석을 이용한 연구는 주로 연구 동향 분석과 소셜미디어 분석 분야에서 진행되어 왔다. 연구 동향 분석 분야에서는 컴퓨터공학, 정보학과 같은 세부 연구 분야의 트렌드를 분석한 연구(Mann, et al., 2006; 박자현 외, 2013)와 알고리즘을 결합하고 분석 효율을 높이고자 하는 연구가 수행되었다(Hillard, et al., 2008). 소셜 미디어 분석 측면에서는 주로 트위터 데이터 분석 위주의 연구(진설아 외, 2013; Hong, et al., 2010)와 분석 모델을 제안하고 개선하는 연구가 수행되었다(박종도, 2015; 배정환 외, 2013; Wang, et al., 2013).

관련 기존 연구를 살펴보면 트위터 데이터에 대한 소셜 미디어 분석 연구는 수행되었으나, 정보보호인식에 관련된 연구는 수행되지 않았으며, 감성 분석을 병행하여 분석하는 연구도 이루어지지 않았다. 따라서 본 연구에서는 정보보호인식을 연구 주제로 트위

터 데이터를 분석하여 주요 이슈를 분석하고 특징을 도출하며, 감성 분석 기법을 적용하여 사용자의 정보 보호인식을 측정하는 지표로 사용하고자 한다. 토픽 모델링 관련 선행 연구의 주제와 연구 결과는 <표 1>과 같다.

### 3. 감성 분석 방법 및 선행 연구

감성 분석이란 텍스트에서 사람들의 주관적 성향과 의견 등을 분석하는 기술을 말한다. 감성 분석의 정확성을 높이기 위해 신뢰도 높은 감성어 사전을 구축하기 위한 연구들이 진행되었다(김유신 외, 2012; 안정국 외, 2015; Taboada, et al., 2011). 소셜 미디어 분석 측면에서는 트위터 감성 분석을 통한 대선 결과 예측, 뉴스와 주가 사이의 유의 관계 확인, 패션 트렌드 파악에 대한 연구(김유신 외, 2012; 이윤주 외, 2014; Tumasjan, 2010)가 수행되었다.

이와 같이 감성 분석 관련 기존 연구에서도 트위터 데이터에 대한 소셜 미디어 분석 분야에 대한 트렌드를 파악하기 위한 연구들은 수행되었으나 정보보호

<표 2> 감성 분석 관련 선행 연구

저자	연구 주제	연구 분야	연구 결과
Baccianella, et al.(2010)	SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining	감성 사전	감성 분석을 위한 사전을 구축하고 이에 대해 평가
Taboada, et al.(2011)	Lexicon-based methods for sentiment analysis		상품평 등에 활용하기 위한 감성 사전을 구축
안정국 외 (2015)	집단지성을 이용한 한글 감성어 사전 구축		집단지성을 활용하여 온라인 감성어 사전 플랫폼을 구축하고 공개
Tumasjan, et al.(2010)	Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment	소셜 미디어 분석	트위터 감성 분석을 통해 대선 결과를 예측
김유신 외 (2012)	뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사결정모형		뉴스와 주가 사이의 유의 관계를 확인
이윤주 외 (2014)	SNS 텍스트 콘텐츠를 활용한 오피니언마이닝 기반의 패션 트렌드 마케팅 예측 분석		트위터 데이터를 분석하여 패션 트렌트 예측에 활용
Hridoy, et al.(2015)	Localized twitter opinion mining using sentiment analysis		트위터에서의 감성 분석을 통해 iPhone6에 대한 사용자 반응 관찰 및 기존 리뷰와의 비교/분석

인식에 관련된 연구는 수행되지 않았으며, 정보보호 인식 연구에서 감성 분석 기법을 적용하여 연구를 진행하면 해당 값을 점수로 환산하여 제시할 수 있다는 장점이 있다. 따라서 본 연구에서는 정보보호인식을 측정하는 지표로 감성 분석 결과를 점수로 변환한 값을 활용하여, 트위터 데이터를 기반으로 사용자 정보보호인식을 분석하는 연구를 수행하고자 한다. 감성 분석 관련 선행 연구의 주제와 연구 결과는 <표 2>와 같다.

### Ⅲ. 연구방법

본 연구는 트위터에 사용자가 게시한 데이터를 적용하여 정보보호 이슈를 도출하고, 점수 결과와 비교 분석하기 위하여 토픽 모델링과 감성 분석 기법을 통해 연구를 진행하였다. 이를 위한 본 연구의 개략적인 연구 절차는 <그림 1>과 같다.

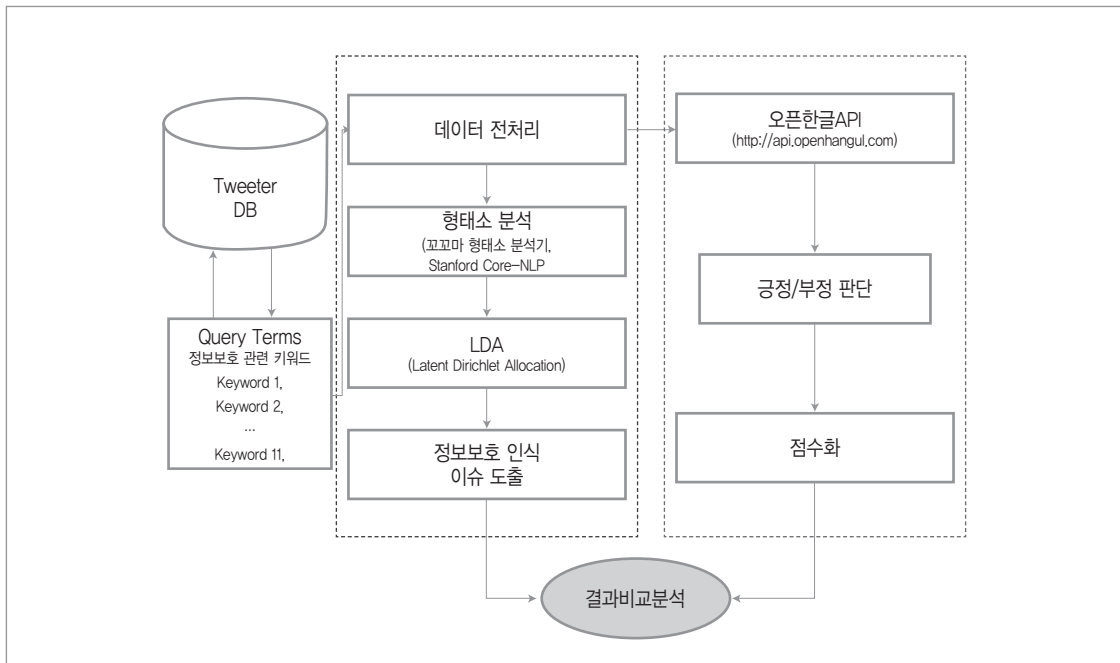
### 1. 연구 키워드 설정

본 연구의 수행을 위하여 트위터에서 데이터 수집을 하기 위한 검색 키워드를 선정하였다. 연구 키워드 선정을 위하여 각 기업의 정보보호 담당자, 보안 컨설턴트, 정보보호 전공자, 총 20명을 대상으로 전문가 인터뷰를 실시하였다. 인터뷰 대상 구성과 응답은 <표 3>, <표 4>와 같다.

