

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호

WO 2024/019205 A1

2024 년 1 월 25 일 (25.01.2024)

WIPO | PCT

(51) 국제특허분류:

G06F 21/62 (2013.01)

G06Q 30/02 (2012.01)

G06F 16/907 (2019.01)

(21) 국제출원번호:

PCT/KR2022/011120

(22) 국제출원일:

2022 년 7 월 28 일 (28.07.2022)

(25) 출원언어:

한국어

(26) 공개언어:

한국어

(30) 우선권정보:

10-2022-0090970 2022 년 7 월 22 일 (22.07.2022) KR

(71) 출원인: 쿠팡 주식회사 (COUPANG CORP.) [KR/KR];
05510 서울특별시 송파구 송파대로 570, 18층, Seoul (KR).

(72) 발명자: 정운기 (JUNG, Un Gi); 05510 서울특별시 송파구 송파대로 570, Seoul (KR). 임이랑 (LIM, Yi Rang); 05510 서울특별시 송파구 송파대로 570, Seoul (KR). 손희연 (SON, Hoe Yeon); 05510 서울특별시 송파구 송파대로 570, Seoul (KR).

(74) 대리인: 특허법인 광장리앤고 (LEE & KO IP); 04532 서울특별시 중구 남대문로 63, 3층, Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

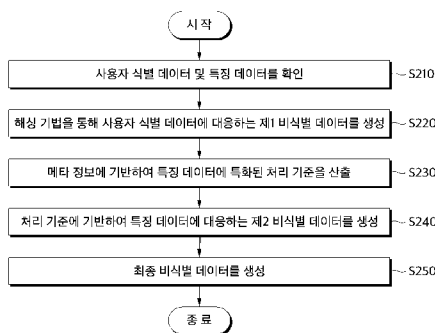
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

공개:

— 국제조사보고서와 함께 (조약 제21조(3))

(54) Title: ELECTRONIC DEVICE AND METHOD FOR DATA PROCESSING

(54) 발명의 명칭: 데이터 처리를 위한 전자 장치 및 그 방법



S210 ... Identify user identification data and feature data
S220 ... Generate first non-identified data corresponding to user identification data through hashing technique
S230 ... Calculate processing criterion specific to feature data on basis of meta information
S240 ... Generate second non-identified data corresponding to feature data on basis of processing criterion
S250 ... Generate final non-identified data

(57) Abstract: An electronic device for data processing and a method therefor are disclosed. The method for data processing according to the present disclosure comprises the steps of: identifying user identification data capable of identifying a service user and feature data corresponding to the user identification data; generating first non-identified data corresponding to the user identification data through a hashing technique; calculating a processing criterion specific to the feature data on the basis of meta information that defines a data processing scheme; generating second non-identified data corresponding to the feature data by processing at least a portion of the feature data on the basis of the processing criterion; and generating final non-identified data including one or more tables having the first non-identified data as a key and at least some of the second non-identified data as values.

(57) 요약서: 데이터 처리를 위한 전자 장치 및 그 방법이 개시된다. 본 개시에 다른 데이터 처리 방법은, 서비스 사용자를 식별 가능한 사용자 식별 데이터 및 사용자 식별 데이터에 대응하는 특징 데이터를 확인하는 단계, 해싱(Hashing) 기법을 통해 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성하는 단계, 데이터의 처리 방식을 정의한 메타 정보에 기반하여 특징 데이터에 특화된 처리 기준을 산출하는 단계, 처리 기준에 기반하여 특징 데이터의 적어도 일부를 처리함으로써 특징 데이터에 대응하는 제2 비식별 데이터를 생성하는 단계, 및 제1 비식별 데이터를 키(key)로, 제2 비식별 데이터 중 적어도 일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별 데이터를 생성하는 단계를 포함한다.

