(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년07월12일
(11) 등록번호 10-2418984
(24) 등록일자 2022년07월05일

(51) 국제특허분류(Int. Cl.)
G06F 21/62 (2013.01) *G06F 16/907* (2019.01)
G06F 16/9535 (2019.01)
(52) CPC특허분류
G06F 21/6254 (2013.01)
G06F 16/907 (2019.01)
(21) 출원번호 10-2020-0159176
(22) 출원일자 2020년11월24일
심사청구일자 2020년11월24일
(65) 공개번호 10-2022-0072113
(43) 공개일자 2022년06월02일
(56) 선행기술조사문헌
김양호 외 2인, '메타데이터를 활용한 개인정보 처리에 대한 의사결정 모델', Journal of The Korea Institute of Information Security & Cryptology, VOL.26, NO.1, Feb. 2016.
이창범, '가명정보에 있어서 "다른 정보"와 "추가 정보"의 차이 및 가명처리의 대상과 범위', 2020 KISA REPORT, VOL.5, 2020.05. pp.42-51.
'개인정보의 비식별 처리', 미국 국립표준 기술연구소, NISTIR 8053, 2015.10.

(73) 특허권자
김수정
서울특별시 은평구 진관1로 21-9, 111동 501호 (진관동, 은평뉴타운 박석고개)
(주)큐브더모먼트
서울특별시 강서구 금낭화로24길 29-5 ,1층(방화동)
(72) 발명자
김수정
서울특별시 은평구 진관1로 21-9, 111동 501호 (진관동, 은평뉴타운 박석고개)
조남열
서울특별시 강서구 강서로45다길 27, 301호 (화곡동)
(74) 대리인
최광석

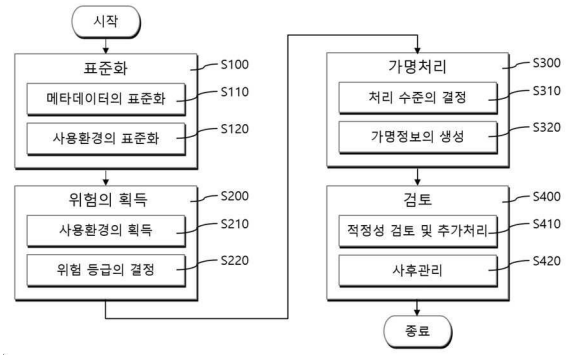
전체 청구항 수 : 총 4 항 심사관 : 구대성

(54) 발명의 명칭 **사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템 및 그 제어방법**

(57) 요 약

본 발명의 실시예는 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법에 있어서, 원천 정보집합물을 획득하는 과정; 상기 원천 정보집합물의 변수를 획득하는 과정; 상기 원천 정보집합물의 변수에 대한 대표명칭을 확인하는 과정; 상기 원천 정보집합물의 변수에 대한 개인정보의 유형을 확인하는 과정; 상기 원천 정보집합물의 개인정보의 유형에 대한 가명처리 수준을 확인하는 과정; 및 상기 원천 정보집합물의 변수를 상기 대표명칭으로 변경하기 위한 개인정보의 메타데이터를 획득하는 과정을 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 제공할 수 있다.

대 표 도 - 도3



(52) CPC특허분류
G06F 16/9535 (2019.01)

명 세 서

청구범위

청구항 1

사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법에 있어서,

원천 정보집합물을 획득하는 과정;

상기 원천 정보집합물의 변수를 획득하는 과정;

상기 원천 정보집합물의 변수에 대한 대표명칭을 확인하는 과정;

상기 원천 정보집합물의 변수에 대한 개인정보유형을 확인하는 과정;

상기 개인정보유형 별로 결정되는 가명처리 기법을 포함하는 가명처리 수준에 대한 정보를 확인하는 과정;

상기 원천 정보집합물의 변수를 상기 대표명칭으로 변경하기 위한 개인정보의 메타데이터를 획득하는 과정;

상기 원천 정보집합물의 변수의 데이터 값을 획득하는 과정;

상기 개인정보의 메타데이터를 이용하여 상기 변수의 데이터 값에 기초해 상기 원천 정보집합물의 변수에 대한 대표명칭을 획득하는 과정; 및

상기 원천 정보집합물의 변수가 상기 대표명칭으로 변경된 정보집합물을 획득하는 과정을 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법.

청구항 2

제1항에 있어서,

상기 원천 정보집합물의 변수에 대한 대표명칭이 없는 경우, 상기 원천 정보집합물의 변수에 대한 새로운 대표명칭을 생성하는 과정을 더 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법.

청구항 3

제1항에 있어서,

상기 원천 정보집합물의 변수에 대한 개인정보유형이 없는 경우, 상기 원천 정보집합물의 변수에 대한 새로운 개인정보유형을 생성하는 과정을 더 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법.

청구항 4

제1항에 있어서,

상기 가명처리 수준에 대한 정보 내에 특정한 개인정보유형에 대한 가명처리 수준에 대한 정보가 없는 경우, 상기 특정한 개인정보유형에 대한 새로운 가명처리 수준에 대한 정보를 생성하는 과정을 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법.

청구항 5

삭제

발명의 설명

기 술 분 야

[0001] 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템 및 그 제어방법에 관한 것으로, 자세히는 개인 정보의 오남용을 차단하기 위하여 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템 및 그 제어방법에 관한 것이다.

배 경 기 술

[0003] 최근 들어, 데이터 3법의 통과로 개인정보보호법에 처음으로 개인정보와 익명정보의 중간 개념인 가명정보의 개념이 도입되었다. 특정 정보주체를 식별할 수 있는 정보는 개인정보, 식별할 수 없는 것은 익명정보, 그리고 추가 정보 없이는 특정 정보주체를 식별할 수 없도록 가명처리된 정보가 가명정보이다.

[0004] 가명정보는 통계작성, 과학적 연구, 공익적 기록 보존 등과 같은 일정한 목적을 위한 처리와 그 정보 결합을 허용했으며, 가명정보는 정보주체의 동의 없이 활용이 가능하다. 가명정보는 가명처리함으로써 원래의 상태로 복원하기 위한 추가정보의 사용·결합 없이 특정 정보주체를 알아볼 수 없는 정보이다.

[0005] 그러나, 가명처리의 취급자별 주관적 판단에 의해 가명처리 기준을 설정하기 때문에 취급자별로 다른 가명정보가 도출되는 문제점을 가진다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 대한민국 공개특허공보 제10-2016-0108993호 (2016.09.21.)

발명의 내용

해결하려는 과제

[0008] 본 발명은 개인정보의 노출 위험에 따른 가명처리 수준을 표준화하고, 개인정보의 노출 위험을 정량적으로 측정하고, 위험에 맞는 가명처리 수준을 결정하고 가명처리함으로써, 신뢰성 있는 가명정보를 제공할 수 있다.

[0009] 본 발명은 정보집합물의 취급자의 전문 지식과 상관없이 재식별 위험의 측정 및 가명처리 수준의 결정을 수행함으로써, 취급자에 따라 동일한 가명정보의 결과를 도출할 수 있다.

[0010] 본 발명은 가명정보에서 특정 정보주체의 식별 위험을 평가할 수 있는 방안을 제시함으로써, 정보주체의 프라이버시를 보고하고 신뢰성 있는 데이터 활용 플랫폼을 제공할 수 있다.

[0011] 본 발명은 개인정보를 포함하는 정보집합물에 대해 동일한 사용환경에 대해 동일한 가명정보의 결과를 도출할 수 있는 정보집합물을 가명 처리하는 시스템 및 그 제어방법을 제공할 수 있다.

[0012] 본 발명은 개인정보의 오남용을 차단하고 신뢰성 있는 가명정보를 제공할 수 있다.

