The Journal of the Convergence Culture Technology (JCCT)
Vol. 2, No. 4, pp.71-76, November 30, 2016. pISSN 2384-0358, eISSN 2384-0366

http://dx.doi.org/10.17703/JCCT.2016.2.4.71
JCCT 2016-11-10

비식별화 정책 비교 및 빅데이터 산업 활성화 방안

De-identification Policy Comparison and Activation Plan for Big Data Industry

이소진*, 진채은*, 전민지*, 이조은*, 김수정*, 이상현**

So-Jin Lee*, Chae-Eun Jin*, Min-Ji Jeon*, Jo-Eun Lee*, Su-Jeong Kim* ,Sang-Hyun Lee*

요 약 본 논문에서는 미국, 영국, 일본, 중국, 한국의 비식별화 정책비교를 통하여 향후 비식별화 규제의 방향과 빅데이터 산업 활성화 방안을 제시한다. 비식별화 기술과 적정성 평가기준을 효율적으로 활용할 경우, 개인의 사생활을 침해하지 않고 개인정보보호법 제약을 피하면서 개인정보를 산업적으로 활용하여 서비스와 기술을 발전시킬 수 있다. 역작용으로는 비식별화 된 데이터들의 집합체로서 개인을 재식별 할 수 있는 재식별화 문제가 발생할 수 있다. 비즈니스 관점에서는 빅데이터 산업이 활성화되기 위해서 규제의 해소와 빅데이터 활용 등을 위한 제도완화가 필요하며 정보보호 측면에서의 보안강화 조치와 제도정비도 필요하다.

주요어 : 빅데이터, 비식별화, 재식별화, 개인정보보호법

Abstract In this study, de-identification policies of the US, the UK, Japan, China and Korea are compared to suggest a future direction of de-identification regulations and a method for vitalizing the big data industry. Efficiently using the de-identification technology and the standard of adequacy evaluation contributes to using personal information for the industry to develop services and technology while not violating the right of private lives and avoiding the restrictions specified in the Personal Information Protection Act. As a counteraction, the re-identification issue may occur, for re-identifying each person as a de-identified data collection. From the perspective of business, it is necessary to mitigate schemes for discarding some regulations and using big data, and also necessary to strengthen security and refine regulations from the perspective of information security.

Key Words : Big-Data, De-identification, Re-identification, Privacy Act

I. 서 론

이세돌 9단과 세기의 바둑대결을 펼쳤던 알파고는 2016년 ICT분야의 핫이슈다. Google DeepMind가 개발한 알파고의 핵심은 딥러닝(Deep Learning) 기계 학습(Machine Learning) 기술이며 방대하게 수집된 빅데이터 분석을 기반으로 하고 있다. IT 트렌드를 주도하고 있는 사물인터넷(Internet of Things), 핀

*준회원, 배화여자대학교 스마트IT학과
**정회원, 배재대학교 산학협력단
접수일자: 2016년 10월 10일, 수정완료일자: 2016년 10월 25일
게재확정일자: 2016년 11월 4일

Received: 10 October, 2016 / Revised: 25 October, 2016
Accepted: 4 November, 2016
**Corresponding Author: 4834396@naver.com
Industry-Academic Cooperation Foundation, Pai Chai University

테크(FinTech), 인공지능(Artificial Intelligence) 등 최신 IT 기술들의 공통점은 방대하게 생성된 데이터를 핵심 기반으로 하는 서비스라는 점이다. 빅데이터의 활용은 기업은 물론 공공, 통신, 금융, 의료, 유통 등 다양한 산업 분야로 빠르게 확산되고 있으며 최신 IT 기술들이 빠르게 성장할수록 빅데이터 시장도 확대되고 있다.[1]

빅데이터 시대에서 새로운 정보를 얻기 위해서는 비식별 데이터의 활용이 필수 불가결한 요소이다. 비식별 데이터란, 데이터 값 삭제, 가명처리, 총계처리, 범주화, 데이터 마스킹 등을 통해 개인정보의 일부 또는 전부를 삭제하거나 대체함으로써 다른 정보와 결합하여도 특정 개인을 식별할 수 없도록 조치한 데이터를 말한다.[2] 빅데이터는 산업체뿐 아니라 사회 전체의 이익을 증대시키는 효과가 있으므로 적극적으로 빅데이터의 활용을 추진할 필요성이 있다. 신 산업 창출과 기업의 경쟁력 강화에 있어서도 빅데이터의 유용성은 충분히인정되는 반면 개인정보보호법상의 개인정보 보호가 빅데이터 시대에 자칫 소홀해질 수 있는 양면성을 지닌다고 할 수 있다. 빅데이터는 대량의 정보를 수집하고 이를 분석하여 새로운 정보를 창출해 내지만 정보의 수집과정이나 정보의 분석결과에 따라 개인의 프라이버시가 침해되기도 한다.[3]