WO 2024/019205 A1

명세서

발명의 명칭: 데이터 처리를 위한 전자 장치 및 그 방법

기술분야

- [1] 본 개시는 개인 정보가 포함된 민감 데이터를 비식별화 처리하는 전자 장치 및 그 방법에 관한 것이다.

배경기술

- [2] 이-커머스(E-commerce)를 비롯한 다양한 서비스를 운영함에 따라, 수많은 사용자의 개인 정보가 포함된 민감 데이터가 축적된다. 보다 증진된 사용자 경험을 제공하고 서비스를 개선하기 위해서는 이러한 사용자의 데이터를 인공지능 모델의 학습에 이용하거나, 통계적으로 처리하는 등, 적극적으로 활용하여야 한다.
- [3] 그러나, 이 과정에서 사용자의 민감 데이터가 노출될 위험도 존재하며, 자신의 민감 데이터가 서비스 제공자에 의해 사용되는 것 자체를 꺼리는 사용자도 많은 것이 사실이다.
- [4] 따라서, 사용자의 민감 데이터가 지닌 특징을 과하게 회색시키지 않으면서도, 누구의 민감 데이터인지 드러나지 않도록 비식별화(익명화 내지는 가명화) 처리할 필요가 있다.
- [5] 관련하여, KR2417044B1 건 등의 선행문헌들을 참조할 수 있다.

발명의 상세한 설명

기술적 과제

- [6] 본 개시는 상술한 문제점을 해결하기 위해 제안된 것으로, 사용자의 데이터가 지닌 특징을 적절히 유지하면서도 각 데이터가 어느 사용자에 대한 데이터인지 식별할 수 없도록 데이터를 처리하는 전자 장치 및 방법을 제공하는 것을 목적으로 한다.
- [7] 본 개시에 의해 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 이하의 실시 예들로부터 또 다른 기술적 과제들이 유추될 수 있다.

과제 해결 수단

- [8] 개시되는 일 실시 예에 따른 데이터 처리 방법은, 서비스 사용자를 식별 가능한 사용자 식별 데이터 및 상기 사용자 식별 데이터에 대응하는 서비스 사용자의 특징 데이터를 확인하는 단계; 해싱(Hashing) 기법을 통해 상기 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성하는 단계; 데이터의 처리 방식을 정의한 메타 정보에 기반하여 상기 특징 데이터에 특화된 처리 기준을 산출하는 단계; 상기 처리 기준에 기반하여 상기 특징 데이터의 적어도 일부를 처리함으로써 상기 특징 데이터에 대응하는 제2 비식별 데이터를 생성하는 단계; 및 상기 제1 비식별 데이터를 키(key)로, 상기 제2 비식별 데이터 중 적어도

일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별 데이터를 생성하는 단계를 포함한다.

