Journal of The Korea Institute of Information Security & Cryptology(JKIISC) ISSN 1598-3986(Print)
VOL.23, NO.5, October 2013 ISSN 2288-2715(Online)
http://dx.doi.org/10.13089/JKIISC.2013.23.5.977

소셜네트워크서비스 개인정보 노출 실태 분석

최 대 선,†‡ 김 석 현, 조 진 만, 진 승 헌, 조 현 숙
한국전자통신연구원

Personal Information Exposure on Social Network Service

Daeseon Choi,†‡ Kim Seok Hyun, CHO, JIN-MAN, Seung-Hun Jin, Hyun-Sook Cho
Electronics and Telecommunications Research Institute

요 약

페이스북과 트위터 한국인 이용자 계정 934만개를 조사하여 개인정보 노출 현황을 분석해보니 다양한 개인정보가 많이 노출되어 있었다. 이메일, 전화번호 같은 식별정보는 1% 미만으로 많이 노출되어 있지 않았지만, 이름, 학교같은 비식별 정보를 통해 개인을 특정할 수 있는 경우가 많았는데, 이름이 유일하여 개인을 특정할 수 있는 계정이 35만개에 달했으며, 2개 이상의 정보를 조합하여 개인을 특정할 수 있는 경우는 297만 명에 달했다. 또한, 동일인이 소유한 페이스북과 트위터 계정의 연결 가능성을 분석하였는데, 동일인일 가능성이 있는 계정 쌍 34만개를 찾을 수 있었다. 계정을 연결할 수 있다는 것은 개인을 식별하고 특정했음을 의미한다. 비식별정보를 통한 특정 가능성과 연결가능성을 통해 식별정보만을 필터링하는 기존 개인정보보호방안에 한계가 있음을 알 수 있다.

ABSTRACT

This paper presents result of researching personal information exposure of Korean twitter and facebook users. Personally identifiable information such as e-mail and phone numer is exposed in the accounts less than 1%. However there are many cases that a person is identified by non personally identifiable information. For example, 350 thousands accounts are distinguished with other accounts because its name is unique. Using combination of information such as name and high school, we can distinguish 2.97 millions accounts. We also found 170 thousands account pairs that are candidate of one users' own account. Linkability between two accounts in two different domains means that the person is identified. Currently, only personally identifiable information is protected by policy. This paper shows that the policy has limited effects under the circumstances that a person can be identified by non personally identifiable information and the account linking is possible.

Keywords: SNS, Pricacy, Personal Information, K-Anonymity, Re-identification

I. 서 론

인터넷에는 많은 개인정보가 노출되어 있다. 게시판, 카페, 블로그 등에 자신의 신상정보와 관심분야, 정치성향, 동태정보가 포함된 글을 올리는 이용자가 많다. SNS 이용이 활성화되면서 더욱 많은 이용자들이 더욱 빈번히 자신의 정보를 게시하고 있다. 그런데 이용자들은 자신이 공개한 범위가 누구에게 노출되는지 잘 알지 못하는 경우가 많다. SNS의 경우 친구들만 볼 수 있다고 생각하고 공개한 정보가 전체 이용자들에게 노출되는 경우를 예로 들 수 있다.

현재 개인정보보호를 위해 주민번호, 계좌번호, 전화번호 등과 같이 그 자체로 개인식별에 사용될 수 있는 식별정보의 노출을 방지하도록 하고 있다. 이를 위해 PC용 파일스캐너 제품과 웹사이트의 식별정보 노출을 탐지하는 프라이버시 스캐너 제품들[2][3]이 나와 있다. 그런데, 이러한 식별 정보 외에도 이름, 거주

접수일(2013년 9월 11일), 게재확정일(2013년 9월 21일)
† 주저자, sunchoi@etri.re.kr
‡ 교신저자, sunchoi@etri.re.kr(Corresponding author)