과제의 해결 수단

[0013] 본 발명의 실시예는 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법에 있어서, 원천 정보집합물을 획득하는 과정; 상기 원천 정보집합물의 변수를 획득하는 과정; 상기 원천 정보집합물의 변수에 대한 대표명칭을 확인하는 과정; 상기 원천 정보집합물의 변수에 대한 개인정보의 유형을 확인하는 과정; 상기 원천 정보집합물의 개인정보의 유형에 대한 가명처리 수준을 확인하는 과정; 및 상기 원천 정보집합물의 변수를 상기 대표명칭으로 변경하기 위한 개인정보의 메타데이터를 획득하는 과정을 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 제공할 수 있다.

[0014] 본 발명의 실시예는, 상기 원천 정보집합물의 변수에 대한 대표명칭이 없는 경우, 상기 원천 정보집합물의 변수에 대한 새로운 대표명칭을 생성하는 과정을 더 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처

리하는 시스템의 제어방법을 제공할 수 있다.

[0015] 본 발명의 실시예는, 상기 원천 정보집합물의 변수에 대한 개인정보의 유형이 없는 경우, 상기 원천 정보집합물의 변수에 대한 개인정보의 새로운 유형을 생성하는 과정을 더 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 제공할 수 있다.

[0016] 본 발명의 실시예는, 상기 원천 정보집합물의 개인정보의 유형에 대한 가명처리 수준이 없는 경우, 상기 원천 정보집합물의 개인정보의 유형에 대한 새로운 가명처리 수준을 생성하는 과정을 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 제공할 수 있다.

[0017] 본 발명의 실시예는, 상기 원천 정보집합물의 데이터의 실제 값을 획득하는 과정; 상기 데이터의 실제 값에 기초하여 상기 데이터의 변수의 대표명칭을 획득하는 과정; 및 상기 데이터의 변수를 상기 대표명칭으로 변경하는 과정을 더 포함하는, 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 제공할 수 있다.

발명의 효과

[0018] 본 발명은 개인정보의 노출 위험에 따른 가명처리 수준을 표준화하고, 개인정보의 노출 위험을 정량적으로 측정하고, 위험에 맞는 가명처리 수준을 결정하고 가명처리 함으로써, 신뢰성 있는 가명정보를 제공하는 효과를 가진다.

[0019] 본 발명은 정보집합물의 취급자의 전문 지식과 상관없이 재식별 위험의 측정 및 가명처리 수준의 결정을 수행함으로써, 취급자에 따라 동일한 가명정보의 결과를 도출하는 효과를 가진다.

[0020] 본 발명은 가명정보에서 특정 정보주체의 식별 위험을 평가할 수 있는 방안을 제시함으로써, 정보주체의 프라이버시를 보고하고 신뢰성 있는 데이터 활용 플랫폼을 제공하는 효과를 가진다.

[0021] 본 발명은 개인정보를 포함하는 정보집합물에 대해 동일한 사용환경에 대해 동일한 가명정보의 결과를 도출할 수 있는 정보집합물을 가명 처리하는 시스템 및 그 제어방법을 제공하는 효과를 가진다.

[0022] 본 발명은 개인정보의 오남용을 차단하고 신뢰성 있는 가명정보를 제공하는 효과를 가진다.

도면의 간단한 설명

[0024] 도 1은 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 블록도를 도시한 것이다.

도 2는 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 상세 블록도를 도시한 것이다.

도 3은 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 도시한 것이다.

도 4는 본 발명의 실시예에 따른 메타데이터를 표준화하는 방법의 흐름도를 도시한 것이다.

도 5는 본 발명의 다른 실시예에 따른 메타데이터를 표준화하는 방법의 흐름도를 도시한 것이다.

발명을 실시하기 위한 구체적인 내용

[0025] 본 발명의 개념에 따른 실시 예들은 다양한 변경들을 가할 수 있고 여러 가지 형태들을 가질 수 있으므로 실시예들을 도면에 예시하고 본 명세서에서 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 개념에 따른 실시예들을 특정한 개시 형태들에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물, 또는 대체물을 포함한다.

[0026] 본 명세서에서 사용한 기술적 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로서, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 본 명세서에 기재된 특징, 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성 요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0027] 본 명세서에서 사용되는 구성요소에 대한 접미사 "모듈(module)" 및 "부"는 명세서 작성의 용이함만이 고려되어 부여되거나 혼용되는 것으로서, 그 자체로 서로 구별되는 의미 또는 역할을 갖는 것은 아니며, 본 발명의 실시예에 따른 방법을 수행하기 위한 하드웨어 또는 상기 하드웨어를 구동할 수 있는 소프트웨어의 기능적 또는 구조적 결합을 의미할 수 있다.

[0028] 이하에서 첨부된 도면을 참고하여, 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명처리하는 시스템 및 그 제어방법에 대해서 설명한다.

[0029] 도 1은 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 블록도를 도시한 것이다.

[0030] 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템은 전자 장치(1000)일 수 있다. 전자 장치(1000)는, 단말기, 디바이스, 전자기기 등으로 호칭될 수 있다. 전자 장치(1000)는, 스마트폰, 태블릿 PC, PC, 스마트 TV, 휴대폰, PDA(personal digital assistant), 랩톱, 미디어 플레이어, 서버, 마이크로 서버, GPS(global positioning system) 장치, 전자책 단말기, 디지털방송용 단말기, 네비게이션, 키오스크, MP3 플레이어, 디지털 카메라, 가전기기 및 기타 컴퓨팅 장치일 수 있으나, 이에 한정되는 것은 아니다. 또한, 이에 한정되지 않으며, 전자 장치(100)는 데이터를 처리하고, 처리된 데이터를 제공할 수 있는 모든 종류의 기기를 포함할 수 있다.

[0031] 도 1에 도시된 바와 같이, 일 실시 예에 따른 전자 장치(1000)는, 저장부(1100), 출력부(1200), 제어부(1300), 통신부(1500), 및 사용자 입력부(1700)를 포함할 수도 있다. 도시된 구성 요소 모두가 전자 장치(1000)의 필수 구성 요소인 것은 아니며, 보다 많은 구성 요소에 의해 전자 장치(1000)가 구현될 수도 있고, 보다 적은 구성 요소에 의해 전자 장치(1000)가 구현될 수도 있다.

[0032] 저장부(1100)는 메모리로 호칭될 수 있으며, 제어부(1300)의 처리 및 제어를 위한 프로그램을 저장할 수 있고, 전자 장치(1000)로 입력되는 정보 또는 전자 장치(1000)로부터 출력되는 정보를 저장할 수도 있다.

[0033] 출력부(1200)는, 오디오 신호 또는 비디오 신호 또는 진동 신호를 출력할 수 있으며, 출력부(1200)는 전자 장치(1000)에서 처리되는 정보를 표시 출력한다. 디스플레이부(1210)는, 사용자의 입력에 대한 응답으로, 응답에 관련된 동작을 실행하기 위한 사용자 인터페이스를 디스플레이할 수 있다.

[0034] 제어부(1300)는 프로세서로 호칭될 수 있으며, 통상적으로 전자 장치(1000)의 전반적인 동작을 제어한다. 예를 들어, 제어부(1300)는, 저장부(1100)에 저장된 프로그램들을 실행함으로써, 사용자 입력부(1700), 출력부(1200), 통신부(1500), 사용자 입력부(1700) 등을 전반적으로 제어할 수 있다.

[0035] 통신부(1500)는, 전자 장치(1000)가 다른 장치(미도시) 및 서버(미도시)와 통신을 하게 하는 하나 이상의 구성요소를 포함할 수 있다. 다른 장치(미도시)는 전자 장치(1000)와 같은 컴퓨팅 장치이거나, 센싱 장치일 수 있으나, 이에 한정되는 것은 아니다.

[0036] 사용자 입력부(1700)는, 사용자가 전자 장치(1000)를 제어하기 위한 데이터를 입력하는 수단을 의미한다.

[0037] 도 2는 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 상세 블록도를 도시한 것이다.

[0038] 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템은 상술한 전자 장치(1000)일 수 있다. 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템은 표준화부(100), 위험(risk) 측정부(200), 가명처리부(300), 검토부(400)를 포함할 수 있다.