- [9] 상기 메타 정보는, 상기 특정 데이터에 포함되는 각 데이터의 종류를 정의하는 데이터 필드 정보 및 상기 각 데이터의 종류에 대응하는 데이터 처리 방식을 정의하는 데이터 비식별화 방식 정보를 포함할 수 있다.
- [10] 상기 데이터 필드 정보는, 상기 서비스 사용자의 성별, 나이, 서비스 페이지 방문 횟수, 서비스 내 주문 건수, 회원 개월 수 중 적어도 일부를 데이터의 종류로서 정의한 정보일 수 있다.
- [11] 상기 메타 정보는, 상기 특정 데이터에 포함되는 데이터 중 적어도 일부의 용도를 정의하는 데이터 용도 정보를 추가로 포함할 수 있고, 상기 데이터 비식별화 방식 정보는, 상기 각 데이터의 종류 및 상기 데이터 용도 정보에 대응하는 데이터 처리 방식을 정의한 정보일 수 있다.
- [12] 상기 데이터 비식별화 방식 정보는, 데이터 값을 이진 변환(binary convert)하는 제1 방식, 데이터 값을 각 구간별 대푯값으로 변환하는 제2 방식, 데이터 값을 근사값으로 변환하는 제3 방식, 데이터 값을 그대로 유지하는 제4 방식 중 적어도 하나를 데이터의 처리 방식으로 정의한 정보일 수 있다.
- [13] 상기 제1 비식별 데이터는, 상기 사용자 식별 데이터와 고유한 값을 갖는 부가 데이터를 기반으로 해시 함수를 통해 생성된 데이터를 포함할 수 있다.
- [14] 상기 부가 데이터는, UUID(Universally Unique Identifier)를 포함할 수 있고, 상기 해시 함수는, SHA(Secure Hash Algorithm) 기반 해시 함수를 포함할 수 있다.
- [15] 상기 UUID는, 타임스탬프 및 MAC 주소(Media Access Control Address)를 기반으로 생성되는 제1 버전 UUID, 상기 제1 버전 UUID를 구성하는 일부 데이터가 대체된 제2 버전 UUID, 상기 서비스 사용자의 이름 및 이름공간(namespace)에 기반하여 MD5(Message-Digest algorithm 5) 기반 해시 함수를 통해 생성된 제3 버전 UUID, 무작위로 생성된 제4 버전 UUID, 상기 서비스 사용자의 이름 및 이름공간에 기반하여 SHA 기반 해시 함수를 통해 생성된 제5 버전 UUID 중 어느 하나에 해당할 수 있다.
- [16] 사용자 식별 데이터는 동일하지만 최종 비식별 데이터가 생성되는 시점이 상이한 경우, 각각의 최종 비식별 데이터의 생성에 이용되는 부가 데이터는 상이할 수 있다.
- [17] 상기 제2 비식별 데이터는, 상기 처리 기준이 존재하는 데이터 필드에 속하는 특정 데이터가 처리 대상인 경우, 상기 처리 기준에 따라 생성되고, 상기 처리 기준이 존재하지 않는 데이터 필드에 속하는 특정 데이터가 처리 대상인 경우, 특정 데이터의 값을 이진 변환함으로써 생성될 수 있다.
- [18] 상기 최종 비식별 데이터는, 생성 시점으로 구별되는 스냅샷 데이터로서, 사전 설정된 데이터 관독 주기마다 확인된 하나 이상의 사용자 식별 데이터와 각각의 사용자 식별 데이터에 대응하는 하나 이상의 특정 데이터가 매칭된 하나 이상의 테이블을 포함할 수 있다.

- [19] 상기 특징 데이터는, 현재 가입 상태인 서비스 사용자의 특징 데이터 및 탈퇴 상태인 서비스 사용자의 특징 데이터에 대응하는 제2 비식별 데이터를 포함할 수 있다.
- [20] 일 실시 예에 따르면, 상기 데이터 처리 방법은, 탈퇴 상태인 서비스 사용자의 특징 데이터를 확인하는 단계; 및 현재 가입 상태인 서비스 사용자의 특징 데이터에 대해 사전 산출된 처리 기준에 기반하여 상기 탈퇴 상태인 서비스 사용자의 특징 데이터의 적어도 일부를 처리함으로써 상기 탈퇴 상태인 서비스 사용자의 제2 비식별 데이터를 생성하는 단계를 더 포함할 수 있다.
- [21] 다른 실시 예에 따르면, 상기 데이터 처리 방법은, 탈퇴 상태인 서비스 사용자의 특징 데이터를 확인하는 단계; 데이터의 처리 방식을 정의한 메타 정보에 기반하여 상기 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준을 산출하는 단계; 및 상기 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준에 기반하여 상기 탈퇴 상태인 서비스 사용자의 특징 데이터의 적어도 일부를 처리함으로써 상기 탈퇴 상태인 서비스 사용자의 제2 비식별 데이터를 생성하는 단계를 더 포함할 수도 있다.
- [22] 상기 제1 비식별 데이터를 생성하는 단계는, 탈퇴 상태인 서비스 사용자의 제1 비식별 데이터를 생성함에 있어서, 탈퇴 상태인 서비스 사용자의 사용자 식별 데이터에 해시 함수를 1차 적용하는 단계; 및 상기 1차 적용 결과와 상기 부가 데이터를 기반으로 해시 함수를 2차 적용하여 상기 탈퇴 상태인 서비스 사용자의 제1 비식별 데이터를 생성하는 단계를 포함할 수 있다.
- [23] 개시되는 일 실시 예에 따른 데이터 처리를 위한 전자 장치는, 트랜시버, 명령어를 저장하는 메모리 및 프로세서를 포함하고, 상기 프로세서는, 상기 트랜시버 및 상기 메모리와 연결되어, 서비스 사용자를 식별 가능한 사용자 식별 데이터 및 상기 사용자 식별 데이터에 대응하는 서비스 사용자의 특징 데이터를 확인하고, 해싱(Hashing) 기법을 통해 상기 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성하고, 데이터의 처리 방식을 정의한 메타 정보에 기반하여 상기 특징 데이터에 특화된 처리 기준을 산출하고, 상기 처리 기준에 기반하여 상기 특징 데이터의 적어도 일부를 처리함으로써 상기 특징 데이터에 대응하는 제2 비식별 데이터를 생성하고, 상기 제1 비식별 데이터를 키(key)로, 상기 제2 비식별 데이터 중 적어도 일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별 데이터를 생성한다.
- [24] 기타 실시 예들의 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

