Journal of The Korea Knowledge Information Technology Society

ISSN 1975-7700

http://www.kkits.or.kr

Analysis of Digital Forensics Technology Trends Based on Big Data

Dong-Suk Hong[1], Sang-Duk Jeon[2], Chan-Ho Kim[1], Han-Gook Kim[1*]

[1]*Department of Technology Commercialization Analysis, KISTI*
[2]*Department of Digital Forensics, Kim & Chang*

ABSTRACT

Digital forensics service market in the US is estimated $976 million in 2012. The growing number of mobile connections and households with at least one computer will continue to propel demand for digital forensics. Otherwise, there are many challenges as the IT world is continually evolving. It is important for the forensics investigator, the forensics toolkits developer and other related experts to monitor and understand technologies trends related to new devices, systems, components, and even future technology trend. Therefore, this paper analyzes emerging issue and trend related to digital forensics technology from papers and news media.

KEYWORDS: Digital forensics, Big data analysis, Emerging issue, IT technologies, News text analysis

ARTICLE INFO: Received 23 January 2014, Revised 10 February 2014, Accepted 14 February 2014.

1. 서 론

디지털 포렌식은 법정 제출용 디지털 증거를 수집하여 분석하는 기술을 뜻하며, 사이버 공간에서의 컴퓨터 범죄를 수사하는 과정을 뜻한다. 최근 컴퓨터 기술이 더욱 복잡해지고 새로운 장비 및 시스템들이 빠르게 진화해감에 따라 이러한 포렌식 기술의 진화는 포렌식 제품과 서비스의 주요한 수요 요인으로 작용해 왔으며, 디지털 포렌식 관련 전문가들 역시 최신 기술과 제품들에 신속히 대응하도록 요구되어지고 있다[1, 2]. 따라서, 본 논문에서는 디지털 포렌식과 관련하여 최근 기술 트렌드를 분석하며 이를 위해 논문과 뉴스 미디어를 활용하였다.

*Corresponding author is with the Department of Technology Commercialization Analysis, KISTI, 66 Hoegi-ro Dongdaemun-gu Seoul, KOREA.
E-mail addresses: hgkim712@kisti.re.kr

2. 빅데이터 분석 기술 연구 동향

2.1 빅데이터 분석 및 관련 기술

위키피디아에 의하면 기존의 데이터베이스 관리 툴과 데이터 처리 어플리케이션으로는 데이터의 수집, 큐레이션, 저장, 검색, 공유, 전송, 분석, 가시화 등에 어려움이 있는 상당히 크고 복잡한 데이터의 집합을 빅데이터*로 정의하고 있다.

Gartner**에 의하면 빅데이터는 volume(데이터 크기), variety(데이터의 다양성), velocity(데이터 전달 속도) 등의 측면에서 특성을 지닌다. 즉, 빅데이터는 데이터의 크기가 급증하고, 데이터 전달 속도가 빠르며 데이터 구조가 다양한 현상을 관찰하여 이러한 현상으로 인해 새로운 도전이 등장할 것으로 전망되고 있다.

더불어 빅데이터는 새로운 가치에 대한 기회를 발견할 수 있는 원천이다. 기업, 공공, 정책, 보건/의료, 금융 등 다양한 분야에서 새로운 가치의 원천이자 지능화 서비스의 핵심 인프라로써 빅데이터의 분석과 활용에 관한 연구가 진행되고 있다[3,4,5].

국내에서 빅데이터의 활용은 현재 과학기술 R&D 부문에 국한되어 왔으나 멀지 않은 미래에 공공 및 산업, 서비스 전분야로 확산되고 새로운 데이터의 생태계가 조성될 것으로 예상되고 있다.

대외적 환경변화와 사회적 현안, 미래의 잠재적 위험성과 기회 등을 신속히 감지하고 정확하게 분석함으로써 현재의 현상 분석뿐만 아니라 미래에 일어날 수 있는 일의 방향성과 트렌드를 예측하기 위하여 빅데이터가 활용될 수 있다. 기업의 경우, 빅데이터로부터 유용한 정보를 찾고 잠재된 정보를 활용하는 기업이 경쟁력을 가지고 시장을 선도할 것이다. 기업은 빅데이타를 활용함으로써 R&D기술 기획 및 마케팅의 효율성 및 효과성을 제고할 수 있을 것이다. 웹 정보, 소셜미디어 정보를 활용하여 급부상하는 이슈를 발굴하고 트렌드를 분석 및 예측할 수 있으며, 이는 기술기획 및 마케팅에 활용될 수 있다. 다양한 측면에서의 빅데이터 분석 및 활용의 목적 및 용도에 대하여 다음 <표 1>에서 정리하였다. 본 논문에서 분석하고 있는 디지털 포렌식의 경우 기업과 공공(보안)분야에서 각각 기술기획 및 국가 위험관리의 효율성을 제고하기 위한 목적으로 빅데이터를 분석 및 활용하는 것과 관련된 연구이다.