방송통신위원회는 2014년 '빅데이터 개인정보보호 가이드라인' 을 발표했다. 해당 가이드라인에서는 여러 경로를 통해 수집한 데이터를 결합해 빅데이터로 활용할 경우, 특정 개인을 식별할 수 없도록 조치해야 한다고 권고하고 있다. 빅데이터를 활용함에 있어서 개인정보가 포함되어 있다면, 기업은 정보주체로부터 목적에 대해 사전 동의를 받아야 한다. 하지만 데이터의 양이 크고 범위가 넓기 때문에 처리 목적과 항목에 대해서 정보주체로부터 사전 동의를 받는 것에 대한 현실적 난관으로, 규제의 해소와 빅데이터 활용 등을 위한 제도 완화가 필요하며 정보보안 측면에서의 보안 강화 조치 및 제도가 필수다.[4]

본 논문에서는 미국, 영국, 일본, 중국, 한국 총 5개 국가의 비식별화 지침을 분석하고 이를 통해 향후 비식별화 규제의 방향과 빅데이터 산업 활성화 방안의 지표를 제안하고자 한다.

II. 국가별 비식별화 정책

그림 1.과 같이 중국을 포함한 주요 국가들에서는 빅데이터를 활용하는데 있어 가장 큰 장애 요인을 개인정보 보호 등의 정책 규제라고 인식하고 있다. 비식별화 정책이 마련되어 있는 주요 5개 국가의 비식별 데이터 활용 규제를 비교해 본다.

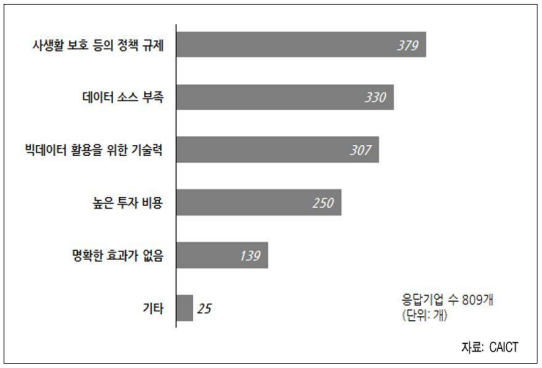

그림 1. 중국 기업 조사 – 빅데이터 활용에 따른 장애 요인
Fig 1. Chinese Enterprise Survey – Barriers to Big Data Utilization

2.1 미국

미국연방거래위원회(FTC)는 2012년 "Protecting Consumer Privacy in an Era of Rapid Change" 라는 권고문을 통해 민간과 공공에게 개인정보의 비식별화 조치 및 규정과 관련한 표준을 제시했다. FTC는 특정 개인, 컴퓨터 및 기타 개인을 식별할 수 있는 장치들과 연관될 수 있는 것은 어떤 정보라도 비식별화의 대상이 되어야 한다고 규정하였다. ①비식별화 기술은 업계가 그 방법을 자율적으로 선택하여 반드시 조치해야 하고, ②정보주체에게 자신의 정보가 비식별화될 수 있다는 사실과 재식별하지 않을 것을 공지하도록 하였으며, ③제3자에게 비식별화 데이터 제공 시 계약상에 재식별 방지를 요구하도록 권고하였다.

2.2 영국

영국은 「1984/1998 데이터보호법」 및 「2000 정보공개법」에 따라 ICO를 설립하여 개인의 프라이버시를 보호하면서도 풍부한 데이터 자원을 활용하는 방법을 도모하고 있다. 2012년 발표된 ICO의 「익명화 규약」은 개인정보를 익명화된 정보로 변경하거나

익명화된 정보가 필요한 기관을 위해 만들어 졌으며, 규약의 권고사항을 이행하면 정보처리자가 익명화된 정보를 공표하더라도 재식별화를 통해 개인정보가 부적절하게 유출될 가능성이 낮다는 합리적인 판단기준을 제시 한다. 「익명화 규약」은 재식별화 과정을 통해 익명화된 데이터가 개인정보로 변환되는 위험에 관한 것이며 재식별에 대한 위험을 점검하기 위한 방법으로 "Motivated Intruder Test"를 활용할 것을 권고한다. 이 검사는 Motivated intruder가 익명화된 정보를 재식별 할 수 있을지를 알아보기 위한 것이다. Motivated intruder는 사전지식 없이 익명화된 정보에서 특정 개인을 식별하기 위해 노력하는 자를 말한다.