- [25] 본 개시에 따르면, 사용자를 식별 가능한 데이터에 대해서 해싱 기법을 통해 비식별화 처리를 하고, 특징을 나타내는 특징 데이터에 대해서 처리 기준에 따라 비식별화 처리를 하여, 두 비식별 데이터를 키-값 테이블로 매칭함으로써, 사용자와 그에 따른 특징을 나타내는 데이터가 매칭된 데이터 셋을 생성할 수

있다.

- [26] 또한 본 개시에 따르면, 탈퇴 상태인 사용자의 데이터를 사전에 1차적으로 비식별화 처리하여 저장함으로써, 탈퇴 회원에 적용되는 법적 규제를 충족하면서도 탈퇴 회원에 대한 데이터를 추후 활용할 수 있는 형태로 보관할 수 있다.
- [27] 발명의 효과는 이상에서 언급한 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 청구범위의 기재로부터 당해 기술 분야의 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [28] 도 1은 일 실시 예에 따른 전자 장치가 작동하는 환경을 나타낸 개략적인 구성도이다.
- [29] 도 2는 일 실시 예에 따른 데이터 처리 방법을 설명하기 위한 흐름도이다.
- [30] 도 3은 일 실시 예에 따라 데이터가 처리되는 파이프라인의 예시도이다.
- [31] 도 4 및 5는 추가적인 실시 예에 따른 데이터 처리 방법을 설명하기 위한 흐름도이다.
- [32] 도 6은 추가적인 실시 예에 따라 데이터가 처리되는 파이프라인의 예시도이다.
- [33] 도 7a 내지 7e는 일 실시 예에 따라 처리되는 데이터의 구조를 나타낸 예시도이다.
- [34] 도 8은 일 실시 예에 따라 데이터를 처리하는 전자 장치를 설명하기 위한 블록도이다.

발명의 실시를 위한 형태

- [35] 실시 예들에서 사용되는 용어는 본 개시에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어들을 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 판례, 새로운 기술의 출현 등에 따라 달라질 수 있다. 또한, 특정한 경우는 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 설명 부분에서 상세히 그 의미를 기재할 것이다. 따라서 본 개시에서 사용되는 용어는 단순한 용어의 명칭이 아닌, 그 용어가 가지는 의미와 본 개시의 전반에 걸친 내용을 토대로 정의되어야 한다.
- [36] 명세서 전체에서 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있음을 의미한다. 또한, 명세서에 기재된 "...부", "...모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하고, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있으며, 도시된 예와 달리 구체적 동작에 있어 명확히 구분되지 않을 수 있다.
- [37] 명세서 전체에서 기재된 "a, b, 및 c 중 적어도 하나"의 표현은, 'a 단독', 'b 단독', 'c 단독', 'a 및 b', 'a 및 c', 'b 및 c', 또는 'a, b, 및 c 모두'를 포괄할 수 있다.