표 1. 주요 분야별 빅데이터 분석

Table 1. The goal on big data analysis by major parts

구분	목적 및 용도
기업	-웹 정보, 소셜미디어 정보를 활용하여 미래사회를 전망하고 기술기획에 반영 -구글검색엔진과 웹데이터를 활용한 급부상이슈 발굴 및 트렌드변화 분석 및 예측 -R&D 기획 및 마케팅의 효율성 및 효과성 제고
공공	-국가 위험관리 및 예측 -다양한 새로운 지적 통찰력을 공공의 목적에 활용
정책	-공공기관의 빅데이터 공유/활용 정책 수립 -기술위험 및 투자손실 요소를 사전에 파악하고 통합적 정책 방향 및 연구개발 전략을 수립
보건/의료	-보건의료 R&D, 임상에서의 활용 등에서 의사결정 지원, 원격진료 수준의 향상, 비용 절감
금융	-자금세탁 추적, 고객 해약방지 등의 리스크 관리 및 소비자 니즈 파악을 통한 고객별 맞춤형 마케팅 목적에 활용

* Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, curation, storage, search, sharing, transfer, analysis, and visualization.

** http://www.gartner.com

다양한 분야에서 빅데이터 분석을 위한 기반 기술로써 관련 주요 기술들은 다음과 같다.

표 2. 빅데이터 분석 관련 기술 분류와 예시
Table 2. Technology classification and examples related on big data analysis

분류	예 시
원본 데이터 저장	대용량 분산 파일 시스템 (Hadoop File System, MogileFS 등)
구조적 데이터 저장	NoSQL(Cassandra, Mongo DB, HBase, Cloudata 등)
배치 분산 병렬 처리	MapReduce(Hadoop), 그래프 분석(Pregel, GoldenORB) 등
데이터스트리밍 프로세싱	S4, Storm 등
데이터 마이닝 및 통계	Mahout, R 등
클러스터링 및 모니터링	ZooKeeper, HUE, Cloumon 등
분산처리	분산관리(ZooKeeper), 분산 큐(Kafka), 분산 캐쉬(Memcached, Redis) 등
기존 솔루션	BI/DW, RDBMS, Streaming DBMS 등
시각화 솔루션	가시화 툴(cognos, D3.js, datawrapper 등)

3V(Volume, Variety, Velocity) 특성 즉, 방대한 데이터 크기, 빠른 데이터 전달 속도, 다양한 데이터 구조 특성으로 빅데이터를 정의했던 Gartner는 다음과 같이 빅데이터 분야의 부상 기술을 소개하고 있다.

<그림 1>에서 나타난 것과 같이 향후 2년에서 5년 사이에 부상할 관련 기술로 Hadoop SQL interface, Document Store Database Management Systems, Context-Enriched Services, Key-value Database Management Systems, Cloud-based Grid Computing, In-Memory Database Management Systems 등이 있다.

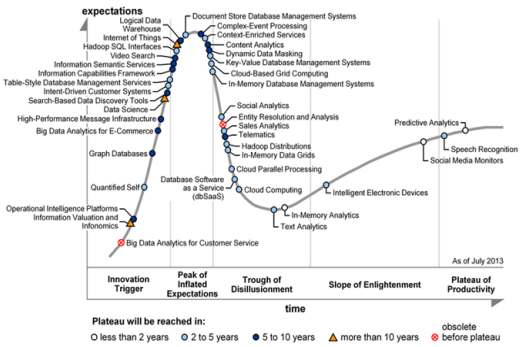

그림 1. 빅데이터 분야의 기술 하이프 싸이클
Figure 1. Hype cycle for big data

2.2 신문 텍스트 분석

주요 미디어 중 하나인 신문의 텍스트는 특허 및 논문 정보와는 다른 특성을 지니고 있다. 신문 텍스트는 특허 및 논문에 비해서는 전문성이 낮으며, 기사화될 수 있는 내용을 선택적으로 보도하는 경향이 있다. 또한 광고를 위한 내용 혹은 중복되어 기재된 정보 등에 대한 정제가 필요할 수 있다. 아래 <표 3>에서는 논문 및 특허, 웹정보, 소셜 미디어, 신문기사 등 정보 출처 및 유형별 빅데이터의 특성과 활용성에 관하여 정리하였다. 본 논문에서는 디지털 포렌식 기술 동향을 분석하기 위하여 논문, 웹정보, 신문기사를 분석, 활용하였다.

표 3. 정보 출처 및 유형별 빅데이터 특성
Table 3. Big data properties by source type

구분	특징 및 활용성
논문 및 특허	-급부상기술영역 탐색, 기술간 연관관계 분석, 기술수명주기 측정, 핵심기술 도출, 공동연구망 분석 등을 위해 분석 -인터넷 정보의 생성과 축적에 비하면 크기와 속도가 낮은 수준 -논문 및 특허데이터의 양이 많아짐에 따라 데이터를 추출, 정제하는 작업이 어려워지고 있으며 데

	이터베이스 복잡성도 증가되고 있음
웹정보	-구글검색엔진과 웹데이터 등을 분석함으로써 미래사회를 전망하고 기술의 발전방향을 예측 -트렌드 변화의 심층분석과 새로운 이슈의 조기 발굴을 위해 활용
소셜 미디어	-페이스북, 트위터 등의 소셜미디어 기반 데이터는 기존의 트랜잭션 데이터보다 상당히 방대하며 비구조적이고 복잡한 구조를 띠고 있음 -제품가격/판매량 예측, 영화 박스오피스 예측, 선거 예측, 정보 확산도 예측, 거시경제 예측 등 다양한 분야에서 주요 이슈를 탐지하고 모니터링하며 미래를 예측하기 위해 활용
신문 기사	-신속하게 정보를 전달하는 속보성이 우수함 -주목받는 주요한 기술과 제품의 혁신활동 현황을 가늠하는데 활용 -기업 활동 및 연구개발 전과정의 다양한 정보 제공 -고객과 대중에 대한 PR 혹은 전달력에 관한 평가, 이슈 모니터링, 트렌드 발굴 등을 통한 전략적 인사이트 제공[6,7]

<그림 2>에서 보는 바와 같이, 미디어 컨텐츠의 분석을 통해 고객 측면, 대중 측면에서 미디어의 효과를 평가할 수 있고, 미디어 소스 측면, 이슈 측면에서 전략적 인사이트를 획득할 수 있다. 미디어 컨텐츠로부터 전략적 인사이트를 획득하기 위한 구체적인 방법으로 이슈 트래킹, 트렌드 발굴, 경쟁자 분석 등이 있다[7]. 본 논문에서는 디지털 포렌식 분야에서의 전략적 인사이트를 얻기 위한 목적으로 신문 미디어를 분석하였다.