2.3 일본

일본의 「개인정보보호법」은 개인정보 정의의 명확화, 개인정보 활용도 제고, 개인정보 보호 강화 등을 주요 골자로 한다. 모호한 개인정보의 개념 때문에 기업들이 개인정보를 회색지대로 인식해 활용이 제대로 이루어지지 못했다는 지적에 따라 개인을 식별할 수 없는 정보는 개인정보가 아니라고 명백히 정의하였다. 개인정보 활용도 제고를 위해 '동의 없이 제공 가능한 익명가공정보' 조항이 신설되었다. 익명가공정보는 특정개인을 식별할 수 없도록 가공한 정보이지만, 정상적인 수단을 통해 복원할 수 없는 수준을 요구하는 것은 아니다. 익명가공정보 취급사업자가 익명가공정보를 취급할 때에는 "개인정보보호위원회"가 정하는 규칙에 따라야 한다. ①안전관리를 위한 조치를 행하고, ②익명가공정보에 포함된 개인정보의 항목을 공표해야 하며, ③제3자에게 제공할 때에는 제3자에게 제공되는 정보가 익명가공정보임을 명시해야 한다. ④해당 익명가공정보를 다른 정보와 조합하거나 입수하는 등의 방법으로 재식별을 해서는 안 된다. 규정을 준수한 익명가공정보는 정보주체의 동의 없이 제3자에게 제공될 수 있다. [6]

2.4 중국

중국의 전국인민대표대회 상무위원회는 총칙, 네트워크 보안 전략 · 기획 · 촉진, 네트워크 운행 보안, 네트워크 정보 보안, 모니터링 경보와 응급 처치, 법률 책임, 부칙 등 모두 7장 68조로 이루어 진 '중화인민공화국 네트워크 안전법(초안)' 중 4장 네트워크 정보보안 항목을 통해 개인정보 보호 사항을 규정하였다. ①네트워크(웹사이트 · SNS · 커뮤니티 등) 운영자(업체 등)는 완전한 사용자 정보보호 제도를 갖춰야 하며, 사용자 개인정보, 프라이버시 및 상업 비밀에 대한 보호를 강화해야 한다. ②네트워크 운영자는 공민(국민) 개인정보를 수집 · 사용할 때 합법적이고 정당하며 필요한 원칙을 따라야 하며, 정보 수집 · 사용의 목적 · 방식 · 범위를 명시하고, 수집자의 동의를 거쳐야 한다. ③네트워크 운영자는 자체 제공하는 서비스와 무관한 공민 개인정보를 수집해서는 안 되며, 법률 · 행정법규의 규정과 쌍방의 약정을 위반해 공민 개인정보를 수집 · 사용해서는 안 된다. 법률 · 행정법규의 규정 또는 사용자와의 약정에 의거해, 보존 중인 공민 개인정보를 처리해야 한다. ④네트워크 운영자는 공민 개인정보를 수집 · 사용할 때 그 규칙을 공개해야 한다. 네트워크 운영자는 수집한 공민 개인정보에 대해 엄격하게 비밀유지를 해야 하며, 유출 · 변조 · 훼손해서는 안 되고 판매 또는 불법적으로 타인에게 제공해서도 안된다. 네트워크 운영자는 기술 조치와 기타 필요한 조치를 취해 공민 개인정보 보안을 확보해야 하고, 수집한 공민 개인정보가 유출 · 훼손 · 분실되는 것을 방지해야 한다. ⑤정보 유출 · 훼손 · 분실 상황이 발생하거나 발생할 가능성이 있을 경우, 즉각 보완조치를 취해야 하며,□영향을 받을 수 있는 사용자에게 알리고 규정에 의거해 유관 주관 부처(기관)에 보고해야 한다. ⑥공민은 네트워크 운영자가 법률 · 행정법규의 규정 또는 쌍방의 약정을 위반하고 자신의 개인정보를 수집 · 사용한 것을 발견했을 경우, 네트워크 운영자에게 그 개인정보를 삭제하라고 요구할 권리가 있다. 또 공민은 네트워크 운영자가 수집 · 저장한 해당 개인정보에 오류가 있을 경우, 운영자에게 정정을 요구할 권리가 있다. ⑦어떤 개인 또는 조직이든 절취 혹은 기타 불법적인 방식으로 공민 개인정보를 얻을 수 없으며, 타인에게 공민 개인정보를 판매하거나 불법적으로 제공할 수 없다. ⑧법에 의거해 네트워크 보안 감독관리 직책을 책임진 부처(기관)은 반드시 직책 이행 중 알게 된 공민 개인정보와 프라이버시, 상업 비밀에 대해 엄격하게 비밀유지 해야 하며, 타인에게 유출 · 판매 혹은 불법적으로 제공해서는 안 된다.[7]