- [38] 이하에서 언급되는 "단말" 또는 "사용자 단말"은 네트워크를 통해 서버나 타 단말에 접속할 수 있는 컴퓨터나 휴대용 단말로 구현될 수 있다. 여기서, 컴퓨터는 예를 들어, 웹 브라우저(WEB Browser)가 탑재된 노트북, 데스크톱(desktop), 랩톱(laptop) 등을 포함하고, 휴대용 단말은 예를 들어, 휴대성과 이동성이 보장되는 무선 통신 장치로서, IMT(International Mobile Telecommunication), CDMA(Code Division Multiple Access), W-CDMA(W-Code Division Multiple Access), LTE(Long Term Evolution) 등의 통신 기반 단말, 스마트폰, 태블릿 PC 등과 같은 모든 종류의 핸드헬드(Handheld) 기반의 무선 통신 장치를 포함할 수 있다.
- [39] 이하의 설명에 있어서, 신호, 메시지 또는 정보의 "전송", "통신", "송신", "수신" 기타 이와 유사한 의미의 용어는 일 구성요소에서 다른 구성요소로 정보, 메시지 또는 정보가 직접 전달되는 것뿐만이 아니라 다른 구성요소를 거쳐 전달되는 것도 포함한다.
- [40] 특히 신호, 메시지 또는 정보를 일 구성요소로 "전송" 또는 "송신"한다는 것은 그 신호, 메시지 또는 정보의 최종 목적지를 지시하는 것이고 직접적인 목적지를 의미하는 것이 아니다. 이는 신호, 메시지 또는 정보의 "수신"에 있어서도 동일하다. 또한 본 개시에 있어서, 2 이상의 데이터 또는 정보가 "관련"된다는 것은 하나의 데이터(또는 정보)를 획득하면, 그에 기초하여 다른 데이터(또는 정보)의 적어도 일부를 획득할 수 있음을 의미한다.
- [41] 또한, 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다.
- [42] 예를 들어, 본 개시의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.
- [43] 아래에서는 첨부한 도면을 참고하여 본 개시의 실시 예에 대하여 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다. 그러나 본 개시는 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시 예에 한정되지 않는다.
- [44] 이하에서는 도면을 참조하여 본 개시의 실시 예들을 상세히 설명한다.
- [45] 실시 예를 설명함에 있어서 본 발명이 속하는 기술 분야에 익히 알려져 있고 본 발명과 직접적으로 관련이 없는 기술 내용에 대해서는 설명을 생략한다. 이는 불필요한 설명을 생략함으로써 본 발명의 요지를 흐리지 않고 더욱 명확히 전달하기 위함이다.
- [46] 마찬가지로 이유로 첨부 도면에 있어서 일부 구성요소는 과장되거나 생략되거나 개략적으로 도시되었다. 또한, 각 구성요소의 크기는 실제 크기를 전적으로 반영하는 것이 아니다. 각 도면에서 동일한 또는 대응하는 구성요소에는 동일한

참조 번호를 부여하였다.

[47] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시 예들을 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시 예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시 예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 명세서 전체에 걸쳐 동일 참조 부호는 동일 구성 요소를 지칭한다.