정보보안에 대한 데이터 수집에 적합한 키워드를 선정하기 위한 목적으로 수행된 전문가 인터뷰 결과,

<표 3> 인터뷰 대상

구분	인원(명)
기업 정보보호 담당자	3
보안 컨설턴트	5
정보보호 전공자	12
합계	20



<그림 1> 연구 절차 개요

〈표 4〉 인터뷰 결과

키워드	횟수	비율(%)	응답 사례
보안	20	100	모바일 보안, IoT 보안 등 이슈가 되는 것에 대한 분석 필요
정보보호	20	100	정보보호라는 키워드는 꼭 포함되어야 할 것
프라이버시	20	100	프라이버시 침해, 프라이버시권 등
개인 정보보호	20	100	개인정보보호, 개인정보보호법 등
개인정보	16	80	개인정보 유출, 개인정보 유출 방지 등
정보 프라이버시	12	60	정보 프라이버시, 정보 준수 등
기밀성	12	60	보안의 3요소인 기밀성, 무결성, 가용성에 대한 응답
무결성	12	60	
가용성	12	60	
부인방지	11	55	부인방지, 전자 서명 등
인증	10	50	공인인증, 인증 관리, 인증 센터 등

보안', '정보보호', '프라이버시', '개인정보보호', '기밀성', '개인정보', '정보', '무결성', '가용성', '부인방지', '프라이버시' 등 총 11개 키워드를 선정하였다. 선정한 한국의 11개 키워드에 매칭이 되도록 영문, 중문 연구 키워드를 한국어 키워드를 번역하여 연구 키워드로 선정하였다. 해당 키워드가 실제 사용되는지 확인하기 위하여 미국의 대표적인 포털 서비스인 구글과 중국의 포털 서비스 업체인 바이두, 그리고 미국과 중국의 정보보호연구들을 통해 번역된 키워드를

보정하였으며 〈표 5〉와 같이 최종 키워드 셋을 선정하였다.

## 2. 데이터 수집 및 전처리

본 연구는 2015년 10월 6일부터 2015년 11월 6일까지 32일간 한국 299,871건, 미국 781,822건, 중국 801,011건, 총 1,882,704건의 트위터 데이터를 수집하여 연구를 진행하였다. 한국과 미국, 중국, 국가별로 11개의 연구 키워드를 포함하고 있는 일반 사용자

〈표 5〉 최종 연구 키워드

한국	미국	중국
보안	security	安全
정보보호	data protection	信息保
프라이버시	privacy	私
개인정보보호	privacy protection	人信息保
개인정보	personal information	人信息
정보 프라이버시	information privacy	信息私
기밀성	confidentiality	机密性
무결성	integrity	完整性
가용성	availability	可用性
부인방지	non-repudiation	不可抵性
인증	authentication	认证

가 작성한 트윗 데이터를 수집하였다. 데이터 수집에는 트위터 API를 이용하였으며, 수집에 활용하기 위한 프로그래밍 언어로 Java와 R을 이용하였다.

트위터는 140자로 정해져 있는 ‘트윗(tweet)’이라는 틀에 사용자의 의견을 담은 소셜 미디어 서비스로, 트윗을 공유하는 기능인 ‘리트윗(retweet)’으로 정보의 전파가 빠르게 일어난다는 특징이 있다(류우종 외, 2013). 이에 따라 사용자들의 관심사와 트렌드를 확인하기 적합한 소셜미디어로 기업에서의 시장 조사뿐만 아니라 많은 연구에서 활용되고 있다(박종도, 2015; 배정환 외, 2013; Hong, et al., 2010; Wang, et al., 2013). 따라서 본 연구에서는 트위터 데이터를 수집하여 한국과 미국, 그리고 중국의 정보보호인식을 확인하고 강화방안을 제시하고자 한다.

본 연구에서는 중국에서 가장 많이 사용되고 있는 웨이보가 데이터 수집이 불가하여 3개국 모두 트위터를 사용하여 데이터를 수집하였다. 시장 조사 전문기관인 Global Web Index에 따르면, 2012년 2분기 중국의 트위터 활성 사용자가 3,550만 명이라고 밝혔으며, 미국은 2,290만 명, 한국은 120만 명이었다(eMarketer, 2012). 또한 Global Web Index는 2014년 3분기 가상 사설망(VPN)을 이용하여 중국 정부의 검열을 우회하여 트위터나 페이스북을 사용하는 이용자가 9,300만 명에 달한다고 밝혔다(Global Web Index, 2014). 따라서 본 연구에서는 중국의 소셜미디어 데이터로 웨이보가 아닌 트위터를 수집하여 연구를 진행하였다.

### 3. 토픽 모델링

본 연구에서는 트위터의 일반 사용자들이 작성한 정보보호 관련 트윗 데이터에서 각국의 정보보호인식을 확인하고 분석하기 위하여 문서 내의 잠재된 주제를 도출하는 기법인 토픽 모델링 기법을 이용하였다. 이를 위해 Java로 구현된 Mallet의 LDA(Latent Dirichlet Allocation) 토픽 모델링 알고리즘을 사용하였다. 각 국가별 정보보호인식에 대한 토픽을 국가별 10개, 총 30개의 주제를 추출하였다.

### 4. 감성 분석 및 스코어링

본 연구는 3개 국어 모두 오픈한글 API(<http://api.openhangul.com>)를 이용하여 감성 분석을 실시하였다. 중국어의 경우 데이터의 직접 분석의 어려움으로 구글 번역기를 통해 영어로 번역하여 분석하였다. 트윗 단위로 긍/부정을 판단하였으며, 트윗을 토픽의 카테고리별로 범주화 하여 카테고리 별 트윗에 대한 긍/부정 트윗의 비율을 계산하였다. 해당 결

$$\text{점수 값(5점 만점)} = \text{트윗 긍/부정비율}(\%) * 5 \quad 2)$$

과를 환산하여 5점 만점의 값으로 점수화 하였다. 공식은 수식(2)와 같다.

또한 각 국가별 점수 결과를 비교하기 위하여 같은 비교 영역을 설정 하였다. 본 연구에서는 해당 비교 기준을 정보보호 관리체계의 국제 표준인 ISO27002:2013의 기술적, 관리적, 물리적 보안의 세 영역으로 설정하였으며, ISO27002 통제 항목의

〈표 6〉 ISO27002 통제 항목

영역	통제항목
기술적 보안 영역	접근 통제, 암호화, 운영 보안, 통신 보안, 시스템 도입, 개발, 유지보수
관리적 보안 영역	정보보호 정책, 정보보호 조직, 인적 자원 보안, 자산 관리, 공급자 관계, 정보보호 사고 관리, 업무연속성 관리의 정보보호 측면, 준거성
물리적 보안 영역	물리적 및 환경적 보안

키워드를 비교 영역을 구분하기 위한 키워드로 사용하였다. ISO27002:2013이란 기존의 보안 표준에 현실을 좀 더 반영하여, 2013년 새롭게 수정된 실무적인 지침을 위한 정보보호 관리체계의 국제 표준 ISO27000의 한 계열로써 공통적으로 적용될 수 있는 보안표준이기 때문에 실 비교를 위한 키워드로써 적합하다고 판단하였다. ISO27002 통제 항목은 <표 6>과 같다.

#### IV. 분석결과

##### 1. 토픽 모델링

###### 1) 한국

다음은 한국 트윗 데이터의 토픽 모델링 분석 수행 결과이다. 10개의 토픽이 도출되었다. 7개의 기술적 보안 영역이 도출되었으며, 3개국 가운데 유일한 물리적 보안 영역의 토픽인 '사내 보안'이 도출되었다.