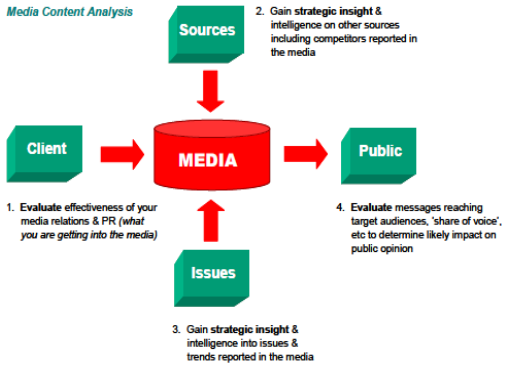

그림 2. 미디어 컨텐츠 분석의 역할과 활용

Figure 2. The four roles and usage of media content analysis

미디어 컨텐츠 분석에 관한 연구 이외에도 신문기사 분석과 관련된 다수의 선행 연구가 존재한다. <표 4>에서 주요 연구를 요약하였다.

표 4. 신문기사 분석 관련 기존 연구

Table 4. The previous study on news articles analysis

구분	내 용
Jim Macnamara (2003)[7]	-신문기사 등 미디어 텍스트의 컨텐츠 분석 방법론 연구
Koppel and Shtrimberg (2006)[8]	-재무적인 시장 성능을 예측하기 위한 목적으로 신문기사를 분석 연구
Kevin Lerman, Ari Gilder, Mark Dredze and Fernando Perira (2008)[9]	-정치적 후보자들에 대한 공공의 의견을 예측하기 위한 목적으로 신문기사를 분석 연구
Tristan Snowsill1, Ilias Flaounas, Tijl De Bie, and Nello Cristianini(2010)[10]	-대용량 신문기사로부터 주요 이벤트를 탐지하는 기술 연구
Omar Ali, Ilias Flaounas, Tijl De Bie, Nick Mosdell, Justin Lewis and Nello Cristianini (2010)[11]	-대용량 신문기사로부터 패턴과 트렌드를 추출하기 위한 텍스트 분석 기술 연구

표 5. 신문기사 분석 관련 주요 이슈
Table 5. News articles analysis related major issues

구분	주요 기술 이슈
대용량 데이터	- 분산 저장 - 텍스트 분석 성능 - 배치 프로세싱 - 클러스터링
비정형 데이터	- 메타데이터 검색 - 중복 및 노이즈 제거 - 개체 추출 정확도 제고 - 구조적 데이터 저장
전달속도	- 리얼타임 프로세싱 - 스트리밍 프로세싱
그 외	- 이슈/인사이트 도출 알고리즘 및 가시화 - 패턴 추출 등

신문 기사는 빅데이터의 특성을 가지며 데이터 크기가 거대하고 데이터 전달 속도가 빠르며 데이터 구조가 다양하므로, 이러한 신문기사 텍스트를 분석하기 위해 해결해야 할 기술적 이슈가 존재한다. 신문기사는 대용량 데이터로 효율적인 분산저장, 텍스트 분석의 성능향상, 배치 프로세싱 등의 기술이 요구된다. 또한 신문기사는 비정형 데이터로써 다양한 데이터 포맷에 대한 메타데이터 검색, 중복 정보 및 노이즈를 제거하기 위한 기술, 개체 추출 정확도 제고를 위한 기술 등이 요구된다. 또한 리얼타임 프로세싱 및 스트리밍 프로세싱 등 빠른 전달속도에 대응하는 기술이 요구된다. 그 밖에 신문기사 분석을 통해 새로운 가치(이슈 및 인사이트)를 도출하기 위한 다양한 알고리즘 및 기술이 요구되며 이러한 가치를 가장 효율적이며 직관적으로 표현할 수 있는 가시화 기술 등도 요구된다.

3. 디지털 포렌식 기술

디지털 포렌식은 증거의 수집, 보존, 분석, 문서화, 그리고 재판 과정에 증거로 제출하기까지의 모든 과정을 포함한다. 디지털 기술의 발달로 증거는 네트워크, 인터넷, 데이터베이스, 모바일 기기, 휘발성 메모리 등 다양한 곳에서 존재하기 때문에 전문성이 더욱 심화되고 대형 시스템의 하드웨어 종류나 운영체제 종류에 따라 다양한 방법론에 대한 연구가 이루어지고 있다[2].

디지털 포렌식은 일반적으로 증거의 압수, 장비의 준비, 포렌식 이미징, 포렌식 조사 및 분석, 문서화, 분석보고서 및 검토의 절차로 이루어진다[12]. 아래 <표 6>에서 이러한 절차의 각 단계에 대한 주요 내용을 기재하였다.