[0039] 표준화부(100), 위험 측정부(200), 가명처리부(300), 검토부(400) 중 적어도 하나는 본 발명의 실시예에 따른 방법을 수행하기 위한 하드웨어 또는 상기 하드웨어를 구동할 수 있는 소프트웨어의 기능적 또는 구조적 결합을 의미할 수 있다. 일 예로, 표준화부(100), 위험 측정부(200), 가명처리부(300), 검토부(400) 중 적어도 하나는 해당 기능을 수행하는 제어부()의 일부일 수 있다. 또한, 다른 예로, 표준화부(100), 위험 측정부(200), 가명처리부(300), 검토부(400) 중 적어도 하나는 제어부()에 의해 수행되는 메모리()에 저장된 소프트웨어의 일부 일 수 있다.

[0040] 표준화부(100), 위험 측정부(200), 가명처리부(300), 검토부(400)는 후술할 도 3의 S100, S200, S300, S400 각각의 과정을 수행하는 프로세서일 수 있으며, 혹은 각각의 과정을 수행하는 저장부()에 저장된 소프트웨어일 수 있다.

[0041] 도 3은 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법을 도시한 것이다. 도 4는 본 발명의 실시예에 따른 메타데이터를 표준화하는 방법의 흐름도를 도시한 것이다. 도 5는 본 발명의 다른 실시예에 따른 메타데이터를 표준화하는 방법의 흐름도를 도시한 것이다.

[0042] 도 3에 도시된 바와 같이, 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법은, 표준화하는 과정(S100), 위험을 측정하는 과정(S200), 가명처리를 수행하는 과정(S300), 가명처리된 가명정보를 검토하는 과정(S400)을 포함할 수 있다.

[0043] 본 발명의 실시예에 따른 사용환경에 대한 위험에 따른 정보집합물을 가명 처리하는 시스템의 제어방법은, 가명정보로부터 특정한 정보주체가 식별될 수 있는 위험을 최소화함으로써, 개인의 프라이버시를 보호할 수 있다.

[0044] S100 과정에서, 시스템은 표준화 과정을 수행할 수 있다. 표준화하는 과정(S100)은 정보집합물을 구성하는 메타데이터를 표준화하는 과정(S110), 정보집합물의 사용환경을 표준화하는 과정(S120)을 포함할 수 있다.

[0045] S110 과정에서, 시스템은 원천 정보집합물(Raw Data-Set)의 개인정보에 대한 메타데이터를 표준화하여 정보집합물(Data-Set)로 생성할 수 있다.

[0046] 도 4를 참조하며, 시스템이 개인정보에 대한 메타데이터를 표준화하는 방법을 설명한다. 시스템은 개인정보의 메타데이터를 저장부() 혹은 통신부()를 통해 획득하기 이전에 개인정보의 메타데이터를 표준화할 수 있다.

[0047] 시스템은 개인정보를 포함하는 원천 정보집합물을 저장부() 혹은 통신부()를 통해 획득할 수 있다(S511).

[0048] 원천 정보집합물(Raw Data-Set)은 정보주체인 개인의 개인정보를 포함할 수 있다. 예를 들어, 정보주체의 '이름', '나이', '주민등록번호', '전화번호' 등과 같은 개인에 관련된 데이터의 집합물을 의미한다. 원천 정보집합물은 데이터를 쉽게 검색할 수 있도록 일정한 규칙에 따라 체계적으로 배열되거나 구성된 데이터의 집합물을 말한다. 정보집합물(Data-Set)은 자료 집합, 데이터 세트라고도 불린다.

[0049] 원천 정보집합물은 데이터베이스의 테이블의 내용이나 통계적 자료 행렬을 의미할 수 있다. 테이블의 모든 필드(컬럼)의 명칭은 객체에 대한 변수들을 의미하고, 각 레코드(로우)는 변수들의 실제 값에 의해 정의된 하나의 객체를 의미할 수 있다. 예를 들어, 특정 객체에 대한 변수들은 사람의 '이름', '나이', '주민등록번호', '전화번호' 등을 포함할 수 있다. 변수들의 실제 값은 정보, 자료, 데이터라고 불리고, '홍길동', '29', '200101-112233', '010-2222-3333' 등의 값을 가질 수 있다.

[0050] 원천 정보집합물은 적어도 하나 또는 다수의 변수들의 실제 값들인 데이터들(자료들, 정보들)로 이뤄질 수 있다. 원천 정보집합물의 개수는 정보집합물에서 변수의 명칭의 레코드를 제외한 나머지 레코드의 수와 일치할 수 있다.

[0051] 시스템은 원천 정보집합물의 변수들을 획득할 수 있다(S512). 시스템은 원천 정보집합물로부터 변수들의 명칭을 획득할 수 있다. 개인정보의 변수들의 명칭은 '이름', '나이', '주민등록번호', '전화번호' 등을 포함할 수 있다.

[0052] 시스템은 변수들의 명칭에 대한 대표명칭이 존재하는지 확인할 수 있다(S513). 시스템은 저장부() 혹은 통신부()를 통해 획득한 변수들의 대표명칭에 대한 정보에 기초하여 원천 정보집합물에서 획득한 변수들이 대표명칭인지 확인할 수 있다.

[0053] 변수들의 대표명칭에 대한 정보는 동일한 종류의 데이터를 지칭하는 변수들의 명칭 집합과 명칭 집합을 대표하는 대표명칭을 포함할 수 있다. 예를 들어 나이에 대한 변수들의 명칭 집합은 ['age', '나이', '연세', 'old']이며, 이를 대표하는 대표명칭은 'age' 일수 있다. 또는, 변수들의 대표명칭에 대한 정보는 변수 마다 매핑된 대표명칭을 포함하는 매칭정보일 수 있다. 예를 들어, [변수-대표명칭]의 구조를 가지며, '나이'-'age', '연세'-'age', 'age'-'age'를 볼 수 있다. 변수들의 대표명칭에 대한 정보는 개인정보의 메타데이터에 포함되며, 개인정보 메타데이터는 저장부()에 저장될 수 있다.

[0054] 또한, 변수들의 대표명칭에 대한 정보는 원천 정보집합물에서 데이터의 변수들에 대한 대표명칭의 정보 및 각 객체들이 어떠한 변수들을 가지는지 정의하는 정보를 포함할 수 있다. 일 예로, 대표명칭의 정보는 객체(특정 정보주체)의 이름, 나이, 주민등록번호, 전화번호에 대한 정보와 같은 개인정보들의 변수들에 대한 표준화된 명칭 ['name', 'age', 'resident registration number', 'mobile number']의 정보를 포함하고, ['name', 'age', 'resident registration number', 'mobile number'] 중 어떠한 개인정보로 대표명칭의 정보가 이뤄질지에 대한 정보가 포함될 수 있다.

[0055] 시스템은 변수의 대표명칭이 존재하지 않는 경우 해당 변수에 대한 대표명칭에 대한 정보를 생성하여 저장부()에 저장할 수 있다(S514). 시스템은 변수의 대표명칭이 존재하는 경우, 다음 과정을 수행할 수 있다.

[0056] 시스템은 변수에 대한 개인정보의 유형이 존재하는지 확인할 수 있다(S515).

[0057] 원천 정보집합물에 저장된 개인정보들은, 데이터의 실제 값(필드의 값)으로서, 데이터의 정보주체(혹은 객체)의 식별의 용이성에 따라 ①직접식별자(고유식별자), ②간접식별자(준식별자), ③속성정보, ④특이정보의 4가지 유형으로 구분될 수 있다.

[0058] “직접식별자”란 해당 정보주체에게만 고유하게 부여되어 있는 정보로써 그 자체만으로 개인 식별성이 강한 정보이고, “준식별자”는 정보주체에게만 고유하게 부여된 정보는 아니지만 보편적으로 널리 이용되고 있는 정보이어서 다른 정보와 결합하면 특정 개인을 식별하기 쉬운 정보이며, “속성정보”는 주로 해당 개인정보처리자만 보유하고 있어 개인정보처리자 이외의 자는 다른 정보와 결합해도 특정 개인을 식별하기 어려운 정보이고, “특이정보”는 해당 정보주체에게 고유하게 부여된 정보는 아니지만 해당 정보주체에 대해서만 해당되는 정보여서 누구든지 쉽게 식별이 가능한 정보를 의미한다.