2.5 한국

한국 방송통신위원회는「빅데이터 개인정보보호 가이드라인」을 통해 빅데이터 처리 과정에서 발생하는 정보의 수집 및 이용 방법, 비식별화, 투명성 확보, 민감정보 및 통신내용 처리금지 등을 규정하였다. ①수집 시부터 개인식별 정보에 대한 철저한 비식별화 조치를 취한 후 수집 · 저장 · 조합 · 분석 및 제3자 제공 등이 가능하고, ②비식별화 조치 후 빅데이터 처리 사실, 목적, 수집 출처 및 정보 활용 거부권의 행사 방법 등을 이용자에게 투명하게 공개하도록 하였으며, ③개인정보가 빅데이터 처리 과정 및 생성정보에 재식별 될 경우, 즉시 파기하거나 추가적인 비식별화 조치를 해야 한다. ④특정 개인의 민감 정보 생성을 목적으로 한 정보 및 통신 내용의 수집 · 이용 · 저장 · 조합 · 분석 등의 처리를 금지하고, ⑤수집된 정보를 저장 · 관리하고 있는 정보 처리시스템에 대한 기술적 · 관리적 보호조치를 적용하도록 하였다. 보호조치란, 침입차단 시스템 등 접근 통제장치 설치, 접속 기록에 대한 위 · 변조 방지 조치, 백신 소프트웨어 설치 · 운영 등 악성프로그램에 의한 침해 방지 조치를 말한다. 그림 2.는 참고문헌[4]를 참고하여 국가별 익명화 지침의 비교를 보여준다.

국가별 익명화 지침 비교

	미국	영국	일본	중국	한국
익명화 지침	비식별화 지침A, 2012년	익명화 규약B, 2012년	「개인정보보호법」 '익명가공정보' 조항 신설, 2015년	중화인민공화국 네트워크 안전법, 2015년	개인정보 비식별 조치 가이드 라인, 2016년
지침 제정 및 감독기관	FTC	ICO	개인정보보호위원회	상무위원회	방송통신위원회
익명화 정보 동의	거부권 없는 고지	거부권 없는 고지	정보항목 공개의무	수신자의 동의	거부권 없는 고지
제3자 제공	계약상 재식별화 방지를 명시	-	익명가공정보라는 사실을 명시	유출,판매,불법적 제공금지	재식별화 위한 관리에 관한 내용을 계약서에 포함
기타	-	Motivated intruder test 권고	-	-	k-익명성 모델 기본적으로 적용

A: Protecting Consumer Privacy in an Era of Rapid Change
B: Anonymisation; managing data protection risk code of practice

그림 2. 국가별 익명화 지침 비교
Fig 2. Comparison of national anonymization guidelines

III. 비식별화 규제개선 및 빅데이터 산업 활성화방안

빅데이터 산업의 활성화를 위해서는 개인정보의 보안 강화와 사용 규제 완화의 절충이 필요하다. 따라서 본 논문에서는 향후 비식별화 규제의 방향과 빅데이터 산업 활성화 방안을 제안한다.

3.1 비식별화 규제 개선 방향

3.1.1 개인정보의 재식별 위험 점검 모델 개발

영국의 익명화 규약에서는 "Motivated Intruder Test"의 활용을 권고하고 있다. 재식별에 대한 위험을 점검하기 위한 모델은 영국에만 유일하게 마련되어 있다. 한국도 영국의 Motivated Intruder Test를 모방하여 재식별에 대한 위험을 점검하는 모델을 개발하고 개인정보 재식별을 방지하기 위한 기준을 제시하여야 한다.