표 6. 포렌식 절차 및 방법론
Table 6. Forensic Procedure and Methods

단계	세부 내용
증거의 압수	-범죄현상에 필요한 장비를 결정할 때에는 수사관과 협의하고 -증거를 압수하기 위한 법적 권한을 검토하며 -범죄현장에서 증거를 가져오는 것이 비현실적인 경우에는 법적 절차에 따라 증거를 복제 또는 이미징하여 획득하고 -잠재적인 용의자, 목격자, LAN 관리자 등에게 압수한 시스템에 대한 정보를 요구하고 -범죄현장은 증거 수집을 위해 체계적으로 그리고 철저히 수색되어야 한다.
장비의 준비	-적절한 증거처리가 유지될 수 있도록 장비가 모니터링되고 문서화되어야 한다. -적당한 운영 장비를 배치하고 -각 장비에 대한 제조자의 운영 매뉴얼이나 관련문서가 접근 가능해야 하며, 분석 또는 이미징 소프트웨어는 사용에 앞서 검증된 것이어야 한다.
포렌식 이미징	-증거의 현재 상태를 문서화하고 증거가 위험원으로부터 오염되지 않도록 노출에 주의해야 한다. -원본증거가 변경되는 것을 방지하기 위해 쓰기방지 기능이 있는 HW나 SW를 사용해야 하며

	-증거획득방법이 포렌식 절차에 적합하고 증명 가능한 것이어야 하며 -원본매체의 비트스트림 이미지를 획득할 수 있는 HW나 SW를 사용하며 -증거분석에 제출된 디지털증거는 데이터 무결성이 유지되어야 한다.
포렌식 조사 및 분석	-분석자는 분석에 필요한 프로세스를 알아내기 위해 의뢰인이 제출한 문서를 검토하고 요청된 분석을 수행하기 위한 법적 권한이 있는지 확인해야 한다. -분석대상이 목표로 하는 자료를 제공하는데 최상의 선택인지 여부를 고려해야 한다. -분석전략은 의뢰인과 분석자가 합의하고 문서화해야 한다. -매체에 대한 분석은 당국의 관리 운용절차에 따라 수행되어야 한다.
문서화	-법적 권한을 증명하는 사본, 연계보관성에 대한 사항, 분석할 증거의 초기 수치, 포장이나 증거의 상태에 대한 정보, 기타 증거에 대한 기술과 사건에 관한 의견 등에 관하여 문서화한다.
분석보고서	-분석자 소속 기관의 요구조건을 충족시켜야 되며 분석자에 의해 작성된 보고서는 의뢰인의 요구사항을 다루고 있어야 한다.
검토	-분석자 소속 기관은 동료에 의한 기술적 검토(peer review) 및 관리상의 검토에 대한 규칙을 제정하고 서면으로 작성된 정책을 가지고 있어야 한다.

또한 디지털 포렌식은 분석 대상에 따라 다음과 같이 몇 가지 포렌식 유형으로 분류된다. 디스크 포렌식, 휘발성데이터 포렌식, 시스템 포렌식, 네트워크 포렌식, 모바일 포렌식, 데이터베이스 포렌식에 대한 간략한 소개와 관련 주요 기술은 다음 <표 7>과 같다.

표 7. 분석 대상 별 디지털 포렌식 기술 분류
Table 7. Digital forensics technology classification by object type to analyze

분류	내용 및 주요 기술
디스크 포렌식	물리적인 저장장치인 하드디스크, 플로피디스크, CD ROM, DVD 등 각종 보조 기억장치에서 증거를 수집하고 분석하는 포렌식
	주요 기술: disk encryption, disk imaging, hash, data recovery 등
휘발성데이터 포렌식	레지스터, 캐시, 메모리, 네트워크 연결 상태, 실행중인 프로그램 상태 등 휘발성 데이터 상의 증거를 수집하고 분석하는 포렌식
	주요 기술: 메모리 상의 숨겨진 데이터 추출, 변경 데이터 추적, 휘발성 데이터 무결성 검증 등
시스템 포렌식	컴퓨터의 운영체제, 응용 프로그램 및 프로세스를 분석하여 증거를 확보하는 포렌식
	주요 기술: file system 분석, process 분석 등
네트워크 포렌식	네트워크 정보와 전송 데이터를 수집하여 필요한 증거를 추출하고 분석하는 포렌식
	주요 기술: 데이터 트래픽 분석, 로그 정보 분석, 웹히스토리 분석, IP 추적 등
모바일 포렌식	휴대폰, PDA, 전자수첩, 디지털 카메라, MP3 플레이어, 캠코더, 휴대용 메모리 카드, USB 등 휴대용 기기에서 필요한 정보를 입수하여 분석하는 포렌식
	주요 기술: 스마트폰 등 모바일 기기의 데이터 수집, 추출, 분석 기술 등
데이터베이스 포렌식	데이터베이스로부터 데이터를 추출, 분석하여 증거를 획득하는 포렌식
	주요 기술: 데이터베이스 복구, 분석, 데이터베이스시스템 제어 등
그 외	회계 포렌식, 암호 포렌식 등