[0059] 시스템은 저장부() 혹은 통신부()를 통해 복수개의 개인정보의 유형분류정보를 획득할 수 있다.

[0060] 복수개의 개인정보의 유형분류정보는 개인정보가 직접식별자, 준식별자, 속성정보, 특이정보 중 적어도 하나의 유형에 속하는지 알려준다. 유형분류정보는 모든 유형의 직접식별자, 준식별자, 속성정보, 특이정보에 대한 개인정보를 포함할 수 있고, 적어도 하나의 유형의 직접식별자, 준식별자, 속성정보, 특이정보에 대한 개인정보만을 포함할 수 있다.

[0061] 일 예로, 복수개의 개인정보의 유형분류정보는 직접식별자, 준식별자, 속성정보, 특이정보에 속하는 개인정보들의 리스트들이 위험등급에 따라 다르게 구성된다. 예를 들어, 특정 개인정보는 1등급에서 준식별자에 대한 정보일 수 있지만, 2등급에서 특이정보에 속할 수 있다. 다른 예로, 복수개의 개인정보의 유형분류정보는 정보집합물의 민감도, 사용자 수준, 정보보호 수준 중 적어도 하나에 기초하여 직접식별자, 준식별자, 속성정보, 특이정보에 속하는 개인정보들의 리스트들이 다르게 구성된다.

[0062] 개인정보의 유형분류정보의 일 예로, 개인정보 중 이름, 사진, 주민등록번호, 전화번호, 이메일주소, IP주소, 차량의 번호판정보 등은 직접식별자에 속하고, 개인정보 중 생년월일, 사망일, 결혼기념일, 직업, 성별, 신용등급, 주소, 우편번호 등은 준식별자에 속하고, 개인정보 중 상품 구매이력, 월별 전화 사용액, 대출 총액, 예금 총액, 보험 구좌수, 고객고유번호 등은 속성정보에 속하고, 개인정보 중 특이 질환자, 초고소득자, 초고령 연령자, 특이 직업, 이동 동선 등은 특이정보에 속할 수 있다.

[0063] 다른 개인정보 유형분류정보의 다른 예로, 개인정보 중 이름, 사진, 주민등록번호, 전화번호, 이메일주소, IP주소, 차량의 번호판정보 등은 직접식별자에 속하고, 개인정보 중 생년월일, 사망일, 결혼기념일, 직업, 성별, 신용등급, 주소, 우편번호 등은 준식별자에 속할 수 있고, 속성정보 또는 특이정보에 속하는 개인정보는 없을 수 있다.

[0064] 시스템은 개인정보의 유형분류정보에 기초하여 원천 정보집합물의 변수가 어떤 유형의 개인정보인지 확인할 수 있으며, 대응하는 개인정보의 유형이 없는 경우 새로운 개인정보의 유형을 생성할 수 있다. 예를 들어, 유형분류정보가 직접식별자 및 준식별자에 대한 개인정보를 포함하되, 속성정보에 속하는 개인정보를 포함하지 않는 경우 속성정보의 유형을 새롭게 생성하고, 새롭게 생성된 속성정보의 유형을 포함하는 새로운 유형분류정보를 생성하여 저장할 수 있다(S516).

[0065] 시스템은 원천 정보집합물의 변수가 개인정보의 유형분류정보에 속하는 경우 다음 과정으로, 유형별 가명처리 수준이 존재하는지 확인할 수 있다(S517). 시스템은 유형별 가명처리 수준에 대한 정보를 저장부() 혹은 통신부를 통해 획득할 수 있다. 시스템은 유형별 가명처리 수준에 대한 정보에 기초하여 개인정보의 유형에 대한 가명처리 수준이 존재하는지 확인할 수 있다. 혹은 시스템은 유형별 가명처리 수준에 대한 정보에 기초하여 유형분류정보에 속한 유형들의 가명처리 수준이 존재하는지 확인할 수 있다.

[0066] 시스템은 유형별 가명처리 수준이 존재하지 않는 경우 새로운 유형별 가명처리 수준에 대한 정보를 생성하여 저장할 수 있다(S518). 새로운 유형별 가명처리 수준에 대한 정보는 데이터의 유형 및 속성에 따른 각 유형별 가명처리 기법, 개인정보의 유형 마다 할당된 가명처리 정보의 식별자를 포함할 수 있다.

[0067] 시스템은 유형별 가명처리 수준이 존재하는 경우, 개인정보의 메타데이터를 획득할 수 있다(S519).

[0068] 개인정보의 메타데이터는 정보집합물(개인정보를 포함함)의 생성 및 사용 목적에 대한 정보 및 정보집합물이 개인정보 중 정보주체를 식별할 수 있는 속성 분류 정보를 포함할 수 있다.

[0069] 개인정보의 메타데이터는 원천 정보집합물의 변수의 정보, 변수의 대표명칭에 대한 정보, 개인정보의 유형에 대한 정보, 유형분류정보, 유형별 가명처리 수준 정보, 유형분류정보에 따른 가명처리 수준 정보를 포함할 수 있다.

[0070] 개인정보의 메타데이터는 변수의 실제 값(예로, 데이터 값)에 대한 정보를 포함할 수 있다. 변수의 실제 값에 대한 정보는 데이터의 유형과 특성에 대한 정보를 포함할 수 있다. 데이터의 타입은 숫자, 문자, 날짜로 구분될 수 있으며, 숫자의 경우 연속형, 이산형 등을 포함할 수 있다. 데이터의 특성은 해당 변수가 가지는 특징을 가질 수 있으며, 일 예로 성별의 경우 2가지의 변수만 존재하는 바이너리 특징을 가지고, 나이의 경우 최소값과 최대값의 특징을 부여할 수 있다.

[0071] 시스템은 개인정보의 메타데이터에 기초하여 데이터의 변수를 확인할 수 있다.

[0072] 일 예로, 시스템은 원천 정보집합물의 데이터의 변수(필드의 명칭)을 개인정보의 메타데이터로 변경하기 위하여 변수의 실제 값(데이터 값)을 확인하고, 데이터의 유형 및 특성에 기초하여 데이터의 변수의 명칭을 유도할 수 있다. 변수의 실제 값에 대한 정보에 기초하여 변수의 실제 값이 어떠한 변수에 해당하는지 확인할 수 있다. 예를 들어, '남' 혹은 'male'의 데이터 값을 보고 문자에 해당하고 성별을 의미하는 바이너리 특징에 속하므로, ['gender'] 변수의 데이터임으로 판단할 수 있다.

[0073] 시스템은 개인정보의 메타데이터를 이용하여 정보집합물을 획득할 수 있다(S520). 시스템은 개인정보에 대한 메타데이터에 기초하여 원천 정보집합물의 데이터의 변수가 변수의 대표명칭과 일치하는 확인한 후에 일치하지 않는 경우 대표명칭으로 변수를 변경할 수 있다. 이를 통해서 생성된 데이터의 집합을 정보집합물(Data-Set)이라 한다.

[0074] 정보집합물(Data-Set)은 원천 정보집합물과 동일하게 테이블 혹은 행렬로 구성될 수 있으며, 테이블의 각 레코드는 제기된 정보집합물의 객체를 의미하고, 모든 필드는 객체에 대한 변수들(혹은 속성들)을 의미할 수 있다.

[0075] 표준화 과정에서 생성된 정보집합물(Data-Set)은 원천 정보집합물를 개인정보의 메타데이터를 이용하여 개인정보의 유형에 대응하는 변수의 대표명칭으로 변수를 변경함으로써 표준화한 정보의 집합물을 의미한다. 혹은 정보집합물은 적어도 2개 이상의 정보집합물의 결합된 정보집합물을 의미한다.