〔표 1〕 페이스북 개인정보 노출 현황

개인정보	노출계정수	비율	개인정보	노출계정수	비율
이름	6,575,571	100.0%	스포츠 팀	359,812	5.5%
성별	6,059,339	92.1%	활동	357,078	5.4%
고등학교	3,139,450	47.7%	인용구	294,686	4.5%
혈액형	2,686,130	40.9%	직책/직위	256,027	3.9%
대학교	2,335,233	35.5%	대학원	199,508	3.0%
직장/직업	1,624,908	24.7%	웹사이트	117,819	1.8%
관심사	1,299,364	19.8%	구사 언어	52,938	0.8%
음악	933,056	14.2%	종교관	43,635	0.7%
TV프로그램	574,500	8.7%	전화번호	41,900	0.6%
영화	558,446	8.5%	이메일	24,469	0.4%
책	467,490	7.1%	대화명	22,296	0.3%
게임	457,165	7.0%	정치관	17,548	0.3%
스포츠 선수	447,492	6.8%	주소	12,834	0.2%
스포츠	365,782	5.6%			

지, 나이, 직업, 직장, 학교 등의 비식별 정보도 개별적으로 혹은 여러 개의 정보를 조합하면 개인을 특정하고 식별할 수 있는 경우가 많다. 또한, 한 이용자의 여러 서비스 계정을 서로 연결할 수 있는데 이를 통해 한 개인에 대해 더 많은 정보를 획득할 수 있으며, 계정을 연결할 수 있다는 것은 개인이 특정되고 식별되었다는 것을 의미한다.

본 논문에서는 실제 SNS 계정들을 조사하여 노출된 개인정보의 종류와 개수를 파악하였고, 노출된 정보를 통한 이용자 특정 가능성과 서비스 간 계정 연결 가능성을 분석하였다. II장에서 조사 방법 및 결과를 제시하고, III장에서는 조사 결과로 파악된 문제점 및 대응 방안에 대해 기술한 뒤, IV장에서 결론을 맺는다.

II. 조사 내용 및 결과

2.1 데이터 수집 및 기본 수치

페이스북 한국인 이용자의 계정 657만 개와 트위터 한국인 이용자 계정 277만 개의 개인정보 노출 현황을 조사하였다(이 계정들이 한국인 이용자 계정의 전부는 아니다). 한국인 이용자를 구분하는 방법은 이름 필드가 한글로 되어 있거나, 영문이름이지만 한글 성을 갖는 경우로 한정하였다. 페이스북은 http://facebook.com/userid로 표시되는 페이지에 노출된 정보를 조사하였다. 노출된 정보는 당연히 이용자가 페이스북 이용자들에게 공개한 정보에 한정되었다. 이 페이지의 정보는 필드 별로 구분되어있다. 그 중 프로파일 필드의 경우는 이용자가 자유롭게 입력한 정보를 담고 있으므로 여러 종류의 정보가 포함되어 있다. 본 조사에서는 프로파일 필드 외에 비교적 정형화된 다른 필드 정보만을 활용하였다. 〔표 1〕은 페이스북 계정에 노출된 개인정보 현황을 보여준다.

이름은 모든 계정에서 필수 공개로 되어 있는 항목이다. 92%의 계정이 성별을 노출하고, 이어 학교, 혈액형, 직장/직업 등의 신상 정보가 노출된 비율이 높았다. 관심사, 좋아하는 음악 등이 노출된 비율은 10% 이상이었다. 이러한 정보를 공개한 것은 이용자의 선택인데, 모든 이용자가 자신의 정보가 모든 페이스북 이용자에게 노출된다는 사실을 알고 공개한 것이라고 보기는 어렵다. 예를 들어 관심사를 공개한 129만 명 모두가 자신의 관심사를 모든 페이스북 이용자에게 의도적으로 공개했다고 할 수는 없을 것이다. 그 외의 TV, 영화, 게임, 스포츠 등이 노출된 비율은 10% 미만이었다. 직책/직위를 공개한 비율은 직장/직업을 공개한 비율보다 많이 낮았다.

종교관, 정치관 등이 공개된 비율은 1% 미만이었다. 이메일, 전화번호, 주소 같이 종래에 개인정보로 생각되던 식별 정보를 노출한 계정의 비율은 0.5% 미만 이었다. 식별정보가 적게 노출되어 있으므로 SNS의 개인정보 노출은 심각하지 않은 것으로 볼 수도 있으나, 식별정보가 아닌 비식별정보로도 개인을 특정할 수 있음을 다음 절에서 밝힌다.