3.1.2 재식별 데이터 관련 법제화 필요

미국의 인터넷 검색 업체 AOL(아메리카 온라인, America Online)은 개인정보 비식별화 작업을 완료한 65만명 이상의 이용자의 검색기록 2천만 개를 대중에게 공개했다. AOL은 재식별화의 가능성을 차단하지 못한 정보를 공개하여 개인정보가 재식별화 되어 인터넷에 유출되는 일이 발생했다. 현재 비식별 데이터의 활용에서 데이터 재식별 문제가 빈번하게 발생하고 있지만 비식별화 규제에서 재식별 데이터에 관한 내용은 규정되어 있지 않다. 개인정보의 보호를 위해서는 재식별 데이터의 유통에 대한 정부의 관리 체계가 필요하다. 재식별 데이터의 생성을 방지하는 방안 마련이 가장 중요하며 명확한 활용 기준 및 범위, 데이터 파기 기준 등의 항목을 포함한 관련 법 제정이 필요하다.

3.1.3 국가적 차원의 개인정보 이용내역 고지 항목 추가

국가적 차원에서 개인정보 이용내역을 SMS나 메신저를 통하여 정보주체에게 의무적으로 고지하는 항목을 개인정보 비식별화 정책에 추가하여야 한다. 개인정보 사용 동의를 받지 않은 곳에서 수집된 비식별 정보를 사용한다면 그림2와 같이 알림서비스를 통해 개인정보 처리 항목과 목적을 고지한다. 재식별화 된 정보나 민감 정보를 포함하고 있다면 거부권을 행사하는 방법도 함께 고지하도록 하고 거부권을 행사하면 해당하는 정보를 파기한다. 이용내역을 의무적으로 고지한다면 사전 동의를 받기 어려운 현실적인 난관

을 해소하고 거부권 행사를 통해 개인정보 유출에 대한 불안감을 낮출 수 있을 것이다.

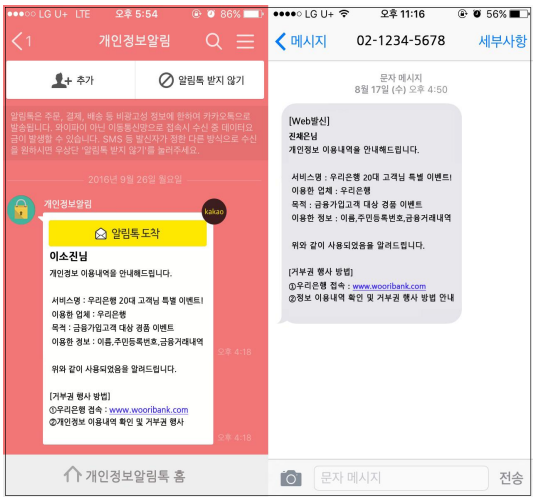

그림 3. SMS와 메신저 알림서비스 화면
Fig 3. SMS and Messenger Notification Service screens

3.2 빅데이터 산업 활성화방안

3.2.1 개인정보 이용 내역 고지 제도를 통한 사전 동의 문제 해소

빅 데이터는 데이터의 양이 크고 범위가 넓기 때문에 정보주체로부터 사전 동의를 받는 것이 현실적으로 어렵다. 현재 개인정보보호 정책은 정보주체에게 사전 동의를 받아야 하고 재식별 된 데이터의 사용은 불가능하다. 이러한 문제가 빅데이터 산업이 발전하지 못하는 가장 큰 원인이다. 앞서 제시한 국가적 차원의 개인정보 이용 내역 고지 제도를 이용하여 개인정보 사전 동의 문제점을 보완한다면 빅데이터를 활용한 산업이 더 발전할 수 있을 것이다.

3.2.2 정부의 공공 데이터 개방 정책 확대

정부는 빅데이터 중에서도 공공 데이터를 공급하여 공공 빅데이터 형성에 주도적인 역할을 해야 한다. 현재 정부가 운영하고 있는 공공 자료 포털은 아직 양질의 빅데이터를 공급하는 역할을 해내기에는 부족한 점이 많다. 빅데이터 시장의 주도국이라 할 수 있는 미국과 영국 정부를 벤치마킹 하여 빅데이터 인프라 구축과 공공 데이터 활용을 통한 시장 창출에 주력해야 한다.

