DB Issue Report

빅데이터와 개인정보보호

주요 내용

I. 배경
II. 기술 및 정책 현황
III. 시사점

KODB 한국데이터베이스진흥원
KOREA DATABASE AGENCY

I 배경

o 빅데이터 시대의 개인정보보호

- 빅데이터 분석으로 사회 현상의 이해와 예측, 경제적 이익을 얻을 수 있을 것이라는 기대에 따라 개인 사생활 보호를 간과
- 빅데이터 분석을 통한 개인 대상 마케팅의 경우, 정보 주체의 개인정보 자기결정권 침해 위험 내재
- 반면 지나친 개인정보보호는 서비스 품질을 저하시켜 개인과 서비스 제공자 모두에게 불이익 초래

o 개인정보 침해 위협[1)]

구분	주요 내용
구글	● 2014년 5월 유럽사법재판소는 '잊혀질 권리'를 인정하는 판결로, 구글의 과도한 개인정보 수집이 개인 인권과 충돌한다는 논란 촉발 ● 2014년 초 프랑스·스페인은 구글의 개인정보 수집 정책이 사생활 보호규정에 위반한다며, 각각 벌금 15만·90만 유로 부과
	● 아동포르노사진을 전송한 개인의 메일을 경찰에 공개하여 논란 발생 ● 아동성폭력 관련 이미지의 특정정보만 분석해 걸러내며 아동성폭력 예방 이외 다른 목적으로 사용하지 않는다고 해명
	● 광고 활용 목적으로 2006년부터 구글 교육용 앱을 사용해온 학생과 교사 약 3천만 명의 지메일 계정과 달력, 클라우드에 보관된 자료 등을 검열

1) ZDNet Korea, 왜 그들을 '21세기 빅브라더'라고 하는가, 2014.7.28, 비즈니스포스트, 구글 또 '빅브라더' 논란에 휩싸여, 2014.8.5, 전자신문, 교실로 간 '빅데이터'..."사생활 침해 vs 맞춤형 교육", 2014.3.25 등을 참고하여 재구성.

구분	주요 내용
구글	● 스트리트 뷰는 동의 없이 제3자의 이미지를 촬영해 사생활 침해 문제가 발생하였고, 촬영하는 와이파이 차량이 무단으로 이메일과 비밀번호 등 개인정보를 수집 ● 방송통신위원회는 2014년 1월 스트리트뷰 관련 정보를 모으면서 이용자 개인의 아이디부터 신용카드 정보까지 무차별적으로 모았기 때문에 구글 본사에 과징금 2억1230만 원 부과
	● 구글 글래스는 타인이 모르는 사이 사전 허가 없이 영상 및 사진을 촬영할 수 있어 개인정보와 콘텐츠 저작권 도용, 사생활 침해 우려(불법 영화 복제 및 몰래카메라 등)
페이스북	● '감정 조작 실험' : 페이스북은 2012년 초 사전 공지없이 68만 9천 3명의 뉴스피드를 조작해 긍정적이거나 부정적인 내용의 피드를 보여주고 사용자들의 심리 반응을 살펴 SNS상에서의 감정전이 현상을 비밀리에 연구 ● 사용자들이 올리는 게시물을 인위적으로 바꿔 영향을 끼칠 수 있다는 점에서 사용자 데이터 보호 측면의 문제
	● 스마트폰의 위치추적 시스템(GPS) 정보를 기반으로 페이스북 친구끼리의 위치를 알려주는 '니어 바이 프렌즈'는 실시간으로 위치정보를 추적해 자동으로 알려준다는 점 때문에 세세한 개인 일정이 전부 노출될 수 있음 ● 앱을 끈 상태에서도 위치정보가 계속 추적된다는 점, 정보이력을 삭제하지 않는 한 해당 정보가 페이스북 서버에 계속 남게 된다는 점 등 문제의 소지
	● 2014년 5월 개발된 페이스북의 '소리 인식 기능'은 스마트폰 사용자가 페이스북 앱을 사용할 때 기기 내장 마이크가 사용자 주변의 소리를 인지(사용자의 기본정보와 함께 기분과 상태까지 파악 가능)
	● 2014년 1월 고객정보를 광고주에게 팔아넘겼다는 의혹을 받아 사용자들로부터 제소