트위터는 이용자가 공개한 프로파일 정보와 위치 정보를 분석했다. 프로파일 정보는 프리텍스트로 되어 있으며 여러 가지 정보가 포함되어 있다. 그 중에서 이름, 나이, 학교, 직업, 직위, 이메일, 전화번호 정보를 추출하였다. 이메일, 전화번호 같은 정형 정보는

정규식[1]을 사용해서 추출이 가능하며, 기존의 개인정보 스캐닝/필터링 제품[2][3]들도, 주민번호와 계좌번호를 포함한 이러한 정형 개인정보를 탐지할 수 있으나 이름, 나이, 학교, 직업, 직위, 등의 비정형 개인정보는 추출이나 탐지가 불가능하다. 본 연구에서는 나이, 학교는 패턴 규칙을 사용해서 추출하였고, 이름, 직업, 직위는 ETRI에서 개발한 개체명 인식기[4]를 사용하여 추출하였다. 개체명 인식기의 경우 미인식과 오인식이 있으므로 이를 이용해 추출한 개인정보에는 오류가 포함될 수 있으며, 모든 노출 정보가 포함되지 않을 수 있다. 트위터 프로파일에는 상기 정보 이외에도 다양한 개인정보가 포함되어 있으나 추출한 정보는 상기 정보에 한정하였다. 트위터 프로파일 및 위치정보 필드에 노출된 개인정보 현황은 [표 2]와 같다.

[표 2] 트위터 개인정보 노출 현황

개인정보	노출 계정 수	비율
이름	1,929,407	69.4%
지역	1,252,289	45.2%
직업	933,056	33.7%
학교	558,446	20.2%
직위	457,165	16.5%
나이	92,291	3.3%
전화번호	5,960	0.2%
이메일	2,376	0.1%

트위터에서 이름이 노출된 계정은 69%이고, 지역(위치)가 노출된 경우는 45%이었다. 직업이 노출된 경우는 33%, 학교 정보는 20%의 계정에서 노출되었다. 트위터에서도 전화번호, 이메일 같은 식별정보가 노출된 경우는 1% 미만이었다. 프로파일 프리텍스트 분석은 미인식된 경우가 많이 있기 때문에 분석기술이 고도화되면 더욱 많은 정보 노출을 확인할 수 있을 것이다.

2.2 특정 가능성

SNS에 노출된 정보에는 전화번호, 이메일 등 직접 식별에 사용될 수 있는 정보도 있다. 이러한 정보들은 종래에도 개인정보로 분류되어 보호 대상으로 하고 있다. 한편, 이러한 식별 가능 정보 이외에도 어떤 개인정보 값은 조사 대상 계정들 가운데서 그 값을 갖는 계정이 한 개만 존재하는 경우가 있다. 예를 들어 페이스북 이용자 중에서 이름의 '김길동'인 계정이 1개만 존재하는 경우다. 이것은 k-anonymity[5]에서 k=1인 경우를 의미한다. 이를 통해 이용자를 특정할 수 있다. 즉 이름이 모두 식별정보는 아니지만, 특정한 이름의 경우는 이름을 통해 식별이 되는 경우가 있는 것이다. 이렇게 어떤 값을 갖는 계정 수가 몇 개인지 정확히 알기 위해서는 각 필드 값을 정규화해야 한다. 이름의 경우도 '김길동', '김 길동', 'Kil-dong Kim', 'Gildong Kim'을 모두 '김길동'으로 정규화해야만 정확한 산정이 가능하다. 이름의 경우는 영어-한글 이름 변환기[6]를 사용하였고, 다른 필드의 경우, 개체명 사전과 규칙을 이용하여 정규화 하였다.

이렇게 개인정보마다 유일 값을 통해 특정할 수 있는 사람 수가 [표 3]에 나타나 있다. 이름의 경우는 207,027개의 유일 값을 가지며, 전체 노출된 이름 중 3.1%가 유일하다. 모든 계정이 이름을 포함하고 있으므로, 이름으로 특정되는 계정의 비율 또한 3.1% 이다. 전화번호와 이메일은 종래에도 식별정보로 분류되어 온 것처럼 유일한 값의 비율이 100%에 육박한다. 하지만 노출된 계정의 수가 적으므로 이를 통해 특정되는 계정의 비율은 1% 미만이다. 직책/직위, 성별, 혈액형은 유일 값을 갖는 경우가 전혀 없었다. 직책명, 성별, 혈액형 등을 유일하게 갖는 사람은 없을 것이기에 이 결과는 당연한 것이다. 한편, 대학교의 경우 국내의 350개의 대학만을 인식하였다. 이때 유일한 값을 갖는 계정은 없었다. 이는 350개의 대학교 중 이를 명시한 이용자가 1명뿐인 학교는 없었다는 의미이다. 고등학교의 경우 36 개 학교는 해당 학교를 명시한 이용자가 1명뿐인 경우가 있었다. 이를 통해 36명을 특정할 수 있다는 의미이다. 트위터의 경우 이름을 통해 특정되는 이용자를 조사하였는데 145,040명이 특정되었다. 이름 하나의 필드 값만으로 두 SNS