4. 디지털 포렌식 기술 동향 분석

4.1 분석 데이터 및 방법론

본 논문에서는 학술적 측면에서의 기술동향을 살펴보기 위해 논문 분석을 수행하고, 웹, 미디어 등 일반적인 측면에서의 기술동향을 살펴보기 위해 웹검색 및 신문기사 분석을 수행하였다. 디지털 포렌식 논문에서 최근 부상하는 이슈를 살펴보기 위하여 SCOPUS* 데이터베이스에서 2000년부터 2012년까지의 'digital forensics'을 포함하는 국내외 논문(제목, 초록, 키워드, 날짜, 국가) 총 1,807건을 다운받아 분석하였다. 또한 신문 기사는 KINDS에서 제공하는 온라인 기사정보와 개별 신문사의 홈페이지에서 제공되는 기사정보를 활용하여 분석하였다. 분석 대상 신문기사는 경향신문, 국민일보, 동아일보, 문화일보, 전자신문, 내일신문, 서울신문, 세계일보, 한겨레, 디지털타임스, 주간조선, 한국일보, 중앙일보, 디지털데일리, 매경이코노미 등 총 40종류 신문사에서 제공된 것으로 2012년부터 2013년까지의 '디지털 포렌식'을 포함하는 최근 국내 신문(제목, 내용, 날짜, 출처) 총 182건을 분석하였다. 웹검색 기반 디지털 포렌식 트렌드 분석을 위해서는 Google Trend 사이트에서 'digital forensics' 검색어로 검색된 결과를 분석, 재구성하였다.

특히, 뉴스 텍스트기반 트렌드 분석 절차 및 방법은 다음과 같다.

1. 원문 데이터 다운로드 및 노이즈 제거
2. 형태소 분석
3. T-score 분석
4. 부상 키워드 도출
5. 가시화

'디지털 포렌식'이 신문의 제목 또는 본문내용에 포함된 기사의 원문 데이터를 다운받아 중복 및 노이즈를 제거하고 형태소를 분석, 추출한 후 '디지털 포렌식' 용어와 동시에 등장한 횟수가 많은 주요 키워드를 부상 키워드로 도출하였다. 가시화 프로그램**을 활용하여 디지털 포렌식, 네트워크 포렌식을 중심으로 부상 키워드들의 네트워크를 분석하였다.

4.2 분석 결과

4.2.1 웹검색 기반 트렌드 분석

전 세계 Google 웹 사용자의 검색 로그를 기반으로 제공되는 동향분석 서비스인 Google Trends를 활용하여 Digital forensics의 웹 검색 동향을 살펴본다. <그림 3>에 따르면 2005년 이후부터 2013년 현재까지 'Digital Forensics' 웹검색 빈도는 전반적으로 증가해오고 있는 것을 알 수 있다.

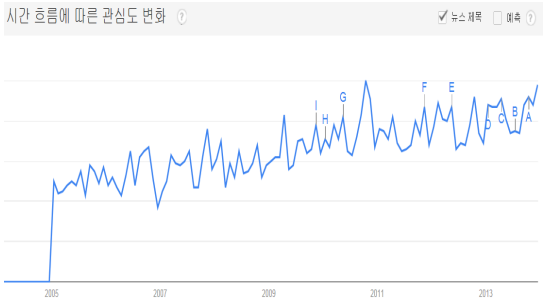

그림 3. 'Digital Forensics'의 웹검색 빈도 시계열변화
Figure 3. Timeline of web search on digital forensics

또한, 'digital forensics'에 대한 지역적인 관심도 분포를 살펴보면 미국, 인도, 영국, 오스트레일리아 순이며, 'digital forensics'과 연관된 검색어로써 최상위 검색어는 'digital computer forensics', 'computer forensics', 'digital forensics jobs', 'digital forensics tools' 등이다. 또

* http://www.scopus.com

** VOSViewer, (http://www.vosviewer.com)

한, 급상승 연관 검색어는 'cyber forensics', 'digital evidence', 'digital forensics analysis' 등이다.

표 8. 'Digital Forensics'의 연관검색어

Table 8. related search keywords with digital forensics

구분	검색어	
최상위 연관검색어	digital computer forensics	100
	computer forensics	100
	digital forensic	85
	digital forensics jobs	35
	digital forensics tools	30
	digital evidence	25
	what is forensics	25
급상승 연관검색어	cyber forensics	급등
	digital evidence	급등
	digital forensic	급등
	digital forensics analysis	급등
	digital forensics degree	급등
	digital forensics jobs	급등
	digital forensics pdf	급등

4.2.2 논문기반 트렌드 분석

디지털 포렌식과 관련된 논문은 2000년대 중반 이후에 급속하게 증가한 것을 알 수 있다. <그림 4>에서 'digital forensics'을 포함하는 논문의 발행 수 현황을 나타내었다. 2006년에 100건을 초과한 이후 2012년 약 341건의 논문이 발행되었다.

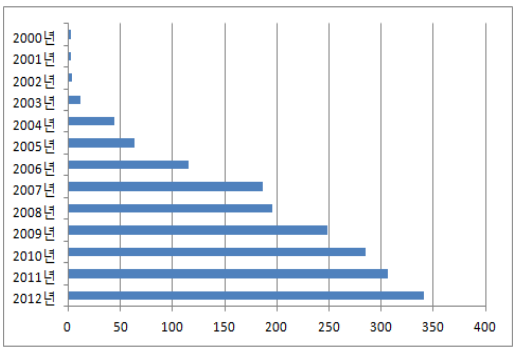

그림 4. 논문의 년도별 발행 수(Digital Forensics)

Figure 4. The numbers of publication papers related to digital forensics by year

디지털 포렌식 관련 논문 중 국가 코드를 포함하고 있는 논문 총 1,777개 논문의 지역별 분포를 살펴보면, 가장 많은 논문을 발행한 국가 top 5는 미국, 영국, 독일, 오스트레일리아, 중국으로 분석되었다. 1위국인 미국의 경우 총 1,111개의 논문이 발행되어 압도적인 논문 발행수를 나타내었으며 그 외 한국은 총 16개 논문이 발행되어 전체 순위 7위를 나타냈다.