구분	주요 내용
애 플	● 2014년 6월 아이폰 사용자가 위치서비스 기능을 꺼도 위치정보가 수집되는 버그 발생, 2010년 6월22일부터 2011년 5월4일까지 이용자들의 동의 없이 위치정보 수집 ● 그러나 재판부는 고객이나 고객의 아이폰을 특정하지 않았기 때문에 애플이 수집한 위치정보가 특정 개인의 위치정보는 아니라고 판단, 사용자 측 손해배상 청구 기각
	● iOS 보안을 연구해온 조나단 지드자스키는 미국 뉴욕에서 개최된 '해커스 온 플래닛 얼스' 컨퍼런스에서 애플이 iOS에 사용자를 모니터링할 수 있는 백도어를 숨겨놓았다고 주장해 파문 ● 애플은 "iOS 진단 기능(사용자의 개인정보 및 보안을 침해하지 않고 기업의 IT 부서, 개발자, 애플이 기술적 문제 해결을 위해 필요한 정보를 제공하며, 동의가 없다면 데이터 전송은 있을 수 없다"고 해명
아 마 존	● 아마존은 고객들이 구입한 물건, 사려고 하는 물품, 쇼핑을 하긴 했지만 사지 않은 물품, 다른 사람들에게 추천한 물품 등에 대한 정보 수집하여 자회사, 제휴회사들끼리 공유
	● 무인 소형 택배기 드론은 남의 집을 엿본다든가 타인의 동의 없는 무단 촬영 등 사생활 침해 문제로 운용 가이드라인을 만들라는 행정명령을 받음
기 타	● 뉴욕시가 추진하는 '티치투원(Teach to One)' 프로그램에 따라 시카고, 뉴욕, 워싱턴 학교는 각 학생의 수학 성취도를 추적 분석, 소프트웨어가 각 학생의 질문·과정별 취약점을 발견 → 결과에 따라 학생별 특성에 따른 개별 학습 프로그램을 운영하고 일대일 온라인 강의 실시 ● 그러나 많은 부모가 아이의 초기 학력 관련 정보가 훗날 미칠 악영향을 두려워하고 있으며, 미국 의회는 학생 데이터 접근을 제어하거나 부모가 데이터 수집 폭을 제한하는 등의 관련 법안 상정 고려 중

Ⅱ 기술 및 정책 현황

o 주요 기술[2)]

단계	필요기술
데이터 수집 단계	● 데이터 수집시 동의 기술 - 개인정보 수집시 동의 지원
	● 데이터 수집시 법률적 위반사항 검토 기술
	● 데이터 수집 거부 기술 - 로봇 등 자동 수집 배제 표준(권고안) - 대용량 크롤링 제한 기술(허용·불허용 의사 표시)
데이터 저장 및 관리 단계	● 데이터 암호화 기술 - 데이터 암호화 기술(공개키 암호화, 대칭키 암호화) - DB 성능 저하가 되지 않는 데이터 암호화
	● 데이터 접근통제(제어) 기술 - 임의 접근통제, 강제 접근통제, 역할기반 접근통제(일반적) 등 - 침입 탐지 및 차단 시스템 - VPN(Virtual Private Network) 등 네트워크 기반 기술 - 사용자 인증·권한 부여 등 계정 관리 기술
	● 데이터 필터링 및 등급 분류 기술 - 데이터 등급별 분류(필터링)·관리기술 - 데이터 자동 필터링(개인정보 자동 비식별 기술)
데이터 처리 및 분석 단계	● 익명화된 데이터 처리 기술 - **PPDM**(Privacy Preserving Data Mining : 프라이버시 보호 분석 기술)* : 데이터 소유자의 프라이버시를 침해하지 않으면서도 데이터에 함축적으로 들어 있는 지식이나 패턴을 찾아내는 기술 - K-anonymity & L-diversity 기술(DB를 분석하여 통계정보만 제공) · K-anonymity : K개 이상의 동일한 데이터를 유지하여 특정인이 추론될 확률을 1/k 이하로 낮추는 기술 · L-diversity : 민감한 데이터의 종류를 L개 이상 유지하는 기술

2) 한국데이터베이스진흥원, 2014, 2014 데이터베이스 백서 및 이재식, 2013, 빅데이터 환경에서 개인정보보호를 위한 기술, Internet & Security Focus 2013 3월호, 한국인터넷진흥원을 재편집.