[표 3] 페이스북 개인정보로 특정되는 사람 수

개인정보	노출 계정 수	유일 값 개수	유일 값 비율	특정 비율
이름	6,575,571	207,027	3.1%	3.1%
전화번호	41,900	41,821	99%	0.6%
이메일	24,469	24,469	100%	0.4%
주소	12,834	8,499	66%	0.1%
별명	22,296	21,469	96%	0.3%
직책/직위	256,027	0	0.0%	0.0%
성별	6,059,339	0	0.0%	0.0%
혈액형	2,686,130	0	0.0%	0.0%
대학교	2,335,233	0	0.0%	0.0%
고등학교	3,139,450	36	0.0%	0.0%

〔표 4〕 페이스북 개인정보 (조합)으로 특정되는 사람 수

개인정보 조합	유일 값 수	특정 비율
이름-고등학교	2,262,410	34.4%
이름-대학교	1,169,170	17.7%
이름-고등학교-대학교	2,975,399	45.2%

서비스에서 특정할 수 있는 개인이 35만 명에 달한다.

두 개 이상의 필드를 조합한 값의 경우는 그 값의 조합이 유일한 경우가 훨씬 많다. 이름과 대학교, 고등학교를 조합하여 특정할 수 있는 사람의 수를 조사한 결과가 〔표 4〕에 표시되어 있다.

이름-고등학교 정보 조합을 이용해 특정할 수 있는 사람의 수는 226만 명이었으며 이는 전체 이용자중 34.4%에 달한다. 이름-대학교 조합은 고등학교 보다 특정성이 떨어지는데 이는 특정 대학교를 명시한 동명이인의 수가 특정 고등학교를 명시한 동명이인의 수보다 적다는 걸 의미한다. 이름-고등학교-대학교 조합의 경우 297만 명을 특정할 수 있어 매우 높은 특정성을 보였다. 여기에 성별, 혈액형 등의 정보를 추가하면 특정할 수 있는 사람 수는 증가할 수 있다.

2.3 ID 연결 가능성

하나의 서비스에 노출된 단편적 개인정보들을 조합하는 것으로 개인을 특정할 가능성은 높아진다. 정보가 많을수록 특정가능성이 높아지는데 여러 서비스에 노출된 개인정보를 취합하면 정보가 더욱 늘어나게 된다. 여러 서비스에 노출된 이용자 정보의 취합을 위해서는 서비스 간의 계정 매핑이 필요하다. 가령, 페이스북에서 kimcs라는 id를 사용하는 이용자가 트위터에서 cskim이라는 id를 사용하는 것을 안다면, 두 서비스에 노출된 개인정보들을 모두 취합할 수 있을 것이다. 한편, 서비스 간 계정을 연결할 수 있다는 것은 서비스 내에서 이용자를 식별하였다는 의미에서 익명화된 사용자를 다시 식별했다는 의미인 re-identification의 한 형태라고 볼 수 있다〔12〕.