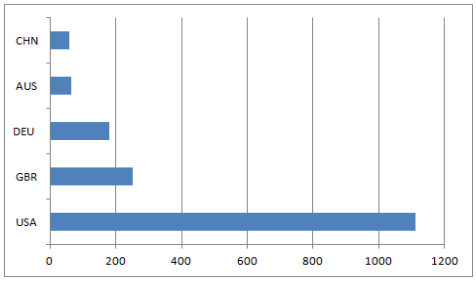

그림 5. 국가별 논문 발행 분포*

Figure 5. The numbers of publication papers by region

다음으로 논문 분석을 통해 시기별로 나타난 주요 부상 키워드를 도출하였다. 이를 도출하기 위하여 시기를 2000년부터 2004년까지 첫 번째 구간, 2005년부터 2008년까지 두 번째 구간, 2009년부터 2012년까지 마지막 세 번째 구간으로 정의하고 각 구간에서 발행된 논문들의 키워드(author keyword, 상위 5개 keywords)를 추출하여 이 들의 발생빈도에 따라 발생빈도가 높은 주요 키워드를 2레벨로 분류하여 표기하였다.

표 9. 시기별 주요 키워드

Table 9. Main keywords according to period

	period1 2000~2004년	period2 2005~2008년	period3 2009~2012년
1레벨	digital evidence, digital	event reconstruction, intrusion	cloud computing, camera identification,

* 미국(USA), 영국(GBR), 독일(DEU), 오스트레일리아(AUS), 중국(CHN)

	tampering, image forensics, network forensics, data mining, digital image forensics, digital investigation, anti forensics, steganalysis, data recovery, file carving, image authentication, multimedia forensics, cybercrime, digital evidence bags, encryption, image processing, incident response, steganography, Microsoft Windows	detection, tamper detection, data hiding, mobile forensics	forgery detection, live Forensics, Image tampering, copy move forgery, JPEG compression, classification, cloud forensics, digital forensic readiness, sensor pattern noise, Android
2 레벨		digital evidence, digital tampering, image forensics, network forensics, digital investigation, steganalysis, digital image forensics, file carving, multimedia forensics, image processing, incident response, steganography, encryption, image authentication,	image forensics, digital image forensics, digital Evidence, anti forensics, digital investigation, file carving, network forensics, data recovery, mobile forensics
		cybercrime, Microsoft Windows	

주) 1 level: 각 구간의 발생빈도 상위 25개 키워드 중에서 이전 시기의 상위 25개 키워드에는 등장하지 않은 신생 주요 키워드.
2 level: 각 구간내 발생빈도 상위 25개 키워드 중에서 1 level에 포함되지 않는 주요 키워드.
단, 주요 키워드에서 digital forensics, computer forensics, computer crime, forensics, security는 제외함.

<표 9>에서 나타난바와 같이 2000년부터 2004년까지 발행된 논문의 키워드 분석을 통해 digital evidence, digital tampering, image forensics, network forensics, data mining, steganalysis, file carving, image authentication, cybercrime, encryption, steganography 등이 주요 키워드에 포함되었고, 2005년부터 2008년까지는 event reconstruction, intrusion detection, tamper detection, data hiding, mobile forensics 등이 주요 키워드로 부상하였다.

2008년까지는 (즉, period1과 period2의 시기까지는) 주요 키워드에 큰 변화가 없었던 것에 비해 2009년 이후에 cloud computing, camera identification, forgery detection, live forensics, cloud forensics, copy move forgery, JPEG compression, classification, digital forensic readiness, sensor pattern noise, android 등 새롭게 떠오른 다양한 부상 키워드를 확인할 수 있다(period3 구간의 1레벨 참고). 즉, 2009년 이후 디지털 포렌식 관련 논문은 양적으로 크게 증가되었을 뿐만 아니라 연구 주제의 범위 역시 큰 폭으로 다양화된 것을 알 수 있다. 한 예로 2009년 이후 부상한 주요 키워드로서 cloud를 포함하는 관련 논문과 Android 관련 논문은 2010년에 처음 등장하였으며, 짧은 기간 내에 전체 관련 키워드 중 발행수 측면에서 상대적으로 높은 비중을 차지하는 부상 키워드로 도출되었다.

4.2.3 뉴스 텍스트기반 트렌드 분석

최근 뉴스 기사에서 언급된 디지털 포렌식 관련 이

슈와 트렌드를 분석하기 위해 2012년부터 2013년까지 발행된 온라인 뉴스 기사 중 '디지털 포렌식'을 포함한 기사의 제목, 내용, 날짜, 출처를 다운받아 제목과 내용의 텍스트를 분석하였다.

<그림 6>은 디지털 포렌식을 중심으로 부상 키워드들의 네트워크를 분석한 것이다. 수사 관련, 정보 보호, 보안 기술 관련, 교육/대학 관련 키워드가 도출되었다.