단계	필요기술
	● 암호화된 데이터 처리 기술 - 순서보존 암호화 기술 : 암호화된 데이터 검색 - 연산보존 암호 기술 : **암호화된 데이터 연산***(MIT 10대 유망기술)
데이터 분석결과 가시화 및 이용 단계	● 이용자 동의 관련 기술 - 이용자 동의 기술 - 빅데이터 분석 결과의 영향력 사전 예측 기술 - 사전 동의없는 경우 사후 동의 지원 기술
	● 분석정보의 이용 모니터링 기술 - 빅데이터 분석 모니터링 기술(이용 내역 고지) - 정보 이용과정 공개 등 모니터링 기술
데이터 폐기 단계	● 데이터 폐기 모니터링 기술 - 데이터 분석, 활용 후 폐기 확인 기술
	● 분산 환경에서 완전한 데이터 폐기 기술(디가우징)

〈참고〉 암호화된 데이터 연산

- 서울대 천정희 교수는 대용량 데이터 상태로 연산이 가능한 제4세대 암호 기술(대용량 데이터를 처리하기 위한 완전 동형 암호 기술) 개발
- 일반적으로 암호화된 정보들은 검색이나 통계처리 등을 위해 모두 복호화해 연산한 후 다시 암호화해 저장해야 하는데, 이는 매우 비효율
- 이 기술은 암호화된 상태 그대로 연산이 가능해 데이터 처리 속도가 높고 외부 유출로부터 데이터의 안전성을 지킬 수 있음
- 또한 비트 단위가 아니라 큰 숫자 단위로 암호화를 하는 것이 가능해 같은 암호문이라도 그 안에 더 많은 원문 정보를 저장할 수 있고, 숫자들 간의 사칙연산이나 다항식 연산과 같은 널리 쓰이는 연산을 직접적으로 수행할 수 있어 훨씬 효율적
- 암호화된 상태에서도 탐색은 물론, 제한된 열람이라든지 통계처리가 가능한 암호를 만들어, 내부자의 데이터 유출 등의 사고를 사전 방지할 수 있음
- 예를 들어 기업이나 보험회사, 국가 등에서 통계나 보험료 등을 계산할 때 암호화된 개인정보를 해제(복호)하지 않고도 처리 가능
- 클라우드와 모바일 정보 이용 환경에서 공공·금융 등 주요 서비스 제공과 관련된 보안 목표를 달성하기 위해 꼭 필요

※ 출처 : 사이언스타임스, 암호화된 정보, 해제 없이 연산한다, 2013.2.14

o 개인정보 제거 방법[3)]

기법	주요 내용
가명처리 (pseudonymisa-tion)	● 개인정보 중 주요 식별 요소를 다른 값으로 대체하여 개인 식별을 곤란하게 함 (예) 홍길동, 35세, 서울 거주, 한국대 재학 → 임꺽정, 30대 서울 거주, 국제대 재학 ● 다른 값으로 대체하는 규칙이 노출되더라도 개인 식별이 불가능해야 함
총계 처리 (Aggregation) 또는 평균값 대체 (Replacement)	● 데이터의 총합 값을 보임으로서 개별 데이터의 값을 보이지 않도록 함 (예) 임꺽정 180cm, 홍길동 170cm, 이콩쥐 160cm, 김팥쥐 150cm → 물리학과 학생 키 합 : 660cm, 평균키 165cm ● 특정 속성을 지닌 개인으로 구성된 단체의 속성 정보를 공개하는 것은 비식별화에 무의미 (예) 에이즈 환자 집단임을 공개하면서 특정 인물 '갑'이 그 집단에 속함을 알 수 있도록 표시하는 것
데이터 값 삭제 (Data Reduction)	● 데이터셋에 구성된 값 중에 필요없는 값 또는 개인 식별에 중요한 값을 삭제 (예) 홍길동, 35세, 서울 거주, 한국대 졸업 → 35세, 서울 거주 주민등록번호 901206-1234567 → 90년대 생, 남자 날짜정보(자격취득일자, 합격일 등)의 연 단위 처리
범주화 (Data Suppression)	● 데이터의 값을 범주의 값으로 변환 (예) 홍길동, 35세 → 홍씨, 30-40세
데이터 마스킹 (data masking)	● 공개된 정보 등과 결합하여 개인을 식별하는데 기여할 확률이 높은 경우, 주요 개인 식별자가 보이지 않도록 처리하여 개인을 식별하지 못 하도록 함 (예) 홍길동, 35세, 서울 거주, 한국대 재학 → 홍**, 35세, 서울 거주, **대학 재학 ● 남아 있는 정보 그 자체로 개인을 식별할 수 없어야 하며 인터넷 등에 공개되어 있는 정보 등과 결합하였을 경우에도 개인을 식별할 수 없어야 함

3) 안전행정부, 공공정보 개방·공유에 따른 개인정보보호지침, 2013.9.

o 미국・유럽 vs 한국

- 유럽 및 미국 등에서 '망각의 권리' 등으로 논의되다 2012년 EU 유럽정보보호지침 개정안에 '잊혀질 권리[4]' 신설

※ EU「General Data Protection Regulation」Article 17 Right to be forgotten and to erasure