본 조사를 통해 수집한 트위터 계정 중에 29,727개는 페이스북 계정을 직접 명시해 놓고 있다. 이 경우는 확실한 계정 연결이 가능하다. 트위터 계정과 페이스북 계정의 identifier가 서로 일치하는 계정 수가 67,175개였다. Identifier의 뒷부분만 다른 경우를 포함하면 93,835개였다. Identifier가 일치한다고 해서 실제로 모든 계정이 일치하는 것은 아니다. 한편, 이름이 양쪽 서비스에서 같은 경우는 몇 가지로 나누어진다. 양쪽 서비스에 그 이름이 1개씩만 있는 경우는 43,310개였다. 이름이 1개씩 만 있는 경우에도 실제로 모든 계정이 동일 이용자의 것은 아니다. 이름이 한쪽 서비스에는 1개만 있으나, 다른 쪽에는 같은 이름을 갖는 계정이 여러 개 있는 경우는 84,978개였다. 이 경우는 다른 서비스에 같은 이름을 갖는 여러 개의 계정 중에 동일인이 있을 수도 있고 없을 수도 있다. 양쪽 다 여러 개의 계정이 존재하는 공통 이름의 개수는 119,086개였다. 이때는 한쪽의 하나의 계정을 하나씩 상대편 서비스의 여러 계정에 대해 일치 여부를 테스트해야 한다. 이러한 후보 쌍의 합계는 34만개이다. 100개의 샘플을 수작업으로 확인한 결과 50% 정도의 일치도를 보였으므로, 상기 후보 쌍 중에 실제 동일인으로 분석되는 경우는 절반인 17만 쌍이 될 것으로 추정할 수 있다. Identifier와 이름 이외에도, 다른 정보를 통해 연결할 수 있는 계정이 있으므로, 실제 찾을 수 있는 계정 연결 쌍은 17만 개 보다 많을 것으로 보인다.

III. 조사 결과 분석

3.1 현황 및 문제점

본 조사 결과로부터 알 수 있는 SNS 개인정보 노출 현황은 다음과 같이 정리할 수 있다.

첫째, SNS를 통한 개인정보 노출이 상당히 많다는 것이다. 이렇게 많은 개인정보가 모두 이용자의 의도에 부합되어 공개된 것이라고 보기는 어렵다. 미국 컨슈머리포트의 조사에 따르면 1,300만 명의 미국인 페이스북 이용자는 프라이버시 설정을 아예 모르거나 이용하지 않고 있다고 한다〔7〕. 따라서 본 조사에서 정보가 노출된 계정의 모든 이용자가 자신이 공개한 정보가 이렇게 모든 이용자에게 노출될 수 있다는 것을 알고 있다고 보기는 어렵다.

둘째, 식별정보는 많이 노출되어 있지 않지만, 여전히 많은 식별이 가능하다. 비식별 정보로 분류되어 오던 정보들도 개별적으로 혹은 타 정보와의 조합을 통해 개인을 특정할 수 있는 경우가 상당히 많음을 알 수 있었다. 물론, 본 조사가 모든 한국인 페이스북 이용자들을 커버하는 것이 아니며, 모든 한국인이 페이스북 이용자가 아니기 때문에, 모든 한국인을 대상으로 한다면, 특정되는 이용자의 숫자가 많이 줄어들 수도 있다. 그러나 상당히 많은 이용자는 여전히 식별 가능하며, 더 많은 단편 정보가 수집될수록 식별 가능

한 숫자가 더욱 증가할 것이다.

셋째, 계정의 연결 가능성도 상당히 높은 것으로 파악되었다. 이용자가 명시한 URL을 통해 많은 수의 계정 연결이 되었으며, 기본적인 단서인 id, 이름만으로도 많은 수의 동일 계정 후보를 찾아 낼 수 있었다. 실제로 동일 계정인지 판단하는 기술[8]을 적용하면, 후보들 중에 실제 동일인을 찾아 낼 수 있다.

이상을 종합하면 SNS 개인정보 노출 문제의 본질은 이용자의 의도와 다르게 개인식별이 될 위험이 있다는 것이다. 이용자는 자신의 정보를 공개할 때 이로 인해 자신이 특정될 수 있는 가능성을 전혀 알 수 없다. 그 이유는 특정 가능성은 자신이 공개한 정보에 의해 결정되는 것이 아니라 다른 이용자들이 공개한 정보에 따라 결정되기 때문이다. 한편, 자신이 공개한 정보의 조합을 통해 자신의 특정 가능성을 추정해 볼 수 있으나, 이를 위해서는 각 서비스에 공개했던 자신의 정보를 모두 기억해야만 하는데 이는 어려운 것으로 보인다.