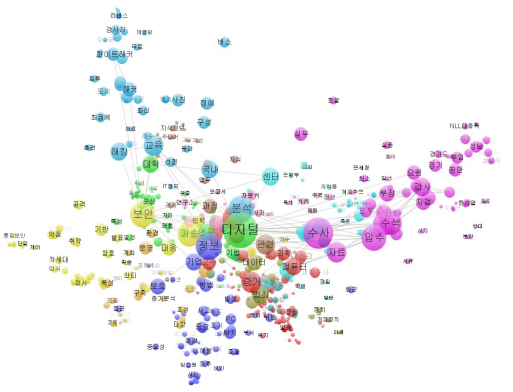

그림 6. 최근 신문기사의 공기어 네트워크 분석('디지털 포렌식'을 중심어로)

Figure 6. Co-occurrence terms network analysis visualization on recent news articles(center keyword is 'digital forensics')

수사와 관련해서는 압수, 자료, 수색, 검사, 계좌추적, 요원, 수사관, 전산 등의 주요 키워드가 있으며 정보 보호와 관련된 PC, 탐지, 예방, 사고, 징후, 접근, 자산, 임직원, 보안사고 등의 키워드가 도출되었다. 또한 보안과 관련된 취약, 대응, 암호, 안티 포렌식, 인식, 통합보안, 공격, 증거분석, 인력, 기술 연구 등의 키워드도 도출되었다. 마지막으로 교육/대학에 관한 키워드로써 해킹, 해커, 훈련, 분석, 센터, 주니어, 최정예, 리버스, 강사진, IT 정보 등의 키워드가 도출되었다.

<그림 7>은 네트워크 포렌식을 중심으로 부상 키워드들의 네트워크를 분석한 것이다. 수사 관련, 네트워크 관련, 디지털/정보기기 관련, 솔루션 기술 관련 키워드가 도출되었다.

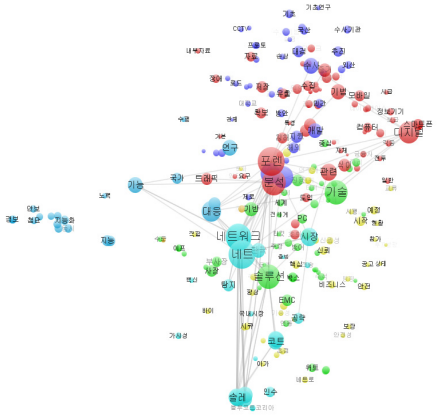

그림 7. 최근 신문기사의 공기어 네트워크 분석('network forensics'를 중심어로)

Figure 7. Co-occurrence terms network analysis visualization on recent news articles(center keyword is 'network forensics')

수사와 관련해서는 수사기관, 대검, CCTV, 기초연구, 대학교 등의 키워드가 도출되었고 네트워크와 관련해서는 탐지, 대응, 지능화, 첨단, 백신, 가시성, 국내시장 등의 키워드가 도출되었다. 다음으로 디지털과 정보기기에 관한 키워드로 스마트폰, 모바일, 침해, 금융사, 컴퓨터, 분석, 유출, 내부자료, 트래픽, 수집, 장애 등의 키워드가 도출되었다. 그 외 솔루션, 기술, 시장, PC 등이 있었다.

5. 결론 및 시사점

디지털 포렌식 분야는 2000년대 중반부터 최근까지 양적인 측면과 내용적인 측면에서 큰 변화와 성장이 있었다. 이러한 추세는 한동안 지속될 것으로 예측되며, 따라서 디지털 포렌식과 관련한 논문, 신문기사 등 빅데이터의 분석을 통해 획득할 수 있는 정보와 가치는 더욱 커질 것으로 예상된다.

본 논문에서는 디지털 포렌식 논문 분석을 통해 2000년 초반에 부상한 주요 키워드, 2000년대 중반에 부상한 주요 키워드, 그리고 20009년 이후 최근에 부상

한 주요 키워드를 도출하였다. 2000년대 초반에는 디지털 증거, 이미지 포렌식, 네트워크 포렌식 등이 주를 이루었고 2000년대 중반에 모바일 포렌식, tamper detection 등이 부상하였으며 최근에는 클라우드 컴퓨팅, 클라우드 포렌식, 카메라 식별, 실시간 포렌식(live forensics), 안드로이드 등의 키워드가 부상하여 시대별로의 주요 트렌드를 알 수 있었다.

또한, 최근 신문 기사의 제목과 본문 텍스트로부터 형태소를 분석, 추출 한 후 디지털 포렌식과 동시 발생 빈도가 높은 연관 키워드를 분석한 결과, 수사와 관련된 키워드뿐만 아니라, 기술적으로는 정보 보호, 보안 기술 관련 키워드가 주요 키워드로 도출되었으며, 그 외 대학, IT 정보, 모의 해킹 등 교육에 관한 키워드도 부상하였음을 알 수 있었다. 한편, 디지털 포렌식 교육을 통한 전문 인력 양성 관련 활동도 활발함을 가늠할 수 있었다.

더불어, 2000년대 초반의 디지털 포렌식 논문 분석 후 주요 부상 키워드 중 하나로 도출되었던 네트워크 포렌식과 동시 발생 빈도가 높은 연관 키워드를 분석한 결과, 수사와 관련된 키워드뿐만 아니라 탐지, 대응, 지능화, 가시성 등 네트워크 관련 키워드와 컴퓨터, 스마트폰, 모바일 등 모바일 정보기기에 대한 키워드가 부상하였음을 알 수 있었다. 최근 스마트폰 등 모바일 기기의 데이터 수집, 분석 기술에 대한 기사와 그 외 네트워크 보안 및 포렌식 관련 솔루션에 대한 기사가 활발히 등장하였던 것으로 보이며, 이를 통해 모바일 기기의 등장이 네트워크 포렌식 기술 수요에 직접적인 영향을 주고 있음을 추정할 수 있다.