'1. 모바일 결제 보안'의 경우, 소액 결제, 신용카드, 설치 등의 단어들에 주로 언급되는 트윗으로, 기술적 보안 영역에 해당하는 토픽이다. 이 토픽은 30

만 원 이상의 카드 결제 시 공인인증서 의무 사용에 따른 불편함과 오류에 대한 이슈이다. '6. 모바일 앱 보안' 토픽에서는 마켓, 액티브 엑스, 링크, 암호화 등이 주요 키워드로 도출되었으며, 링크를 통해 이뤄지는 어플리케이션의 설치나 연동, 액티브 엑스 사용이 불가하여 사용할 수 없는 서비스 등을 주요한 이슈로 요약할 수 있다. 2015년 3월 18일 금융위원회의 전자금융감독규정 개정에 따라 공인인증서 의무사용 조항이 삭제되었으나, 아직 많은 온라인 서비스와 결제 시스템에서는 공인 인증서를 대체하지 못하고 있으며, 이에 따라 이용자의 불편이 가중된다는 것이다.

'3. 피싱 사기'와 '4. 파밍 사기' 토픽은 사회공학 공격의 주요한 방법으로, 링크, 발견, 사건, 불안, 사고, 사기, 사이트 수정, 등이 주요한 키워드이다. 국내에서 피싱과 파밍에 대한 이슈와 불안감을 확인할 수 있다. 피싱 사기와 파밍 사기, 두 토픽은 기술적 보안 영역의 토픽으로 도출되었다. 새로운 피싱과 파밍 사기가 알려짐에 따라 사용자의 불안감이 고조되며, 이는 사회적 비용으로 나타난다. 따라서 정부에서는 정책적 노력과 홍보에 앞장서야 할 것이며, 개

<표 7> 토픽 모델링 수행 결과-한국

번호	토픽	키워드
Technical protection area		
1	모바일 결제 보안	보안, 신용카드, 설치, 업그레이드, 소액, 법, 예러, 민원
2	앱스토어 본인인증	보안, 본인, 성공, 아이디, 보관, 앱스토어, 실패, 보호
3	피싱사기	보안, 사고, 사기, 보고, 사이트, 수정, 로그인, 솔루션
4	파밍사기	보안, 실시간, 링크, 발견, 사건, 업체, 메모, 불안
5	모바일 페이 인증	보안, 애플, 삼성페이, 서비스, 문제, 실명인증, 승인, 앱인증
6	모바일앱 보안	보안, 마켓, 액티브 엑스, 반영, 로그, 링크, 사용자, 암호화
7	본인인증 서비스 불편	개인정보, 아이폰, 명의, 아이폰, 엉터리, 불편, 로그인, 문자 인증
Administrative protection area		
8	병원정보 유출	개인정보, 목적, 병원, 연구소, 보도, 배포, 본인, 명단
9	사내 보안	보안, 사내, 보안위협, 보안검색, 발급, 시스템, 백업, 변경
Physical protection area		
10	앱스토어 본인인증	보안, 본인, 성공, 아이디, 보관, 앱스토어, 실패, 보호

개인 역시 이러한 사기를 피할 수 있도록 주의를 기울여야 할 것이다.

‘2. 앱스토어 본인인증’에 해당하는 트윗은 앱스토어와 아이디, 본인 인증의 성공 및 실패에 대한 언급이 이뤄지는 트윗이다. 이와 유사한 토픽으로 ‘7. 본인인증 서비스 불편’ 토픽이 도출되었다. 해당 토픽의 트윗에서는 아이폰 본인인증 서비스, 로그인, 문자 인증 등이 주요 키워드로 도출되어 본인인증 서비스에 대한 불편이 주요하게 제기되었음을 확인할 수 있다. ‘5. 모바일 페이 인증’ 토픽은 삼성 페이, 애플, 실명인증, 문제 등의 주요 키워드로 이뤄져 있으며, 공통적으로 인증의 불편함에 관련된 이슈라고 할 수 있다. 세 토픽 모두 기술적 보안 영역의 토픽으로 도출되었다. ‘8. 정보보호 목표’ 토픽에서는 정보보호, 목표, 수준, 불편, 비용이 주요한 키워드로 도출되었으며, 관리적 보안 영역의 토픽으로 도출되었다.

‘9. 병원 정보 유출’과 ‘10. 사내 보안’ 토픽은 각각 약학정보원 의료정보 연구 이용 유출 사건과 삼성페이의 핵심기술을 보유하고 있는 삼성전자 자회사인 루프페이의 사내망 해킹에 따른 이슈에 대한 토픽이다. 병원 정보 유출의 경우 관리적 보안 영역의 토픽

이며, 사내 보안의 경우 물리적 보안 영역의 토픽으로 도출되었다. 한국의 토픽 모델링 수행 결과는 <표 7>과 같다.

2) 미국

<표 8>은 미국 트윗 데이터에 토픽 모델링 분석을 수행한 결과이다. 총 10개의 도출된 토픽 중, 2개의 기술적 보안 영역 토픽과 3개국 가운데 가장 많은 8개의 관리적 보안 영역의 토픽이 도출되었다.

‘1. 온라인 인증 시스템’과 ‘2. 모바일 지문 인증 시스템’ 토픽의 경우, 각 non-Repudiation, Oath, System과 Authentication, Android, Fingerprint가 주요한 키워드로 기술적 보안 영역의 토픽으로 도출되었다. ‘3. 데이터 보호법 개정 요구’ 토픽과 ‘6. 개인정보 국외 이전’ 토픽에서는 Information, Law, Act와 Data, Protection, law 등의 단어가 주로 언급되었다. ‘3. 데이터 보호법 개정 요구’ 토픽에서는 미국의 사이버 감청의 범위와 개인정보 이전에 대한 부분을 다루고 있다. 두 토픽은 유럽사법재판소가 미국과 체결했던 Safe Harbor 조약을 무효화하여 미국 기업이 유럽 국가들의 개인정보를 국외 이전할 수 없

<표 8> 토픽 모델링 수행 결과-미국

번호	토픽	키워드
Technical protection area		
1	온라인 인증 시스템	Security, Government, Oath, Order, Non-repudiation, State, Secrecy, System
2	모바일 지문 인증 시스템	Security, Authentication, Android, System, Fingerprint, Key, Apple, Password
Administrative protection area		
3	의료정보 연구 이용	Security, Privacy, Law, Digital, Information, Id, Secrecy, Act
4	데이터 보호 기준 개정	Security, Confidentiality, Protection, Datum, Health, Law, Patient, Doctor
5	개인정보 국외 이전	Security, Officer, Integrity, Question, Commission, Protection, Journalistic, Nation
6	디지털 개인정보 보호법 개정	Security, Protection, Data, EU, Court, Privacy, Law, Rule
7	소셜 미디어 개인정보 이용 동의	Security, Privacy, People, Information, Privacy, Time, Person, Policy
8	정보보호정책 수립 의무화	Security, Media, Confidentiality, Integrity, Nation, Secrecy, Blog, Plan
9	사이버 감청	Security, Penalty, Privacy, Competition, Element, Director, Web, Client
10	온라인 인증 시스템	Security, Integrity, Right, Value, Money, Trust, Character, Person

게 된 이슈에 따른 것이다.