[48] 처리 흐름도 도면들의 각 블록과 흐름도 도면들의 조합들은 컴퓨터 프로그램 인스트럭션들에 의해 수행될 수 있음을 이해할 수 있을 것이다. 이들 컴퓨터 프로그램 인스트럭션들은 범용 컴퓨터, 특수용 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서에 탑재될 수 있으므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비의 프로세서를 통해 수행되는 그 인스트럭션들이 흐름도 블록(들)에서 설명된 기능들을 수행하는 수단을 생성하게 된다. 이들 컴퓨터 프로그램 인스트럭션들은 특정 방식으로 기능을 구현하기 위해 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 지향할 수 있는 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 메모리에 저장되는 것도 가능하므로, 그 컴퓨터 이용 가능 또는 컴퓨터 판독 가능 메모리에 저장된 인스트럭션들은 흐름도 블록(들)에서 설명된 기능을 수행하는 인스트럭션 수단을 내포하는 제조 품목을 생산하는 것도 가능하다. 컴퓨터 프로그램 인스트럭션들은 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에 탑재되는 것도 가능하므로, 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비 상에서 일련의 동작 단계들이 수행되어 컴퓨터로 실행되는 프로세스를 생성해서 컴퓨터 또는 기타 프로그램 가능한 데이터 프로세싱 장비를 수행하는 인스트럭션들은 흐름도 블록(들)에서 설명된 기능들을 실행하기 위한 단계들을 제공하는 것도 가능하다.

[49] 또한, 각 블록은 특정된 논리적 기능(들)을 실행하기 위한 하나 이상의 실행 가능한 인스트럭션들을 포함하는 모듈, 세그먼트 또는 코드의 일부를 나타낼 수 있다. 또, 몇 가지 대체 실행 예들에서는 블록들에서 언급된 기능들이 순서를 벗어나서 발생하는 것도 가능함을 주목해야 한다. 예컨대, 잇달아 도시되어 있는 두 개의 블록들은 사실 실질적으로 동시에 수행되는 것도 가능하고 또는 그 블록들이 때때로 해당하는 기능에 따라 역순으로 수행되는 것도 가능하다.

[50] 도 1은 일 실시 예에 따른 전자 장치가 작동하는 환경을 나타낸 개략적인 구성도이다. 도 1을 참조하면, 전자 장치(110) 및 데이터베이스(120)는 연결된 네트워크(130)를 통해 서로 통신하거나, 외부의 다른 장치와 통신할 수 있다. 네트워크(130)는 근거리 통신망(Local Area Network; LAN), 광역 통신망(Wide Area Network; WAN), 부가가치 통신망(Value Added Network; VAN), 이동

통신망(mobile radio communication network), 위성 통신망 및 이들의 상호 조합을 포함하며, 도 1에 도시된 각 구성 주체가 서로 원활하게 통신을 할 수 있도록 하는 포괄적인 의미의 데이터 통신망이며, 유선 인터넷, 무선 인터넷 및 모바일 무선 통신망을 포함할 수 있다. 무선 통신은 예를 들어, 무선 랜(Wi-Fi), 블루투스, 블루투스 저 에너지(Bluetooth low energy), 지그비, WFD(Wi-Fi Direct), UWB(ultra wideband), 적외선 통신(IrDA, infrared Data Association), NFC(Near Field Communication) 등이 있을 수 있으나, 이에 한정되는 것은 아니다.