[0076] S110 과정에서, 다른 예로, 시스템은 원천 정보집합물(Raw Data-Set)의 개인정보에 대한 메타데이터를 표준화하여 정보집합물(Data-Set)로 생성할 수 있다.

[0077] 도 5를 참조하며, 다른 예로, 시스템이 개인정보에 대한 메타데이터를 표준화하는 방법을 설명한다. 시스템은 개인정보의 메타데이터를 저장부() 혹은 통신부()를 통해 획득하기 이전에 개인정보의 메타데이터를 표준화할 수 있다.

[0078] 시스템은 원천 정보집합물을 저장부() 혹은 통신부()를 통해 획득할 수 있다(S611).

[0079] 시스템은 원천 정보집합물의 특정 변수의 실제 값들(필드의 값들)을 획득할 수 있다(S612). 예를 들어, '생년월일' 필드의 값들은 '990211', '001212', '880102' 등을 포함할 수 있다.

[0080] 시스템은 저장부() 혹은 통신부()를 통해 획득한 데이터 타입의 대비정보에 기초하여 변수의 실제 값인 데이터의 타입을 확인할 수 있다. 데이터 타입이란 숫자, 문자, 날짜 와 같은 정보를 의미하며, 숫자의 경우 연속형, 이산형의 유형으로 구분될 수 있고, 문자의 경우 단어, 문장, 단락의 유형으로 구분될 수 있다.

[0081] 시스템은 원천 정보집합물의 변수의 실제 값인 데이터의 타입이 데이터 타입의 대비정보 내에 존재하지 않는 경우 원천 정보집합물의 변수의 데이터의 실제 값을 정의하는 새로운 데이터 타입을 생성하여 저장할 수 있다(S614).

[0082] 시스템은 원천 정보집합물의 변수의 실제 값에 대한 데이터 타입이 존재하는 경우 데이터의 특성을 확인할 수 있다(S615). 시스템은 데이터 특성의 대비정보를 저장부() 혹은 통신부()를 통해 획득할 수 있다. 시스템은 데이터 특성의 대비정보 내에 원천 정보집합물의 실제 값인 데이터의 특성이 속하는지 확인할 수 있다.

[0083] 데이터의 특성이란 특정 변수(필드)가 가지는 정보로써, '나이' 변수는 숫자이면서, 최소값과 최대값을 가지는 특징을 가지며, '성별' 변수의 경우 데이터가 '남' 혹은 '여'로 2가지만 가질 수 있는 특징을 가진다.

[0084] 시스템은 원천 정보집합물의 변수의 실제 값에 대한 데이터 특성이 존재하지 않는 경우 원천 정보집합물의 변수의 데이터의 실제 값을 정의하는 새로운 데이터 특성을 생성하여 저장할 수 있다(S616).

[0085] 시스템은 원천 정보집합물의 변수의 실제 값에 대한 데이터 타입 및 특성에 기초하여 경우 원천 정보집합물의 변수의 대표명칭을 확인할 수 있다(S617). 시스템은 원천 정보집합물의 변수의 대표명칭이 없는 경우 새로운 변수의 대표명칭을 생성할 수 있다(S618).

[0086] 시스템은 개인정보의 메타데이터를 획득할 수 있다(S519). 개인정보의 메타데이터는 정보집합물(개인정보를 포함함)의 생성 및 사용 목적에 대한 정보 및 정보집합물이 개인정보 중 정보주체를 식별할 수 있는 속성 분류 정보를 포함할 수 있다.

[0087] 개인정보의 메타데이터는 변수의 실제 값(예로, 데이터 값)에 대한 정보를 포함할 수 있다. 변수의 실제 값에 대한 정보는 데이터의 타입과 특성에 대한 정보를 포함할 수 있다. 데이터의 타입은 숫자, 문자, 날짜로 구분될 수 있으며, 숫자의 경우 연속형, 이산형 등을 포함할 수 있다. 데이터의 특성은 해당 변수가 가지는 특징을 가질 수 있으며, 일 예로 성별의 경우 2가지의 변수만 존재하는 바이너리 특징을 가지고, 나이의 경우 최소값과 최대값의 특징을 부여할 수 있다.

[0088] 시스템은 대표명칭이 있는 경우 해당 원천 정보집합물의 변수를 대표명칭으로 변경할 수 있다(S620). 시스템은 개인정보의 메타데이터를 이용하여 정보집합물을 획득할 수 있다. 시스템은 개인정보에 대한 메타데이터에 기초하여 원천 정보집합물의 데이터의 변수를 대표명칭으로 변경할 수 있다. 이를 통해서 생성된 데이터의 집합을 정보집합물(Data-Set)이라 한다.

[0089] 이를 통해서 시스템은 원천 정보집합물의 데이터 값들을 분석함으로써 데이터의 타입 및 특성을 확인하고, 확인된 데이터의 타입 및 특성에 기초하여 데이터의 변수를 확인한 후에 변수의 대표명칭으로 변경함으로써, 기존 변수의 명칭에 관계없이 데이터의 값을 이용해 대표명칭으로 변수를 변경할 수 있다.

[0090] 참고로, 원천 정보집합물은 표준화되기 전의 정보집합물로, 동일한 데이터 값(실제 값)을 표현하는 변수의 명칭이 다른 경우 서로 다른 원천 정보집합물일 수 있다. 다만, 서로 다른 원천 정보집합물들이라도 표준화를 통해서 동일한 정보집합물이 될 수 있다. 예를 들어, 제1 원천 정보집합물과 제2 원천 정보집합물이 '홍길동', '29', '200101-112233', '010-2222-3333'와 같은 서로 동일한 객체에 대한 실제 값을 가지더라도, 제1 원천 정보집합물에서 해당 객체에 대한 변수들의 용어가 '이름', '나이', '주민등록번호', '전화번호'이고, 제2 원천 정보집합물에서 해당 객체에 대한 변수들의 용어가 '이름', 'age', '주민등록번호', '전화번호'인 경우 제1 원천 정보집합물과 제2 원천 정보집합물은 서로 다른 원천 정보집합물로 분류된다.

[0091] 도 3에 도시된 바와 같이, S120 과정에서, 시스템은 정보집합물의 사용환경을 표준화할 수 있다.

[0092] 시스템은 정보집합물의 사용환경의 표준화 정보를 정보집합물의 취급자의 입력, 저장부(), 또는 통신부()로부터 획득할 수 있다. 정보집합물의 사용환경은 정보집합물이 이용 및 활용되는 처리 과정의 모든 요소를 포함할 수 있다.

[0093] 가명정보 내의 존재하는 특정 정보주체는 정보집합물을 이용하는 사용자의 배경지식에 의해 재식별되거나 가명정보와 함께 사용되는 다른 정보집합물에 의해 재식별될 가능성이 있다. 그러므로, 본 발명은 가명정보가 사용되는 사용환경을 표준화함으로써, 가명정보로부터 정보주체의 재식별 가능성을 통제할 수 있다.

[0094] 정보집합물의 사용환경의 표준화는 정보집합물에 대한 민감도 수준의 표준화, 사용자 수준의 표준화, 정보집합물의 사용시 정보보호 수준의 표준화를 포함할 수 있다.

[0095] 정보집합물에 대한 민감도 수준이란 가명정보의 원천데이터인 정보집합물의 민감도를 평가하는 방법을 표준화하여 등급화된 수준을 말하며, 가명처리 대상이 되는 정보집합물의 생성시기, 취급 정보주체 수, 시계열 등의 데이터 형태, 고유식별자와 준식별자 존재 등을 종합적으로 평가한 수준을 포함할 수 있다.

[0096] 또한, 사용자 수준은 사용자의 데이터 관련 능력을 등급화한 수준을 말하며, 정보집합물을 이용하는 사용자의 데이터 분석 능력으로, 구체적으로 데이터 사이언티스트, DBA 등 전문가와 영업자료 활용 등 단순 업무처리자 등으로 구분한 수준을 포함할 수 있다.