또한, '7. 디지털 개인정보 보호법 개정'과 '9. 정보 보호정책 수립 의무화'토픽에서는 각각 Information, Privacy, Policy와 Privacy, Security, Penalty, Director에 대한 언급이 주로 이뤄진 트윗이었으며, '7. 디지털 개인정보 보호법 개정' 토픽은 캘리포니아의 디지털 개인정보 보호법 개정에 따른 이슈에 따른 것이다. '9. 정보보호정책 수립 의무화' 토픽은 미국의 자동차 제조사에서 정보보호정책을 제출하지 않는 경우 벌금을 부과할 수 있는 법안을 미하원이 입법 추진한 것에 따른 이슈이다. 두 토픽 모두 관리적 보안 영역의 토픽으로 도출되었다. '4. 의료정보 연구 이용'과 '8. 소셜 미디어 개인정보 이용 동의' 토픽은 Health, Protection, Datum, Doctor와 Media, Blog, Integrity, Confidentiality가 주요 키워드로 도출되었으며, '5. 데이터 보호 기준 개정'에서는 Security, Protection, Commission이 주요한 키워드로 유럽 정보보호 위원회의 데이터 보호 기준 개정에 대한 트윗들이 주를 이루었다. '8. 소셜 미디어 개인정보 이용 동의' 토픽의 경우, 23개 주에서 발의되어 2016년 개정이 의논 중인 소셜 미디어 관련 개인정보보호법안에 관한 이슈로 온라인 계정에 대한 개

별적 동의를 강제하는 것이 주요 골자이다. '10. 사이버 감청' 토픽은 Person, Right, Trust가 주요 키워드로 나타났으며, 미국 국가안전보장국의 사이버 감청에 따른 개인정보수집과 개인정보권리에 관한 이슈이다. 미국의 토픽 모델링 수행 결과는 <표 8>과 같다.

### 3) 중국

다음은 중국 트윗 데이터의 토픽 모델링 분석 수행 결과이다. 각 기술적, 관리적 보안 영역에서 5개의 토픽이 도출되었다. 웹, DB 관련 토픽과 사이트 검열이 특징적인 토픽으로 도출되었다.

'1. 웹 서버 보안'과 '3. 웹 애플리케이션보안', '4. 웹 트래픽' 토픽에서는 공통적으로 Security와 Web이 언급되고 있었으며, Data와 Internet도 주요 키워드로 도출되었다. '2. DB 데이터 보안'의 경우 DB, Data, Firewall이 주요 언급되는 단어였으며, 웹과 DB에 대한 4가지 토픽들은 지난 2015년 7월 6일 중국 사이버 보안법 채택에 따른 이슈들에 대한 트윗이다. '5. 사이버 공격 방어' 토픽에서는 Security, Attack, Defense가 주로 언급된 단어였으며, '9. 사이트 검열'에서는 Security, Scan, Check가 주

<표 9> 토픽 모델링 수행 결과-중국

번호	토픽	키워드
Technical protection area		
1	웹 서버 보안	Security, Web, Safety, Support, Option, Person, Data, Internet
2	DB 데이터 보안	Security, DB, Work, Data, Safety, Firewall, Time, Domain
3	웹어플리케이션 보안	Security, Web, Site, Shop, App, Id, Work, Information
4	웹 트래픽	Security, Traffic, Stop, Internet, Information, Injury, Web, Data
5	사이버 공격 방어	Security, Attack, Location, Speed, Password, Type, Time, Defense
Administrative protection area		
6	정보보호사고 관련 법안	Security, Data, Mobile, Disclosure, Person, Information, Thing, Anxiety
7	정보통신산업 관련 처벌	Security, Original, Law, Company, Accident, Internet, Product, Act
8	사이트 검열	Security, Fine, Government, Data, Method, Communication, Penalty, Law
9	정보보호사고	Security, Law, Support, Scan, System, Check, Site, Datum
10	웹 서버 보안	Security, Safety, Fake, Information, Law, Prevention, Problem, Crime

요 키워드였다. 중국의 개정된 헌법 291조에 따르면, 인터넷과 같은 온라인 매체를 통해 허위 정보를 유포할 경우 3년에서 7년에 이르는 중형에 처하며, 지난 8월 텐진항 폭발사고 이후 인터넷 통제 수위를 대대적으로 강화하고 있다. ‘7. 정보보호사고 관련 법안’, ‘8. 정보통신 산업 관련 처벌’, ‘10. 정보보호사고’에서는 각각 Security, Accident, Law와 Security, Communication, Penalty, 그리고 Security Crime, Prevention이 주요한 키워드였다. ‘6. 모바일 정보유출 우려’ 토픽의 경우, Security, Information, Data, Mobile 등이 주요 키워드로 도출되었다. 웹과 DB 관련 4개 토픽과 사이트 검열이 특징 토픽으로 도출되었다. 이는 중국의 사이버 보안법 채택 이슈로 도출된 토픽이며, 검열의 경우, 온라인 검열이 구글과 페이스북, 트위터에서 이슈가 되었던 중국의 특징적인 이슈이다. 중국의 토픽 모델링 수행 결과는 <표 9>와 같다.

#### 4) 국가별 비교

각 국가에서 도출된 토픽들을 살펴보면, 한국의 경우 기술적 보안 영역에서 가장 많은 토픽이 도출되었으며, 미국에서는 관리적 보안 영역의 토픽이 8개로 도출되었고, 중국에서는 기술적 보안 영역과 관리적 보안 영역에서 각각 5개의 토픽이 도출되었다.

한국의 토픽들을 살펴보면, ‘2. 앱스토어 본인인증’의 경우 국내 한 모바일 간편 결제 서비스의 앱스토어 인증 불가 문제에 대한 트윗들이 주를 이뤘으며, 유사한 토픽으로 도출된 ‘7. 본인인증 서비스 불편’에 해당하는 트윗에서는 본인인증 서비스에 대한 불편이 주요하게 제기되었다. 미국 또한 ‘1. 온라인 인증 시스템’ 토픽이 도출됨에 따라, 한국과 미국은 인증 서비스, 인증 시스템에 서비스 불편 사항 개선과 기술적 보완이 필요할 것으로 보인다.

한국의 ‘3. 피싱 사기’와 ‘4. 파밍 사기’ 토픽은 신

<표 10> 토픽 모델링 수행 결과—국가별 비교

토픽	국가	감성 분석 점수	토픽	국가	감성 분석 점수
Technical protection area			Administrative protection area		
모바일 결제 보안	Korea	1.7	정보보호 목표	Korea	3.1
앱스토어 본인인증	Korea	2.0	병원정보 유출	Korea	1.4
파밍 사기	Korea	1.5	데이터 보호법 개정 요구	USA	2.6
피싱 사기	Korea	1.4	의료정보 연구 이용	USA	3.4
모바일 페이 인증	Korea	1.7	데이터 보호 기준 개정	USA	2.6
모바일앱 보안	Korea	1.5	개인정보 국외 이전	USA	2.5
본인인증 서비스 불편	Korea	1.7	디지털 개인정보 보호법 개정	USA	2.5
온라인 인증 시스템	USA	3.7	소셜 미디어 개인정보 이용 동의	USA	2.3
모바일 지문 인증 시스템	USA	3.4			
웹 서버 보안	China	3.0	정보보호 관련 처벌 감독	USA	3.1
DB 데이터 보안	China	3.2	데이터 보호법 개정 요구	USA	3.7
웹 애플리케이션 보안	China	2.6	모바일 정보유출 우려	China	2.4
웹 트래픽	China	2.3	정보보호사고 관련 법안	China	2.4
사이버 공격 방어	China	2.2	정보통신산업 관련 처벌	China	2.4
Physical protection area			사이트 검열	China	2.6
사내 보안	Korea	1.5	정보보호사고	China	2.4