3.2 관련 연구

한국인터넷진흥원은 국내 트위터 이용자 계정 200개를 대상으로 개인정보 노출 현황을 조사하여 상당수의 계정에서 개인정보가 공개되고 있음을 확인했다[9]. 이름(88%), 인맥정보(86%), 사진 등 외모정보(84%), 위치정보(83%), 관심분야 등 취미정보(64%), 스케쥴 정보(63%), 가족 정보(52%) 등이 노출되어 있었으며 의료정보(29%), 정치성향 정보(19%) 등 민감 정보가 노출된 비율도 높았다. 여기서는 프로파일뿐 아니라 타임라인도 분석하였는데, 수작업으로 소량의 샘플 데이터를 분석한 점이 본 연구와 다르다. [10]은 다양한 매체 및 데이터 분석을 통해 정보 제공자가 의도하지 않은 사생활 침해가 가능한 것을 보였다.

[11]에서는 SNS에 노출된 개인정보 취약점 4가지로 1) 개인정보를 누구나 쉽게 접근할 수 있다. 2)동일 사용자의 정보를 다른 사이트에서 수집/조합하여 더 많은 정보가 수집될 수 있다. 3) 공개 또는 수집된 정보를 추론하여 숨겨진 정보를 유추할 수 있다. 4) 익명화된 데이터에서 사용자의 id가 식별될 수 있다는 점을 들고 있다. 또한, 본 연구와 유사하게 실제 SNS 이용자 데이터를 통한 특정 가능성 및 id 연결 가능성을 조사하였다. [12]에서는 서로 다른 서비스의 계정을 연결하는 연구 결과를 소개하였다. 계정의 속성값 즉 개인정보 들을 이용하거나 친구관계를 이용하여 계정 간의 일치 여부를 판정하고 있다.

한편, 정부에서는 공공 데이터를 공개할 때 식별정보를 제거하도록 하는 가이드라인[13]을 발표하였다. 이 가이드라인에서는 비식별 정보 조합으로 인한 개인 재식별 가능성을 제거하도록 요구하고 있다.

3.3 대응방안

이용자의 의도와 다르게 개인식별이 될 위험에 대응하기 위해서는 노출된 정보를 모두 수집하여 이용자 특정 가능성을 분석해야 한다. 특정 가능한 경우 특정에 사용된 정보를 삭제하거나 일부를 마스킹하면 특정 가능성을 제거할 수 있다. 이는 기존에 노출된 정보로 인한 위험 이외에도 추가로 정보를 공개할 때도 해당된다. 추가 노출 정보를 더해 특정 가능성을 평가하면, 그 정보를 공개해도 괜찮은 것인지 아닌지 판단할 수 있고, 이 판단에 따라 필터링을 수행하면 된다. 한편, [14]에서 밝힌 것처럼 노출을 이용해 추가적인 정보를 추론할 수 있으며 추론된 정보들을 노출 정보에 더해 이용자 특정에 사용할 수 있다. 이러한 위험에 대응하기 위해서는 노출 정보 위험 분석 이전에 가능한 추론을 실시해서 추론 가능한 정보를 모두 도출해야 한다. 이러한 위험도 분석 과정은 네트워크 취약성을 분석하기 위해 실제 공격을 실시해 보는 과정과 유사하다고 할 수 있다. [15]에서는 이러한 위험 분석 연구를 소개하고 있다.

IV. 결론

본 논문에서는 페이스북 657만 개와 트위터 277만 개의 한국인 이용자 계정을 조사하여 개인정보 노출 현황을 분석하였다. 바로 식별에 이용될 수 있는 식별정보는 1% 미만이 노출되어 있으나, 기존에 비식별 정보라고 생각되던 정보로 개인을 특정할 수 있는 경우가 3% 이상이고 다른 정보와 조합을 통해 개인을 특정할 수 있는 경우가 45%에 달하였다. 또한 34만 개의 페이스북 계정과 트위터 계정 연결 가능 후보를 도출하여 각각의 개인정보를 결합할 수 있는 가능성을 확인했다. SNS 개인정보 노출은 심각한 문제이며, 다른 인터넷 서비스를 추가로 분석하면 위험성이 더욱 커질 수 있다. 이에 대응하기 위해서 기존의 식별정보 이외에 비식별 정보의 조합 및 계정연결로 인한 프라이버시 위험에 대한 고려와 대비가 필요하다.

참고문헌

〔1〕 http://en.wikipedia.org/wiki/Regular_expression- Regular expression

〔2〕 프라이버시스캐너, http://wdigm.com

〔3〕 SafePrivacy, http://www.nicstech.com/new/sub_02_01_03.html

〔4〕 이창기, 장명길, "Structural SVMs 및 Pegasos 알고리즘을 이용한 한국어 개체명 인식," 인지과학, 21(4), pp.665-667, 2010년 12월

〔5〕 Latanya sweeney, "k-Anonymity : A Model For Protecting Privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, Oct. 2002.