향후 클라우드 컴퓨팅 등 다양한 IT 기술의 지속적인 발전으로 인해 디지털 포렌식에 대한 수요를 촉진할 것으로 전망되며, 더불어 포렌식 수사관, 포렌식 툴킷 개발자 및 그 외 포렌식 관련 전문가들은 관련 기술 동향을 모니터링하기 위해 빅데이터를 효율적으로 분석하고 이해해야만 할 것이다.

References

[1] Wikipedia, http://en.wikipedia.org/wiki/Digital_forensics.

[2] S.-D. Jeon, D.-S. Hong, and K.-J. Han, *Digital forensics technology trend and prospect*, The journal of Information Policy, Vol. 13, No. 4, pp. 3-19. 2006.

[3] J. H. Lee, *Data big-bang, big data trend,* Journal of Communications & Radio Spectrum, Special Issue, pp. 43-55. 2013.

[4] S. M. Baek, *Future insight: New industrial prospects and policy direction of Health industry using Big Data*, The Korea Health Industry Development Institute Brief, Vol. 84, pp. 1-20, 2013.

[5] J. I. Beom, S. J. Choe, and D. H. Song, *Best practice and Insight of big data*, Nonghyup Economic Research Institute 2013 CEO Focus 312, pp. 1-38, 2013.

[6] D.-S. Hong, Y.-W. Park, J. T. Lee, H.-G Kim, and D. H. Lim, *Technology commercialization emerging issue analysis*, The proceeding of International Conference on Convergence Content 2013, Vol. 11, No. 2, Japan, 2013.

[7] Macnamara, J.R, *Media content analysis: Its uses; benefits and best practice methodology,* Asia Pacific Public Relations Journal, Vol. 6, No. 1, pp. 1-34, 2003.

[8] M. Koppel and I. Shtrimberg, *Good News or bad news? Let the market decide*, Computing Attitude and Affect in Text: Theory and Applications, Vol. 20, pp. 297-301, 2006.

[9] K. Lerman, A. Gilder, M. Dredze and F. Perira, *Reading the Markets: Forecasting public opinion of political candidates by news analysis*, The proceeding of 22nd

International Conference on Computational Linguistics, pp. 473-480, 2008.

[10] T. Snowsill, I. Flaounas, T. D. Bie, and N. Cristianini, *Detecting Events in a million New York Times articles*, LNAI 6323, Springer, Heidelberg, pp. 615-618, 2010.

[11] O. Ali, I. Flaounas, T. D. Bie, N. Mosdell, J. Lewis and N. Cristianini, *Automating news Ccontent analysis: An application to gender bias and readability*, Workshop on Applications of Pattern Analysis, UK, pp. 36-43, 2010.

[12] SWGDE(Scientific Working Group on Digital Evidence), *Best practices for computer forensics version 2.0*, 2006.

빅데이터 분석을 통한 디지털 포렌식 기술 동향 분석

홍동숙[1], 전상덕[2], 김찬호[1], 김한국[1]

[1] *한국과학기술정보연구원 기술사업화분석실*

[2] *김・장 법률사무소*

요 약

미국의 디지털 포렌식 서비스 시장은 2012년 976 백만 달러로 추정된다. 모바일과 컴퓨터 사용이 늘어나면서 디지털 포렌식에 대한 수요를 촉진하고 있다. 그러나 IT 기술의 지속적인 진화에 따른 많은 도전 과제들이 존재한다. 포렌식 수사관, 포렌식 툴킷 개발자 및 그 외 포렌식 관련 전문가들은 새로운 장비, 시스템, 구성 요소와 관련된 기술 동향, 심지어 미래 기술 동향을 모니터링하고 이해해야만 한다. 따라서 본 논문에서는 신문 기사, 논문 등 빅데이터를 활용하여 디지털 포렌식 기술과 관련하여 최근 부상하는 이슈 및 트렌드를 분석한다.

감사의 글

본 연구는 2013년도 한국과학기술정보연구원의 재원으로 혁신형 기술사업화 정보지원체제 구축 사업의 지원을 받아 수행된 연구임(No. K-13-L03-C02-S03).

Dong-Suk Hong received the Ph.D. degree in the Department of Computer Science and Engineering from Konkuk University in 2008. From 2008 to 2009 she was a researcher at Institute of TMS(Telecommunication Multimedia and SoC) in Yonsei University. She has been a senior researcher at KISTI since 2009. Her current research interests include technology commercialization, big data analysis, emerging issue analysis, intelligent system, and digital forensics technology.

E-mail address: dongsuk.hong@gmail.com

Sang-Duk Jeon received the Engineering Master of degree in the Department of Computer Engineering from Yonsei University in 2003. From 2001 to 2006 he was a Computer Investigator at Digital Forensic Center in the public Supreme Prosecutor's Office. He has been a senior manager at Digital Forensic Team in Kim & Chang since 2007.

E-mail address: zauri3@naver.com

Chan-Ho Kim received the Ph.D. degree in the Department of Economics from Hannam University in 2012. From 1991 to 2000 he was a researcher at KINITI(Korea Institute of Industry and Technology Information). He has been a senior researcher at KISTI since 2001. His current research interests include technology commercialization, intelligence global commerciali zation system, and success and failure of commercialization of technology.

E-mail address: chkim@kisti.re.kr

Han-Gook Kim received the Ph.D. degree in the Department of Industrial Engineering from Tokyo Institute of Technology in 2007. From 2008 to 2009 he was a professor in Woosong University. He has been a senior researcher in the Department of Technology Commercialization Analysis at KISTI since 2009. His current research interests include information system analysis, global technology commercialization, and big data analysis.

E-mail address: hgkim712@kisti.re.kr