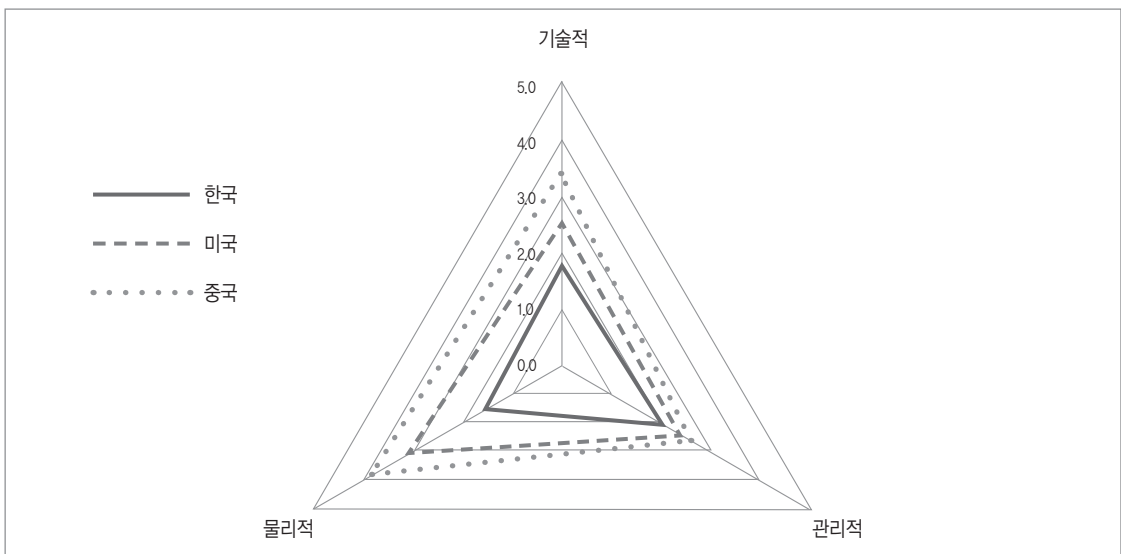
중 보이스 피싱 사기에 따른 이슈이며, 중국에서는 '6. 모바일 정보유출 우려'가 주요 토픽으로 도출되었다. 특히 '3. 피싱 사기'와 '4. 파밍 사기'의 경우 한국과 중국에서 정보유출과 그에 따른 피해를 예방하기 위한 정책적 노력과 홍보, 그리고 개개인의 경각심이 필요할 것이다. 한국의 '8. 정보보호 목표' 토픽에서는 정보보호 목표 수준과 정보보호에 따른 불편사항, 비용에 대한 토픽이 도출 되었다. 따라서 보안성과 편의성을 모두 고려한 정보보호 목표 수준에 대한 논의가 필요할 것으로 판단된다. 한국에서의 '9. 병원정보 유출' 토픽은 약학정보원 의료정보 유출 사건에 따른 이슈로 도출 되었다. 미국에서의 '4. 의료 정보 연구 이용' 토픽은 태아에게서 얻은 의료 정보를 연구에 사용하는 것에 대한 이슈였다. 중국의 경우 한국, 미국이 모바일, 온라인 인증 시스템 관련 이슈가 주요한 것과 달리 '1. 웹 서버 보안', '2. DB 데이터 보안', '3. 웹 애플리케이션 보안', '4. 웹 트래픽'과 같은 토픽들이 도출되었는데, 이는 지난 2015년 7월 6일 중국 사이버 보안법 채택에 따른 토픽으로 추출되었으며, 또한, '9. 사이트 검열'이 특징적인 토픽으로 도출되었다. 이에 따라 최근 국제 공공 보안 박람회를 비

롯하여 해외 진출의 주요 국가인 중국에서의 웹 방화벽 등의 장비와 솔루션을 바탕으로 시장 개척을 도모할 수 있을 것으로 생각된다. 국가별 비교 결과는 <표 10>과 같다.

## 2. 감성 분석 및 스코어링

### 1) 국가별 비교

본 연구에서는 ISO27002:2013의 기술적, 물리적, 관리적 보안 영역에 따라 수집한 트윗을 감성 분석하였으며, 5점 만점의 점수로 환산하였다. 환산한 점수는 5점에 가까울수록 긍정적인 트윗이 많은 것이며, 0점에 가까울수록 부정적인 트윗이 많은 비중을 차지하는 것을 의미한다. 한국의 기술적 보안 영역 점수는 1.8점, 관리적 보안 영역의 점수는 2.1점, 물리적 보안 영역의 점수는 1.6점이다. 미국의 경우 기술적 보안 영역이 2.5점, 관리적 보안 영역의 점수는 2.4점, 물리적 보안 영역의 점수는 3.1점으로 도출되었다. 중국에서는 기술적 보안 영역이 3.4점, 관리적 보안 영역의 점수가 2.7점, 물리적 보안 영역의 점수가 3.8점이었다. 스코어링 결과에서는 세 영역 모두



<그림 2> 정보보호인식에 대한 감성분석 점수 비교

에서 중국이 가장 높은 점수로 측정되었으며, 다음으로 미국, 한국 순으로 측정되었다. 국가별 점수 결과는 <그림 2>와 같다.

2) 통제 항목별 비교

세부적으로 기술적, 관리적, 물리적 보안 영역의 ISO27002:2013의 통제 항목별 점수를 측정하였다. 점수 결과의 세부 항목별 비교는 <표 11>과 같다.

한국과 미국의 ‘접근 통제’와 ‘암호화’의 점수는 각각 1.6, 1.5와 1.1, 0.9로 측정된 반면, 중국의 점수는 3.8점과 4.5점으로 높게 측정되었다.

이는 미국에서의 구글과 애플 등이 도입한 메시지 암호화 정책에 따른 논란으로 테러 대응에 어려움이 있다는 정부 측과의 대립에 따른 이슈가 영향을 미친 것으로 보이며, 또한 미국이 환태평양경제동반자 협정(TPP)을 통해 중국의 해킹과 디지털 보호 무역 주

의를 금지하겠다는 이슈에 따라 자국 정보보호에 대한 서로 다른 입장 차가 반영된 것으로 보인다.

‘업무 연속성 관리의 정보보호 측면’의 경우, 미국에서만 측정되고 한국과 중국에서는 측정되지 않았는데, 이는 상대적으로 정보보호 전문가나 현업의 담당자가 사용하는 용어이며, 미국에서는 9.11 테러 당시 업무 연속성 관리가 기업의 생사를 좌우했던 경험이 있기에 데이터 복구의 중요성이 매우 중요하게 여겨지는 것으로 생각된다.

3) 정보보호인식 강화

분석 결과 주요 이슈들에서 총 5가지 문제점과 그에 따른 대안을 도출할 수 있었다. 분석을 통한 문제점 도출 및 대안은 <표 12>와 같다.