- [51] 전자 장치(110)는 데이터베이스(120)에 저장된 데이터 또는 그 사본을 비식별화하기 위한 장치이다. 본 개시에서 '비식별화'는 개인 정보를 비롯한 민감 데이터가 누구의 데이터인지 식별할 수 없도록 변경하는 방식, 메커니즘을 총칭하며, '식별할 수 없다'는 것은 데이터가 사전 설정된 기준에 따라 비식별화되거나, 비식별화와 관련하여 공인된 기관, 위원회 등의 승인을 받을 수 있는 상태가 된 것을 의미할 수 있다. 이와 같이 비식별화된 데이터는 인공지능 모델의 기계 학습에 사용될 수도 있다.
- [52] 전자 장치(110)는 비식별화를 위한 연산을 수행하는 하나의 서버일 수 있으나, 실시 예에 따라서는 복수의 서버로 구성될 수도 있으며, 하나 이상의 서버와 외부의 클라우드 서버가 전자적으로 통신 가능하도록 연결된 형태로 구성될 수도 있다.
- [53] 데이터베이스(120)는 서비스를 이용하는 사용자의 개인 정보가 저장되는 매체이다. 구체적으로, 데이터베이스(120)에는 사용자의 성별, 나이, 서비스 이용 횟수, 서비스 가입 일자, 서비스 이용 일자, 아이디, 비밀번호 또는 회원 시리얼 번호 등을 포함하는 민감 데이터가 저장될 수 있다. 관련하여, 데이터베이스(120)에 데이터가 저장되는 방식은 다양하게 설계될 수 있으며, 데이터베이스(120)는 서로 관련된 데이터 포인트에 대한 액세스를 저장 및 제공하는 관계형 데이터베이스(RDB, Relational Database)일 수 있으나, 실시 예에 따라서는 비관계형 데이터베이스(NoSQL)일 수도 있다.
- [54] 전자 장치(110)는 데이터베이스(120)로부터 데이터를 수신하거나, 데이터베이스(120)에 저장된 데이터를 읽을 수 있고, 해당 데이터를 이용하여 연산을 수행할 수 있다. 또한, 전자 장치(110)는 상기 연산 결과를 다시 데이터베이스(120)에 전송하거나, 전자 장치(110) 내 저장 공간에 저장하거나, 별도의 저장 매체에 전송할 수 있다. 비록, 도 1에는 데이터베이스(120)가 1개인 것처럼 도시되어 있으나 이는 이해의 편의를 돕기 위한 것으로, 실시 예에 따라 데이터베이스(120)의 수는 복수일 수 있다.
- [55] 이상과 관련하여, 이하 도면들을 통해 보다 상세히 설명하기로 한다. 도 2 내지 7e에 도시된 방법은 예를 들어, 상술한 전자 장치(110)에 의해 수행될 수 있다.
- [56] 도 2는 일 실시 예에 따른 데이터 처리 방법을 설명하기 위한 흐름도이다.
- [57] 단계 S210에서, 전자 장치(110)는 서비스 사용자를 식별 가능한 사용자 식별 데이터 및 사용자 식별 데이터에 대응하는 서비스 사용자의 특징 데이터를

확인한다. 본 개시에서, '사용자 식별 데이터'는 서비스 사용자마다 고유하게 부여된 회원 시리얼 번호, 회원 아이디 또는 그 밖의 식별자를 의미할 수 있다. 또한, '특징 데이터'는 서비스 사용자의 개인 정보 중 '사용자 식별 데이터'를 제외한 나머지 정보를 통칭하며, 예시적으로 사용자의 성별, 나이, 서비스 이용 횟수, 회원 개월 수 등을 포함할 수 있다.