[0097] 저장부()는 정보집합물을 사용할 사용자들에 대한 저장정보를 저장할 수 있다. 시스템은 저장부()에 저장된 사용자들에 대한 정보에 기초하여 사용자 수준을 표준화할 수 있다. 예를 들어, 사용자들에 대한 저장정보는 사용자의 기술영역, 직업, 전문가 여부, 기업유무, 기업 형태, 업무영역 등에 기초하여 분류된 복수의 사용자 수준

들에 대한 정보를 포함할 수 있다. 시스템은 정보집합물을 사용할 사용자 정보를 입력 받으면, 사용자에 대한 저장정보와 대비하여 해당 사용자의 사용자 수준을 결정할 수 있다.

[0098] 또한, 정보보호 수준은 사용자 및 비인가자를 통제하기 위하여 등급화한 수준을 말하며, 정보집합물을 대상으로 접근통제, 사용 이력 수집 분석 등 내부의 정보를 보호하기 위해 기술적, 물리적, 관리적 정보보호를 평가한 수준을 포함할 수 있다.

[0099] 저장부()는 정보집합물을 사용할 때 개인정보의 보호수준에 대한 저장정보를 저장할 수 있다. 시스템은 저장부()에 저장된 보호수준에 대한 정보에 기초하여 개인정보의 보호수준을 표준화할 수 있다. 예를 들어, 정보집합물의 사용시 보호수준에 대한 저장정보는 정보집합물을 대상으로 접근통제, 사용 이력 수집 분석 등 내부의 정보를 보호하기 위해 기술적, 물리적, 관리적 정보보호를 평가한 수준에 대한 정보를 포함할 수 있다. 시스템은 정보집합물을 입력 받으면, 보호수준에 대한 저장정보와 대비하여 해당 정보집합물의 사용시 보호수준을 결정할 수 있다.

[0100] S200 과정에서, 시스템은 위험(리스크)을 획득할 수 있다. 위험을 획득하는 과정(S200)은 사용환경을 획득하는 과정(S210)과 위험의 등급을 결정하는 과정(S220)을 포함할 수 있다.

[0101] S210 과정에서, 시스템은 사용환경을 측정하기 위하여 정보집합물의 민감도, 사용자 수준, 보호수준에 대한 정보들 중 적어도 하나를 정보집합물의 취급자의 입력 혹은 저장부()로부터 획득할 수 있다. 정보집합물이 사용될 환경이란 가명정보 내의 개인정보들의 민감도 수준, 정보집합물을 실제 사용하는 사람들의 대한 수준, 가명정보에 접근성 혹은 사용권한에 대한 정보에 따라서 조성되고, 가명정보가 사용되는 환경을 의미한다.

[0102] S220 과정에서, 시스템은 위험의 등급을 결정할 수 있다. 시스템은 획득된 사용환경에 대한 정보에 기초하여 위험 점수 혹은 위험의 등급을 결정할 수 있다. 시스템은 앞서 측정한 개인정보들의 민감도 수준, 가명정보를 사용하는 사용자 수준, 가명정보에 대한 보호 수준 중 적어도 하나를 S120 과정에서 표준화한 사용환경의 민감도 수준, 사용자 수준, 보호 수준과 대비하여 사용환경에 대한 위험도를 결정할 수 있다.

[0103] 또한, S300 과정에서, 시스템은 가명처리를 수행할 수 있다. 가명처리를 수행하는 과정(S300)은 정보집합물에 대한 처리 수준을 결정하는 과정(S310)과, 가명처리를 수행하는 과정(S320)을 포함할 수 있다.

[0104] S310 과정에서, 시스템은 획득한 사용환경 자체 혹은 위험 점수(등급)에 기초하여 정보집합물에 대한 처리 수준에 대한 정보를 결정할 수 있다. 정보집합물에 대한 처리 수준에 대한 정보는 기본 처리 수준에 대한 정보와, 가중치 수준의 정보를 포함할 수 있다.

[0105] 기본 처리 수준에 대한 정보는 개인정보(데이터의 실제 값)를 가명처리하기 위해 매핑정보를 포함할 수 있다. 이 경우 개인정보는 준식별자에 해당하는 정보일 수 있다. 매핑정보은 특정한 변수의 데이터 혹은 특정한 데이터 구조(유형, 특성)를 특정한 형태의 다른 데이터로 변경시킬 수 있는 정보를 포함한다. 즉, 기본처리 수준에 대한 정보는 개인정보를 추가정보를 사용하지 않고는 재식별이 불가능하게 가명처리하는 방법에 대한 정보를 포함한다.

[0106] 일 예로, 생년월일에 대한 기본 처리 수준에 대한 정보는 생년월일에 대한 데이터가 '110101', '990112', '031212'인 경우 앞의 2자리를 기초로 각각 '2011년생', '1999년생', '2003년생'을 도출할 수 있는 정보 내지 알고리즘을 포함할 수 있다.

[0107] 가중치 수준의 정보는 개인정보에 적용되는 기본 처리 수준에 부가적으로 적용되는 가중치를 의미한다. 가중치 수준의 정보는 획득한 사용환경 혹은 위험 등급에 따라 부가적으로 적용될 수 있다. 예를 들어, 가중치 수준의 정보는 위험 등급에 따라서 특정한 개인정보에 적용되는 가명 정도의 가중치를 올려서 비식별정도를 증가시킬 수 있으며, 가중치를 내려서 식별정도를 증가시킬 수 있는 정보를 포함할 수 있다.

[0108] 시스템은 위험 등급에 따라 정보집합물에 대한 가명처리 수준을 결정할 수있다.

[0109] 일 예로, 시스템은 위험 등급에 따라 특이치 제거 처리수준을 결정할 수 있다 구체적으로, 연속형의 숫자는 정규분포 형태를 띠며 이 경우 시그마 분포에 들어있는 특이치를 제거하는데, 3시그마 분포는 가장 낮은 위험한 경우 제거하고, 2시그마는 중간 위험한 경우 제거하고, 1시그마는 가장 위험한 경우 제거하도록 처리할 수 있다. 시스템은 정규분포의 시그마 분포에 따라 위험 등급을 결정할 수 있으며, 결정된된 위험 등급에 따라서 특이치 데이터를 제거할 수 있다.

[0110] 다른 일 예로, 시스템은 위험 등급에 따라 차등 개인정보보(Differential Privacy)의 민감도 생성 수준을 결정

할 수 있다. 민감도 생성 수준은 간접식별자들을 묶어서 동일한 정보(K-익명성의 K값)들로 구성된 클래스를 형성할 때 클래스를 구성하는 요소들의 숫자를 기준(평균 또는 특정 K값)으로 할 수 있다. 시스템은 민감도 생성 수준에 기초하여 위험 등급에 따른 차등 개인정보보호의 정도르 결정할 수 있다.

[0111] S320 과정에서, 시스템은 가명처리를 수행할 수 있다. 시스템은 가명처리 시에 개인정보의 최소처리원칙을 준수하기 위하여 처리환경, 사용환경, 처리 목적, 정보의 성격 등에 관한 정보에 기초하여 정보집합물에 대한 가명처리를 수행하여 가명정보를 생성할 수 있다. 시스템은 기본 처리 수준에 대한 정보와 가중치 수준의 정보에 기초하여 정보집합물에 대한 가명처리를 수행하여 가명정보를 생성할 수 있다.

[0112] 가명정보란 개인정보를 가명처리 함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보를 의미한다. 가명처리란 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것을 의미한다.

[0113] 가명처리 방식으로는 휴리스틱 가명화, 암호화, 교환 방법이 있다. 휴리스틱 가명화는 식별자에 해당하는 값을 기 설정된 규칙에 따라 가공해 개인정보를 숨긴다. 암호화는 일정한 규칙의 알고리즘을 적용해 암호화함으로써 개인정보를 대체하는 방법이다. 교환은 기존의 데이터베이스의 레코드를 사전에 정해진 외부 변수(항목)값과 연계해 바꾸는 방법이다. 가명처리의 방식은 정보집합물에 대한 처리 수준에 대한 정보에 의해서 결정될 수 있다.