첫 번째로 공인 인증서 사용에 따른 불편과 비효율성에 관한 문제점을 도출할 수 있었다. ‘앱스토어 본

<표 11> 점수 결과 - 세부 항목 비교

통제 항목 (ISO27002:2013)	점수		
	Korea	USA	China
Technical Protection Area			
접근 통제	1.6	1.1	3.3
암호화	1.5	0.9	4.5
운영 보안	2.0	3.8	3.2
통신 보안	1.7	3.7	3.0
시스템 도입, 개발, 유지보수	2.3	2.9	3.0
Administrative Protection Area			
정보보호 정책	2.5	2.1	3.0
정보보호 조직	2.3	3.6	2.9
인적 자원 보안	2.5	3.5	2.9
자산 관리	2.4	2.0	2.2
공급자 관계	-	-	-
정보보호 사고 관리	1.5	1.3	2.5
업무 연속성, 정보보호 관리 측면	-	1.6	-
준거성	-	2.6	2.6
Physical security Area			
물리적 및 환경적 보안	1.6	3.1	3.8

〈표 12〉 분석을 통한 문제점 도출 및 대안

문제점	대안
<b>정부</b>	
공인인증서 사용에 따른 불편 및 비효율성	공인인증서 대체 서비스 의무화 및 금융사고 배상책임보험 가입 의무화
신중 보이스 피싱 사기에 따른 피해 및 우려	사기 방지 교육 강화
기술 유출에 대한 우려	기술유출 사전예방 지원, 및 사후 구제
<b>기업</b>	
의무 사용 회사 앱의 개인정보 침해	의무 사용 앱 개인 정보보호 가이드라인 수립
<b>학계</b>	
공인인증서 사용에 따른 불편 및 비효율성	국내외 사례 분석 및 대체 서비스 제안
정보보호에 따른 불편사항, 비용	유저블 시큐리티 논의

인 인증' 토픽의 경우 국내 한 모바일 간편 결제 서비스의 앱스토어 인증 불가 문제에 대한 트윗들이 주를 이루었으며, '본인인증 서비스 불편' 토픽은 아이핀과 같은 본인인증 서비스가 느리거나 오류가 발생해 이용자 불편이 야기되어 제기된 이슈로 감성 분석값 역시 2점과 1.7점으로 낮은 결과를 보였다. 또한 '모바일 결제 보안' 토픽의 주요 이슈는 30만 원 이상의 카드 결제 시 공인인증서를 의무적으로 사용 해야 하는 불편함과 해당 기능을 이용할 때 발생하는 오류가 불편하다는 이슈로 감성 분석 점수는 1.7점으로 측정되었다. 2015년 3월 18일 금융위원회는 전자금융감독규정을 개정하여 '공인인증서 또는 이와 동등한 수준의 안정성이 인정되는 인증 방법'을 사용하여야만 한다는 의무조항을 삭제하였지만 그럼에도 불구하고 아직 많은 사이트와 결제 시스템에서는 공인 인증서를 대체할 방법을 적용하지 못하고 있으며, 사고 발생 시의 책임 소재 문제로 공인 인증서 대체 서비스를 쉽게 지정하지 못하고 있다는 것을 확인할 수 있었고, 이에 따라 사용자 불편이 야기되고 있음을 확인할 수 있었다. '모바일 페이 인증' 토픽 역시도 페이 서비스 인증 시에 발생하는 오류에 대한 불만 사항과 휴대폰 인증이 필요하다는 불편에 대한 이슈가 대부분으로 감성 분석 점수 역시 1.7점으로 낮게 측정되었다. 따라서 정부 측면에서 기존의 공인인증서 의무

사용 조항을 폐기하는 것에 그칠 것이 아니라, 보다 현실적인 가이드라인과 기업의 대안 지정을 의무화하는 정책이 제정이 필요한 것으로 된다.

두 번째로 신중 보이스 피싱 사기에 따른 피해 우려와 문제점을 도출할 수 있었다. '피싱 사기'와 '파밍 사기'는 신중 보이스 피싱 사기에 따른 이슈로 감성 분석 점수 역시 각 1.5점과 1.4점으로 전체 측정치에서 가장 낮은 수준으로 측정되었다. 따라서 정부 측면에서의 피해를 예방하기 위한 정책적 노력과 홍보를 적극적으로 할 필요성이 있으며, 개개인 역시 보다 높은 경각심이 요구됨을 판단할 수 있었다.

세 번째로는 정보보호에 따른 불편과 비용 문제를 도출할 수 있었다. '정보보호 목표' 토픽은 정보보호 목표 수준과 정보보호에 따른 불편사항, 비용에 대한 내용이 주요한 이슈로 전체적으로 긍정적인 토픽이 더 많이 측정되었지만, 부정적인 토픽도 38%에 달하고 있다. 따라서 보안 수준은 유지하면서도 사용자의 편의성은 높인다는 개념인 '유저블 시큐리티(Usable Security)'에 대한 논의가 필요하다는 것을 판단할 수 있었다(머니투데이, 2015).

네 번째로는 의무사용 앱이 개인정보 침해 문제점을 도출할 수 있었다. '모바일 앱 보안' 사용이 강제된 회사 앱이 개인 정보 침해 논란에 휩싸인 이슈이다. 감성 분석 점수 측정 결과 1.5점으로 가장 낮은 수준

〈표 13〉 정보보호인식 강화를 위한 방안

주체	방안(대안)	내용
정부	공인인증서 대안 지정 의무화	기업이 적용할 수 있는 현실적인 가이드라인과 기업의 공인인증서 대안 지정을 의무화 하는 정책 제정 필요
	금융사고 배상책임보험 가입 의무화	사고 발생 시의 책임 소재 문제는 공인인증서의 대체 서비스를 쉽게 지정하지 못하는 원인중 하나로, 금융사의 과도한 리스크에 대한 안전장치 필요
	실시간 대국민 홍보 강화	새로운 피싱, 파밍의 경우 쉽게 예방하기 어려우므로 관계당국의 금융사, 기업과의 다각적인 공조체계와 실시간 홍보 체계 구축 필요
	의료정보 연구 이용법 개정 및 감독 강화	의료정보 연구 이용법 개정을 통하여 관련 산업 발전 및 감독 강화
	기술유출 사전예방 지원	기술보호 상담 및 보안 시스템 구축 지원 등 기술유출 사전예방을 위한 지원 강화
	기술유출 피해기업 사후구제	기술 분쟁 중재 위원회 기능 강화를 통한 피해기업에 대한 사후구제 지원
	회사 앱 개인정보보호 가이드라인 수립	회사 앱의 개인정보보호를 강제하는 가이드라인 수립 필요
기업	회사 앱 개인정보 침해 자체 기준 수립	회사 앱의 개인정보 침해에 대한 자체 기준 수립 필요
	서비스 사용성 및 대체 기능 강화	페이 서비스의 사용성 개선 및 대체 서비스 및 기능 강화 필요
학계	공인인증 대체 서비스 관련 국내외 사례 분석 및 대안 제안	공인인증 대체 서비스의 국내외 사례에 대한 분석 및 기업에서 적용 가능한 방안 탐색 제시 필요
	‘유저블 시큐리티’ 논의 필요	보안 수준은 유지하면서도 사용자의 편의성을 높인다는 개념인 ‘유저블 시큐리티’에 대한 논의를 통해 사용자의 편의성과 보안성을 모두 충족시킬 수 있는 방안 제시 필요

으로 측정되었으며, 기업 자체적으로 회사 앱의 개인정보 침해 기준을 세울 필요성이 있을 뿐만 아니라, 정부 측면에서도 회사 앱의 개인정보보호 가이드라인을 구축할 필요성이 있다.