3.2.3 빅데이터 관련 성공사례 홍보

빅데이터 도입에 대한 인식이 긍정적으로 변화하고 있으며, 이를 활용하여 빅데이터 산업을 활성화하고자 하는 기대와 빅데이터 성공사례에 대한 관심이 높아졌다. 미국의 유료 동영상 콘텐츠 유통 플랫폼인 넷플릭스(Netflix)에서는 자체 제작한 드라마인 '하우스 오브 카드(House of cards)'를 흥행 성공시켰다. 빅데이터 분석을 통해 시청자들이 선호하는 배우를 캐스팅하여 드라마를 제작하였으며 시청 패턴을 분석하여 시청자들이 다시보기나 몰아보기를 한다는 사실을 도출해내어 에피소드를 한꺼번에 공개한 점이 성공 요인이었다. 일본의 생활 잡화 판매점인 무인양품(MUJI)은 모바일 애플리케이션이나 웹을 통해 고객의 온·오프라인 행동데이터를 분석하여, 최적 할인율의 쿠폰을 실시간으로 제공하는 형식으로 오프라인 방문자 수를 몇 주 만에 대폭 증가시켰다. O2O(Online to Offline)전략에 빅데이터 솔루션을 활용하여 위치기반 서비스 사업에 새로운 시장을 개척하였다. 국내에서는 CJ대한통운이 물류의 전반적인 관리와 통제를 효과적으로 하기 위해 빅데이터를 통합물류시스템에 적용시키는 등의 활용 사례가 있다. 이처럼 빅데이터 산업 활성화를 위하여 다양한 방식으로 관련 성공사례를 분석하고, 적극적인 홍보 및 활용을 통해 새로운 사업을 창출하도록 하여야 한다.

IV. 결 론

본 논문에서는 비식별화 정책이 마련된 5개 국가의 정책을 비교하고 향후 비식별화 규제의 방향과 빅데이터 산업 활성화 방안을 제시하였다. 빅데이터가 다양한 산업 분야로 발전하고 있지만 새로운 시대에 정보를 얻기 위해서는 비식별 데이터의 활용을 위한 제도 완화가 필요한 상태에 있다. 국가별 개인정보보호법 비교를 통하여 기존의 규제를 개선할 방안으로 영국의 Motivated Intruder Test를 모방하여 재식별에 대한 위험을 점검하는 모델을 개발하고 재식별 데이터의 생성을 방지하기 위한 방안을 마련해야한다. 재식별 데이터의 유통에 대한 정부의 관리체계가 필요하며, 명확한 재식별 데이터 활용 기준 및 범위, 데이터 파기 기준 등의 항목을 포함한 관련 법 제정이 필

요하다. 국가적 차원에서 개인정보 이용내역을 SMS나 메신저를 통하여 정보주체에게 의무적으로 고지해야 한다. 앞서 제시한 비식별화 규제 개선을 통해 사전 동의 문제점을 해소하고 빅데이터 시장의 주도국인 미국과 영국 정부를 벤치마킹하여 국가적 차원의 공공 데이터 공급이 이루어진다면 빅데이터 산업 활성화에 중요한 밑바탕이 될 것으로 기대된다. 홍보가 부족한 빅데이터 관련 산업을 다방면으로 성공사례 분석을 하여 적극적인 홍보 및 활용을 한다면 새로운 빅데이터 사업을 창출하는 계기가 될 것이다. 향후 연구에서는 본 논문에서 제안한 빅데이터 산업의 활성화 방안을 기반으로 하는 데이터 재식별 방지에 대한 연구와 새로운 빅데이터 산업의 개발이 필요할 것으로 판단된다.

References

[1] Fasoo.com, “Big Data Utilization and Privacy”, 2016.
[2] Yang Hyun Chul, “Effects of applying personal information de-identification technology on intention to use big data : big data providers and consum”, Kwangwoon University, 2016.
[3] Son Young Wha, “Privacy Policy in Big Data Age", Korean Business Law Society, 2014.
[4] Kim Tae Hyung, “Big Data Utilization and Privacy”, Security News, 2016.
[5] Information and Communication Technology Promotion Center, “Big Data Utilization Status in China”, 2016.
[6] NICE Credit Information CB Research Institute, “Country-specific de-identification guidelines for using Big Data”, 2016.
[7] Ongi Hong, “China Cyber Security Act”, Internet Information Control · Security Emphasis Security News, 2016.