[0114] 또한, S400 과정에서, 시스템은 가명처리 후에 가명정보를 검토할 수 있다. 가명처리된 가명정보를 검토하는 과정(S400)은 가명정보에 대한 적정성 검토 및 추가적인 가명처리를 수행하는 과정(S410)과 가명정보에 대한 사후관리를 수행하는 과정(S420)을 포함할 수 있다.

[0115] S410 과정에서, 시스템은 가명정보가 목적달성을 위해 적절한 수준으로 가명처리가 이루어졌는지 검토할 수 있으며, 정보주체에 대한 재식별 가능성이 없는지 판단할 수 있다. 시스템은 정보주체에 대한 재식별 가능성이 있다면, 리스트 등급의 상향 조절하여 추가적인 가명처리를 수행할 수 있다. 이로써, 특정주체에 대한 재식별 위험을 최소화하고, 취약점을 보완할 수 있다.

[0116] S420 과정에서, 시스템은 S410 과정의 검토결과로 정보주체에 대한 재식별 가능성이 없어서 적정으로 판단된 경우에 가명정보에 대한 실사용 정보를 획득하여 기 설정된 기준과 대비하여 가명정보에 대한 실사용에 문제 여부를 판단할 수 있다. 기설정된 기준은 개인정보보호법 등과 관련된 법령을 준수하기 위해 가명정보를 사용하는 기술적 기준, 가명정보를 사용하는 사용자의 관리적 기준, 가명정보가 사용되는 환경등의 물리적 기준을 포함할 수 있다.

[0117] 본 발명은 도면에 도시된 실시 예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

[0119] 표준화부(100)

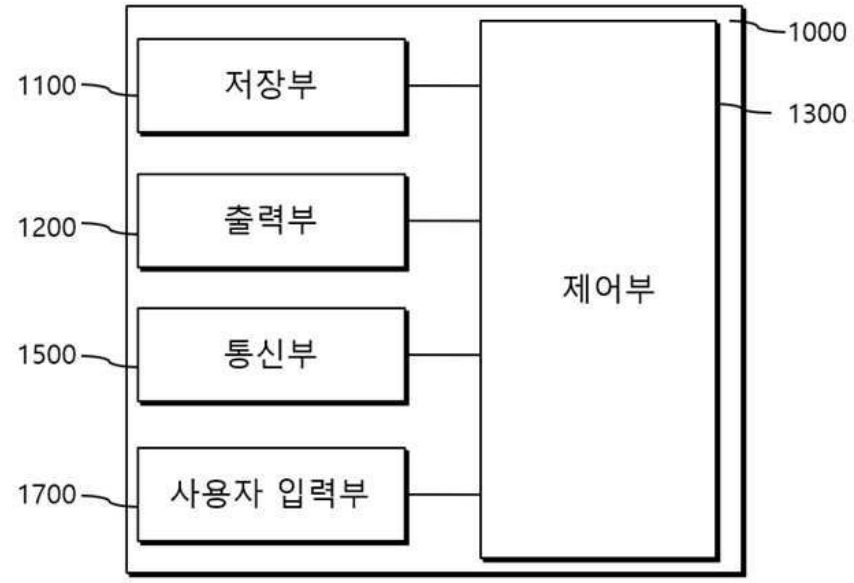
위험 측정부(200)

가명처리부(300)

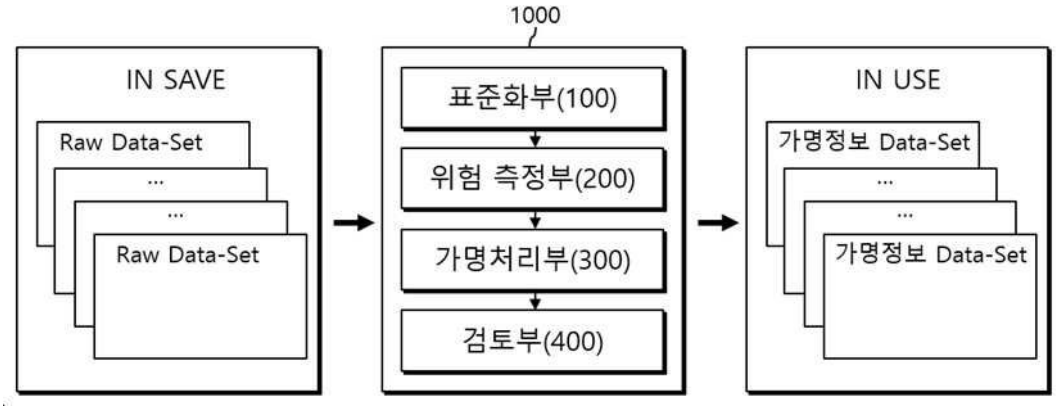
검토부(400)