- 미국 캘리포니아주, 2013년 9월, 18세 이하 미성년자에 한해 인터넷 서비스업체에 자신 관련 기록물을 지우거나 숨기도록 요청할 수 있는 법안 통과[5]

- 미국은 2014년 5월까지 프라이버시 보호를 위한 빅데이터 정책을 검토, '소비자 프라이버시 권리 장전(2012년 오바마 행정부 제안)'의 통과 촉구, 사이버보안 입법안(2011년 제안)의 이행, 전자 커뮤니케이션에 관한 프라이버시법 개정 등을 제안[6]

- 한국은 미국・유럽과는 반대로 안행부에서 '빅데이터 개인정보보호 가이드라인'을 준비하는 등 개인정보보호를 완화하려는 시도

※ 개인정보보호법 제36~37조 개인정보 정정·삭제 규정이 '잊혀질 권리'와 유사

o 빅데이터 개인정보보호 가이드라인

- 2013년 12월 18일 방송통신위원회와 한국인터넷진흥원이 빅데이터 개인정보보호 가이드라인(안) 발표 후 지금까지 지속 수정 중

- 개인정보보호위원회에서 「개인정보보호법」및 「정보통신망 이용촉진 및 정보보호 등에 관한 법률」의 규정과 입법 취지에 부합하지 아니하는 일부 내용을 포함하고 있으므로 개인정보보호 관련 법률의 내용과 체계에 부합하도록 재검토하도록 권고

4) 본인이 원하는 경우, 온라인 상의 모든 개인정보를 삭제할 수 있는 권리. (출처 : 지성우, 기본권 이론적 관점에서의 '잊혀질 권리(Right to be fotgotten)'에 대한 시론적 고찰, 언론중재, 언론중재위원회, 2011)
5) 매일경제, 잊혀질 권리' 비즈니스 활황…남이 퍼간 사생활 다 지워드립니다. 2013.10.18.
6) 한국인터넷진흥원, 미국 백악관, 빅데이터 활용과 프라이버시 보호 강화를 위한 빅데이터 정책 검토, 인터넷 및 정보보호 동향, Vol.2. 2014.

- (주요 내용) 동의 없이 공개된 개인정보 수집, 이용 내역정보 수집, 공개된 개인정보 및 이용내역정보 등을 활용하여 새로운 개인정보 생성, 공개된 개인정보의 제3자 제공이 정보주체의 동의없이 가능하도록 규정

- (주요 반론) 제한 없이 일반 공중에게 공개되었다 하여 개인정보보호법이나 정보통신망법의 규제를 받지 않는다거나 구체적 사정을 고려함이 없이 해당되는 모든 정보주체가 그 수집에 필요한 동의를 한 것으로 해석하거나 간주할 수는 없으며, '공개된 개인정보'라 하여 당연히 정보주체의 동의 없이 제3자에게 제공할 수 있다고 볼만한 법률적 근거가 없음

- 정보 수집은 옵트아웃 방식을 적용하여 현행법과 달리 사업자의 의무를 면제해주는 점이 문제이며, 정보 이용시 비식별화 의무는 사업자가 서비스를 하기 힘든 과도한 규정이라는 지적

III 시사점

o 빅데이터 환경에 맞는 새로운 기술과 정책 필요

- '데이터 암호화·접근통제' 등의 기술은 많이 성숙되어 있으나, '데이터 이용 동의·모니터링' 등은 연구 필요
- 식별이 불가능해도 여러 경로를 통해 모은 개인정보를 취합, 분석할 경우 식별이 가능해질 수 있어 이에 대한 보완 필요
- 개인정보보호를 위해서는 옵트인(Opt-In) 방식이 효과적이나 빅데이터 분석을 위해서는 개인정보 제공의 인센티브를 받거나 삭제 요청할 수 있는 옵트아웃(Opt-Out) 방식의 보완 필요
- 개인정보를 비식별화하여 개인의 사전·사후 동의 없이 데이터를 유통 할 수 있는 옵트온(Opt-On) 처리 기술 연구[7)]

< 유통용 데이터 생산 처리 방법>

o 개인 비식별화 절차

①개인정보 비식별화	⇒	② 검증 및 인증	⇒	③ 유통 및 폐기
개인정보 비식별화 알고리즘 적용		알고리즘 준수여부 검증 및 인증		기술·운영적 지침

o 개인정보 비식별화 원칙(N:1원칙)

· 유일성 제거 : 개인식별 가능 정보 제거
· 동일성 제거 : 개별 개인정보 ! = 대표레코드
· 연속성 제거 : 유사 개인 그룹핑의 연속성 제거

연락처 : 02-3708-5361, taehoon@kodb.or.kr

7) 연세대학교 이원석 교수의 의견임.