다섯 번째로는 기술 유출 우려에 관한 문제점을 도출할 수 있었다. ‘사내 보안’은 삼성전자의 삼성페이에 사용된 핵심 기술인 카드정보 마그네틱 전송 기능을 보유한 자회사인 루프페이의 해킹에 따른 이슈로 감성 분석 점수는 1.5점으로 측정되었다. 삼성페이는 8월 20일부터 약 348억 원이 결제되었을 만큼 많이 사용되고 있는 서비스로 이에 대한 사내망 해킹 이슈가 일반 사용자들이 많이 언급하게 된 것으로 보인다. 산업통상자원부와 중소기업청에 따르면, 기술 유출을 방지하고 국가 핵심기술을 보호하기 위하여

2015년 12월, ‘산업기술 유출방지 및 보호에 관한 중합계획’과 ‘중소기업 기술 보호 지원계획’을 확정하였으며, 또한 기술 유출범죄에 대한 형량 강화 등을 추진할 계획이라고 밝혔다. 따라서 정부에서는 이와 같은 계획에 맞추어 기술유출 사전예방을 위한 지원과 피해 기업에 대한 사후 구제, 그리고 내실화를 지속적으로 추진해 나가야 할 것이다. 정보보호인식 강화를 위한 세부 방안 내용은 〈표 13〉에서 확인할 수 있다.

## V. 결론

### 1. 연구 결과 토의 및 시사점

본 연구는 트위터 데이터를 이용해 정보보호인식을 분석하기 위하여 전문가 인터뷰를 바탕으로 정보 보호 관련 11개의 키워드를 선정하여 연구를 진행하였다. 한국 트위터 데이터 299,871건을 포함하여 미국, 중국의 트위터 데이터 총 1,882,704건에 대해 정보 보호 관련 이슈들을 분석하였다. 토픽 모델링을 통해 10개의 주제를 추출하여 상대적인 분석을 실시하였고, 국가별, 이슈별, 항목별 점수 결과를 제시하였다.

분석결과, 한국의 주요 토픽이 기술적 보안 영역에서 도출됨을 확인하였으며, 파밍과 피싱 사기, 모바일 앱 보안, 루프 페이지 사내망 해킹 사건에 따른 사내 보안 이슈, 병원정보 유출 이슈 등이 1.5점 전후로 측정되는 것을 확인하였다. 이에 사회 공학의 대표적인 사기 방법인 파밍과 피싱, 그리고 모바일 시대의 이슈, 정보 유출 이슈가 중요하게 논의되고 있음을 확인하였다. 이와 같이 기존의 설문 분석이나 현업에서 많이 이용하고 있는 보안 관련 기사를 스크랩하는 방법이 아닌, 비정형 데이터의 분석을 통하여 정보보호 인식과 주요 이슈를 파악할 수 있음을 확인할 수 있었다.

본 연구는 비정형 데이터인 트위터의 일반 사용자가 작성한 트윗을 텍스트 마이닝 기법인 토픽 모델링과 감성 분석을 통해 분석하였다. 이에 본 연구의 학술적 시사점은 다음과 같다. 첫째, 기존에 주로 행해진 설문 분석을 통한 연구 방법이 아닌(김영곤, 2010; 백민정 외, 2010; 손승희, 2013; 이은곤 외, 2012; Bulgurcu, et al., 2010; Humaidi, et al., 2015), 비정형 데이터를 분석하는 데이터 분석 기법인 토픽 모델링과 감성 분석을 적용하여 정보보호인식과 주요 이슈를 확인하여 학문의 다양성 확보에 기여했다는 의의가 있다. 특히 감성 분석의 적용을 통하여, 단순한 이슈 확인에 그친 것이 아니라 점수를 통해 이슈에 대한 사용자들의 긍/부정을 분석하였다.

둘째, 기존의 정보보호인식 관련 연구들이 정보보호 인식과 정보보호행위에 영향을 미치는 요인과 상관 관계에 대한 연구들에 국한된 것(김영곤, 2010;

백민정 외, 2010; 손승희, 2013; Bulgurcu, et al., 2010; Humaidi, et al., 2015)과는 달리 일반 사용자들의 정보보호인식 자체에 대해 연구를 수행하였으며, 실제 데이터를 바탕으로 한 이슈 분석과 대안을 제시하였다는 의의가 있다.

셋째, 본 연구는 도출된 정보보호 이슈를 기반으로 국가 간 비교를 수행하였으며, 이를 바탕으로 한국의 정보보호인식 강화방안을 탐색하였다. 따라서 향후 정보보호인식에 대한 추가적인 국가 간 비교 연구와 시계열적 분석에 적용할 수 있을 것이다.

본 연구는 정보보호인식을 분석하기 위하여 트위터 API를 통한 실제 데이터를 활용한 연구로, 기존의 설문이나 기사 스크랩 등의 방법이 아닌 온라인에서 수집 가능한 실제 데이터를 활용하였다는 점에서 실무적 의의가 있다. 본 연구의 연구 모델을 활용하여 국내 이슈 및 해외 시장 분석에 활용할 수 있을 것으로 기대한다. 본 연구의 실무적 시사점을 아래와 같이 제시할 수 있다.

먼저 정보보호분야가 아닌, 다른 산업분야에서도 적용할 수 있는 방법을 제시하고 있다는 점이다. 본 연구의 방법은 정보보호인식을 분석하기 위하여 사용하였지만, 일반산업에서의 비정형 데이터를 통한 산업의 전반적인 인식, 주요 이슈 분석에 사용할 수 있다는 의의가 있다.

또한, 정보보호인식에 대한 이슈를 도출하고, 이에 대한 정부, 기업, 학계 측면에서의 정보보호인식 강화 방안을 제시했다는 점에서 실무적 의의를 지닌다. 본 연구의 방법을 통하여 실무에서의 이슈 분석 및 방안 도출뿐 아니라, 각계 측면에서의 종합적인 분석이 가능할 것이다.

## 2. 연구의 한계점 및 향후 연구 방향

본 연구는 텍스트 마이닝 기법을 활용하여 정보보호 인식과 관련 이슈를 확인하고 이를 감성 분석을 통해 점수화하여 제시하였다. 하지만 다음과 같은 한계

점을 지니고 있으며, 이는 향후 연구에서 보완되어야 할 것이다.

첫째, 본 연구에서 사용된 데이터의 한계점이다. 중국의 정보보호인식을 분석하고자 하였으나, 데이터 수집의 어려움으로 인하여 기존에 중국에서 가장 많이 사용되고 있는 웨이보가 아닌 트위터 데이터를 사용하였으며 데이터의 볼륨 자체는 충분히 확보하였으나, 해당 데이터들이 대표성을 갖지 못할 수 있다는 문제점이 있다. 또한, 중국어 데이터의 직접 분석이 불가하여 구글 번역기를 통해 영어로 번역하여 분석하였다. 이는 문맥상의 차이가 발생할 수 있다. 본 연구에서 사용한 기법인 토픽 모델링과 감성 분석 기법의 세부 분석 단위가 단어이기 때문에 해당 차이는 최소화될 수 있을 것이나, 중국 데이터의 직접 분석을 통한 비교 연구를 수행한다면 더욱 정확한 반응을 확인할 수 있을 것이다.

둘째, 국가별 점수의 직접 비교가 어렵다는 문제점이 있다. 본 연구의 결과로 도출된 감성 분석에 따른 점수값은 사용자의 해당 의견에 대한 긍/부정 비율을 계산하여 도출된 값이며, 그에 따라 상대적으로 불만을 위주로 의견이 표출되고 있는 한국과 전체적으로 가장 높은 점수가 측정된 중국의 경우 국가별 정보보호 상황과 사용자의 차이에 따른 보정이 필요하다고 생각된다. 이를 보정하기 위한 추가적인 연구를 진행하거나 머신러닝 등의 방법을 활용한 연구가 가능할 것이다.