〔6〕 김석현, 최대선, 진승헌, "비정형 사용자 이름의 정형화된 한글 이름 변환 방법 연구," 정보과학회 2013 추계학술발표회논문집, 게재예정, 2013년 11월

〔7〕 "Facebook & your privacy Who sees the data you share on the biggest social network," http://www.consumerreports.org/cro/magazine/2012/06/facebook-your-privacy/index.htm, Jun. 2012.

〔8〕 박준범, 최대선, 진승헌, "페이스북과 트위터 이용자 계정 연결 방법," 정보과학회 2013 추계학술발표회논문집, 게재예정, 2013년 11월

〔9〕 방송통신위원회, "트위터를 통한 개인정보 유형별 노출 현황,"http://old.kcc.go.kr/user.do?mode=view&page=P05030000&dc=K04030000&boardId=1042&boardSeq=30643, 2011년 1월

〔10〕 정영수, "Big Data 시대의 프라이버시 보호," NIA Privacy Issues, 제7호, 2012년12월

〔11〕 Yuhao Yang, Jonathan Lutes, Fengjun Li, Bo Luo, Peng Liu, "Stalking Online: on User Privacy in Social Networks," CODASPY '12 Proceedings of the second ACM conference on Data and Application Security and Privacy, pp. 37-48, Feb. 2012.

〔12〕 A. Narayanan, V. Shimatikov. "Deanonymizing Social Networks," Proceedings of the 30th IEEE Symposium on Security and Privacy, pp. 173-187, May. 2009.

〔13〕 안전행정부, "공공정보 개방 공유에 따른 개인정보 보호 지침," http://www.mospa.go.kr, 2013년 9월

〔14〕 M. Kosinski, et.al. "Private traits and attributes are predictable from digital records of human behavior," Proceedings of the National Academy of Sciences of the United States of America, vol. 110, no.15, pp. 5802-5805, Mar. 2013.

〔15〕 최대선, 김석현, 조진만, 진승헌, "빅데이터 개인정보 위험 분석 기술," 정보보호학회지, 제13(2), 2013년 6월

〈저자소개〉

최 대 선 (Daeseon Choi) 정회원
1995년: 동국대학교 컴퓨터공학과 학사
1997년: 포항공과대학교 컴퓨터공학과 석사
2009년: 한국과학기술원 전산학과 박사
1997년~1999년: 현대전자/현대정보기술 연구소 선임
1999년~현재: 한국전자통신연구원 책임연구원
〈관심분야〉 인증, 개인정보보호, 빅데이터 분석

김 석 현 (Kim, Seok Hyun) 정회원
2008년: 충주대학교 전자통신학과 학사
2010년: 전남대학교 정보보호협동과정 석사
2010년~현재: 한국전자통신연구원 선임연구원
〈관심분야〉 통신공학, 정보보호, Social Network Security

조 진 만 (CHO, JIN-MAN) 종신회원
1989년: 충남대학교 계산통계학과 학사
1991년: 충남대학교 전자계산학과 석사
1991년~현재: 한국전자통신연구원 책임연구원
〈관심분야〉 개인정보보호, 스마트카드

진 승 헌 (Seung-Hun Jin) 종신회원
1993년: 숭실대학교 전자계산학과 학사
1995년: 숭실대학교 전자계산학과 석사
2004년: 충남대학교 컴퓨터과학과 박사
1994년~1996년: 대우통신
1996년~1999년: 삼성전자
1999년~현재: 한국전자통신연구원 인증기술연구실장/책임연구원
〈관심분야〉 컴퓨터/네트워크 보안, PKI, ID 관리, 개인정보보호, 모바일 지불결제 보안

조 현 숙 (Hyun Sook Cho) 종신회원
1979년: 전남대학교 수학교육과 학사
1989년: 충북대학교 컴퓨터과학과 석사
2001년: 충북대학교 컴퓨터학과 박사
1982년~현재: 한국전자통신연구원 사이버보안연구단 단장/책임연구원
〈관심분야〉 암호학, 보안 프로토콜, 네트워크 보안