도면

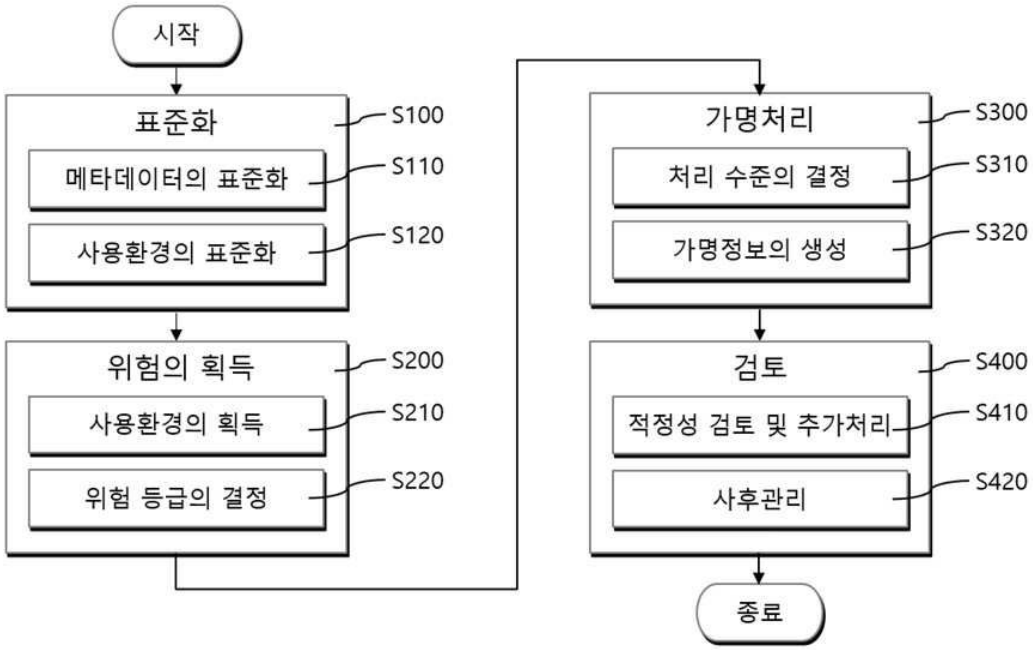
도면1


1000
1100
저장부
1200
출력부
1500
통신부
1700
사용자 입력부
제어부
1300

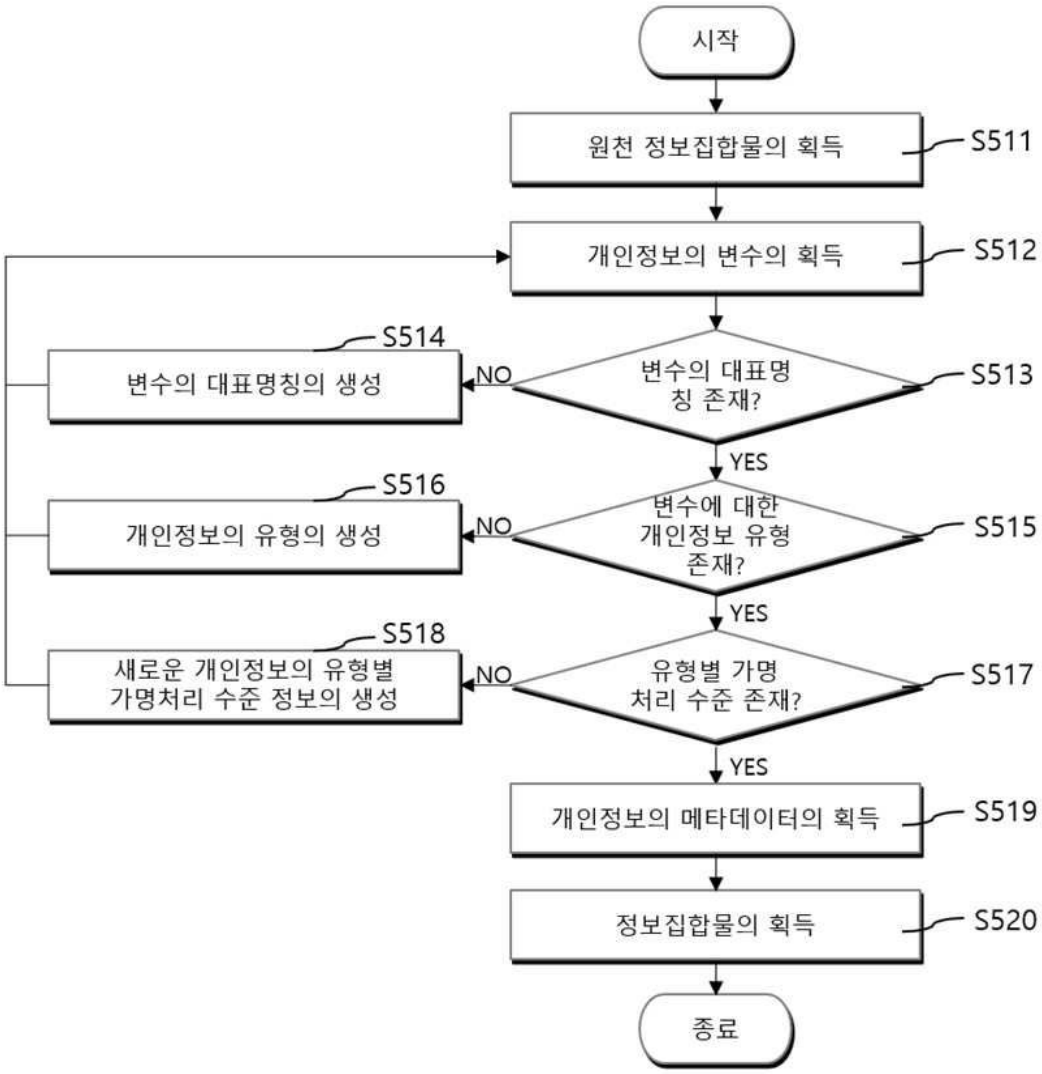
도면2


1000
IN SAVE
Raw Data-Set
...
...
Raw Data-Set
표준화부(100)
위험 측정부(200)
가명처리부(300)
검토부(400)
IN USE
가명정보 Data-Set
...
...
가명정보 Data-Set

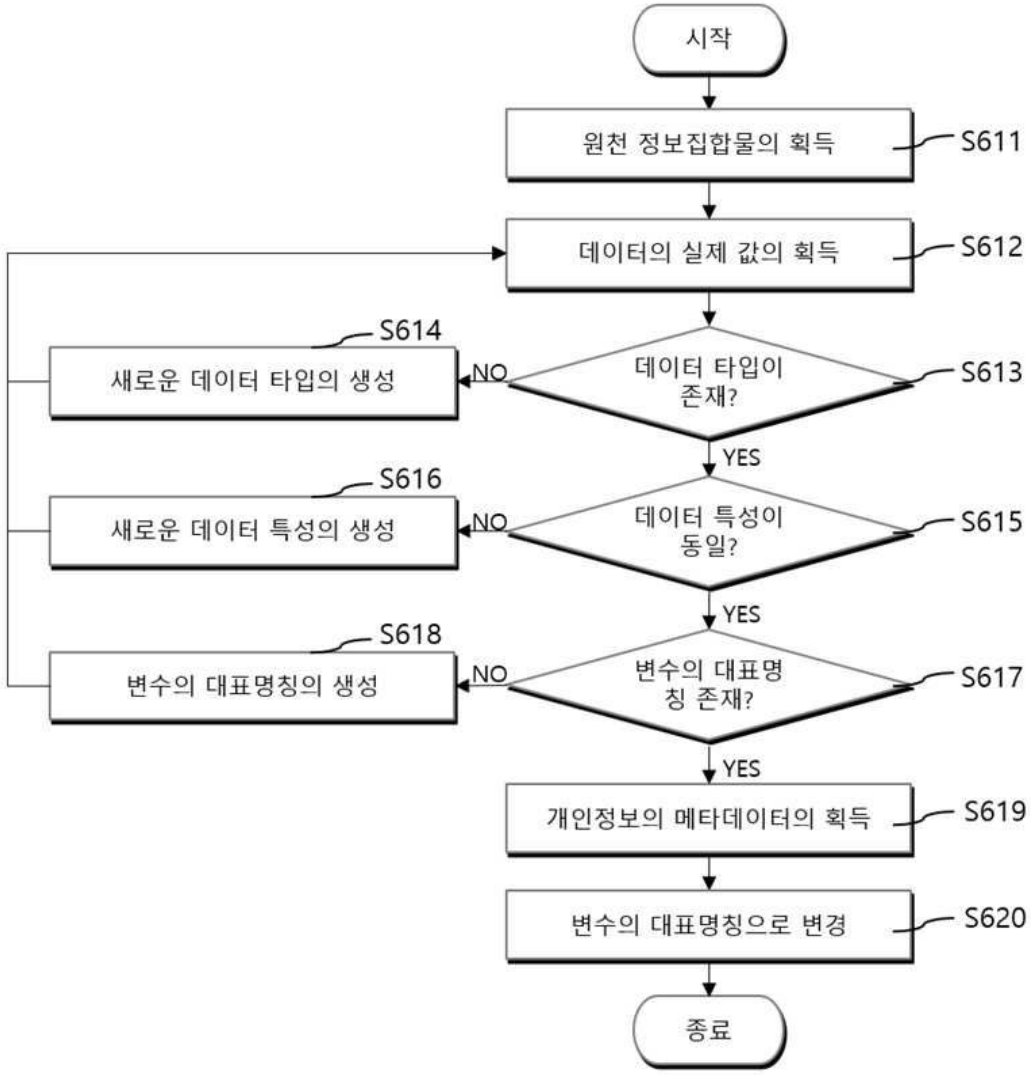
도면3


시작
표준화
S100
메타데이터의 표준화
S110
사용환경의 표준화
S120
위험의 획득
S200
사용환경의 획득
S210
위험 등급의 결정
S220
가명처리
S300
처리 수준의 결정
S310
가명정보의 생성
S320
검토
S400
적정성 검토 및 추가처리
S410
사후관리
S420
종료

도면4


시작
원천 정보집합물의 획득
S511
개인정보의 변수의 획득
S512
변수의 대표명
칭 존재?
S513
NO
변수의 대표명칭의 생성
S514
YES
변수에 대한
개인정보 유형
존재?
S515
NO
개인정보의 유형의 생성
S516
YES
유형별 가명
처리 수준 존재?
S517
NO
새로운 개인정보의 유형별
가명처리 수준 정보의 생성
S518
YES
개인정보의 메타데이터의 획득
S519
정보집합물의 획득
S520
종료

도면5


시작
원천 정보집합물의 획득
S611
데이터의 실제 값의 획득
S612
데이터 타입이 존재?
S613
NO
새로운 데이터 타입의 생성
S614
YES
데이터 특성이 동일?
S615
NO
새로운 데이터 특성의 생성
S616
YES
변수의 대표명칭 존재?
S617
NO
변수의 대표명칭의 생성
S618
YES
개인정보의 메타데이터의 획득
S619
변수의 대표명칭으로 변경
S620
종료