셋째, 연구에 시간에 대한 분석이 이루어지지 않은 점이다. 본 연구는 32일간의 데이터를 수집하여 분석하였으나, 추후 시계열적으로 분석하고, 트윗이 올라온 시간에 따른 분석을 수행한다면 더욱 발전된 연구 결과를 제시할 수 있을 것이라 기대된다.

#### ■ 참고문헌

김영곤 (2010). “항공 응용 분야: 정보보안 거버넌스의 구성요소가 종업원의 보안 인식과 행위에 미치는 영향

에 관한 연구.” 「한국항행학회논문지」, 14(6): 935-950.

김유신 · 김남규 · 정승렬 (2012). “뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자의사결정모형.” 「지능정보연구」, 18(2): 143-156.

류우중 · 하종우 · Alam, M.H., · 이상근 (2013). “토픽 모델링 기법을 이용한 트위터 트렌드 추출.” 「한국정보과학회 학술발표논문집」, 191-193.

머니투데이 (2015). “통념의 틀을 벗어나라, 만물이 소통한다.” <http://www.mt.co.kr/view/mtview.php?type=1&no=2014122314007122101&outlink=1>. (검색일: 2015.01.16).

박자현 · 송민 (2013). “토픽 모델링을 활용한 국내 문헌 정보학 연구 동향 분석.” 「정보관리학회지」, 30(1): 7-32.

박종도 (2015). “커뮤니티 기반 Q&A서비스에서의 질의 할당을 위한 이용자의 관심 토픽 분석에 관한 연구.” 「정보관리학회지」, 32(3): 397-412.

배정환 · 손지은 · 송민 (2013). “텍스트 마이닝을 이용한 2012년 한국대선 관련 트위터 분석.” 「지능정보연구」, 19(3): 141-156.

백민정 · 손승희 (2010). “조직의 정보윤리실천이 구성원의 정보보안 인식과 행동에 미치는 영향에 관한 연구.” 「경상논총」, 28(4): 119-145.

손승희 (2013). “스마트워크 근무환경 특성이 개인의 정보보안인식 행동에 미치는 영향에 관한 연구: 자율성과 이동성 그리고 책임을 중심으로.” 「경상논총」, 31(4): 17-39.

아이뉴스 (2015). “보안업계 상반기 실적 보니... 열매는 거의 없어.” [http://news.inews24.com/php/news\\_view.php?g\\_serial=915323&g\\_menu=020200&rrf=nv](http://news.inews24.com/php/news_view.php?g_serial=915323&g_menu=020200&rrf=nv). (검색일: 2015.08.23).

아이티투데이 (2015). “유독 잠잠했던 대형 보안사고 막았나 숨겨졌나?” <http://www.ittoday.co.kr/news/articleView.html?idxno=66598>. (검색일: 2015.12.04).

안정국 · 김희웅 (2015). “집단지성을 이용한 한글 감성어 사전 구축.” 「지능정보연구」 21(2): 49-67.

유홍렬 · 홍모세 · 권태경 (2013). “피싱 및 파밍 공격에 의한 다수의 패스워드 유출 요인에 관한 연구.” 「정보보호학회논문지」, 23(6): 1225-1229.

이윤주 · 서지훈 · 최진탁 (2014). “SNS 텍스트 콘텐츠를 활용한 오피니언마이닝 기반의 패션 트렌드 마케팅 예측 분석.” 「한국정보기술학회논문지」, 12(12): 163-

- 170.
- 이은곤 · 최지은 · 이호근 (2012). “계좌 이용 과정과 결과의 투명성이 온라인 뱅킹 이용자의 보안 인식에 미치는 영향.” 『한국경영정보학회』, 14(3): 75-97.
- 임채호 (2006). “효과적인 정보보호인식제고 방안.” 『정보보호학회지』, 16(2): 30-36.
- 전은정 · 김학범 · 염홍열 (2012). “일본 · 중국의 개인정보보호 법 · 제도 동향.” 『정보보호학회지』, 22(5): 90-105.
- 조승연 · 최지은 · 이규현 · 김희웅 (2015). “고객 온라인 구매후기를 활용한 추천시스템 개발 및 적용.” 『Information Systems Review』, 17(3): 157-164.
- 진설아 · 허고은 · 정유경 · 송민 (2013). “트위터 데이터를 이용한 네트워크 기반 토픽 변화 추적 연구.” 『정보관리학회지』, 30(1): 285-302.
- 차운정 · 이지혜 · 최지은 · 김희웅 (2015). “소셜미디어 토픽모델링을 통한 스마트폰 마케팅 전략 수립 지원.” 『지식경영연구지』, 16(4): 69-87.
- 최판암 · 이민형 (2014). “지역기관의 산업보안 활동이 기술유출방지에 미치는 영향.” 『한국지방 자치연구』, 16(1): 119-141.
- Baccianella, S., Esuli, A. & Sebastiani, F. (2010). “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining.” *LREC* No.10: 2200-2204.
- Bauer, S. & Chudzikowski, K. (2015). “Mind the Threat! A Qualitative Case Study on Information Security Awareness Programs in European Banks.” *Americas Conference on Information Systems*.
- Blei, D., Ng, A. & Jordan, M. (2003). “Latent dirichlet allocation.” *Journal of machine Learning research*, 3: 993-1022.
- Bulgurcu, B., Cavusoglu, H. & 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., Son, JY. & 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 (2012). “Twitter Grows Stronger in Mexico.” <http://www.emarketer.com/Article.aspx?R=1009370&ecid=a6506033675d47f881651943c21c5ed4>. (Retrieved on September 24, 2012).
- Eminağaoğlu, M., Uçar, E. & Eren, Ş. (2009). “The positive outcomes of information security awareness training in companies - A case study.” *information security technical report*, 14(4): 223-229.
- Global Web Index (2014). “90 million VPN users in China have accessed restricted social networks.” <http://www.globalwebindex.net/blog/vpn-in-china>. (Retrieved on September 24, 2015).
- Griffiths, T. L. & Steyvers, M. (2004). “Finding scientific topics. Proceedings of the National Academy of Sciences.” 101(suppl 1): 5228-5235.
- Hillard, D., Purpura, S. & Wilkerson, J. (2008). “Computer-assisted topic classification for mixed-methods social science research.” *Journal of Information Technology & Politics*, 4(4): 31-46.
- Hong, L. & Davison, B. D. (2010). “Empirical study of Topic Modeling in twitter. In Proceedings of the first workshop on social media analytics.” *ACM*, 80-88.
- Hridoy, S.A.A., Ekram, M.T., Islam, M.S., Ahmed, F. & Rahman, R.M. (2015). “Localized twitter opinion mining using sentiment analysis.” *Decision Analytics*, 2(1): 1-19.
- Humaidi, N. & 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.
- Kim, S.Y., Song, S.J. & 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.
- Mann, G.S., David, M. & Andrew, MC. (2006).

- “Bibliometric impact measures leveraging topic analysis.” *Digital Libraries, JCDL'06. Proceedings of the 6th ACM/IEEE-CS Joint Conference on, IEEE.*, 65–74.
- Siponen, M.T. (2000). “A conceptual foundation for organizational information security awareness.” *Information Management & Computer Security*, 8(1): 31–41.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K. & Stede, M. (2011). “Lexicon-based methods for sentiment analysis.” *Computational linguistics*, 37(2): 267–307.
- Tumasjan, A., Sprenger, T.O., Sandner, P.G. & Welpe, I.M. (2010). “Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment.” *ICWSM*, 10: 178–185.
- Wang, X., Zhu, F., Jiang, J. & Li, S. (2013). “Real time event detection in twitter.” *In Web-Age Information Management*, 502–513