- [58] 일 실시 예에 따르면, 전자 장치(110)는 데이터베이스(120)에서 사용자 식별 데이터 및 특징 데이터를 독출하거나, 이를 복사한 사본을 전자 장치(110)에 임시 저장할 수도 있다.
- [59] 일 실시 예에 따르면, 전자 장치(110)는 사전 설정된 데이터 판독 주기마다, 혹은 사전 설정된 판독 조건이 충족될 때마다, 특정 사이즈의 배치(batch)에 해당하는 사용자 식별 데이터 및 특징 데이터를 확인할 수 있다. 이때, 배치 사이즈는 사전에 설정되는 것일 수도 있고, 데이터 판독 기간에 따라 결정되는 것일 수도 있다. 이에 따라, 전자 장치(110)는 한번의 확인 시 둘 이상의 사용자에 대한 서로 다른 사용자 식별 데이터를 확인할 수 있고, 각 사용자에 대한 특징 데이터를 확인할 수도 있다.
- [60] 단계 S220에서, 전자 장치(110)는 해싱(Hashing) 기법을 통해 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성한다.
- [61] 일 실시 예에 따르면, 제1 비식별 데이터는 (1)사용자 식별 데이터와 (2)고유한 값을 갖는 부가 데이터를 기반으로 해시 함수를 통해 생성된 데이터를 포함할 수 있다. 예컨대, 전자 장치(110)는 사용자 식별 데이터와 부가 데이터를 병합하여 해시 함수에 입력함으로써 제1 비식별 데이터를 생성할 수 있다.
- [62] 상기 실시 예의 일 예시로서, 부가 데이터는 UUID(Universally Unique Identifier)를 포함할 수 있으며, 해시 함수는 SHA(Secure Hash Algorithm) 기반 해시 함수를 포함할 수 있다. 이에 따르면, 전자 장치(110)는 사용자 식별 데이터와 UUID를 병합하여 SHA 기반 해시 함수에 입력함으로써 제1 비식별 데이터를 생성할 수 있다. 대표적으로, 해시 함수는 SHA-256 기반의 해시 함수를 포함할 수 있으며, UUID는 다음의 5가지 버전의 UUID를 포함할 수 있다. 다만, 해시 함수 및 UUID의 종류는 본 개시에 기술된 것으로 한정되지 않음에 유의해야 한다.
- [63] (1) 제1 버전 UUID: 타임스탬프 및 MAC 주소(Media Access Control Address)를 기반으로 생성되는 UUID
- [64] (2) 제2 버전 UUID: 제1 버전 UUID를 구성하는 일부 데이터를 대체함으로써 생성되는 UUID
- [65] (3) 제3 버전 UUID: 서비스 사용자의 이름 및 이름공간(namespace)에 기반하여 MD5(Message-Digest algorithm 5) 기반 해시 함수를 통해 생성되는 UUID
- [66] (4) 제4 버전 UUID: 무작위로 생성되는 UUID
- [67] (5) 제5 버전 UUID: 서비스 사용자의 이름 및 이름공간에 기반하여 SHA 기반 해시 함수를 통해 생성되는 UUID

- [68] 한편 일 실시 예에 따르면, 제1 비식별 데이터의 생성에 이용되는 부가 데이터와 관련하여, 사용자 식별 데이터는 동일하지만 최종 비식별 데이터가 생성되는 시점이 상이한 경우, 각각의 최종 비식별 데이터의 생성에 이용되는 부가 데이터는 상이할 수 있다. 즉, 전자 장치(110)는 동일한 서비스 사용자의 개인 정보를 비식별화 할지라도, 비식별화 시점(최종 비식별 데이터가 생성되는 시점)에 따라서 상이한 부가 데이터를 이용함으로써, 비식별화 시점에 따라 최종 비식별 데이터를 다르게 생성할 수 있고, 이로써 서비스 사용자를 기준으로 한 비식별화 효과에 비식별화 시점을 기준으로 한 비식별화 효과가 더해질 수 있다.
- [69] 단계 S230에서, 전자 장치(110)는 데이터의 처리 방식을 정의한 메타 정보에 기반하여 특정 데이터에 특화된 처리 기준을 산출한다.
- [70] 일 실시 예에 따르면, 메타 정보는 (1)특정 데이터에 포함되는 각 데이터의 종류를 정의하는 데이터 필드 정보 및 (2)각 데이터의 종류에 대응하는 데이터 처리 방식을 정의하는 데이터 비식별화 방식 정보를 포함할 수 있다. 이때, 데이터 필드 정보 및 데이터 비식별화 방식 정보를 포함하는 메타 정보는 최종 비식별 데이터의 소비자(consumer)가 생성한 설정 정보 또는 전자 장치(110)가 소비자를 고려하여 자체 생성한 설정 정보에 따라 정의될 수 있다.
- [71] 관련하여, 데이터 필드 정보는 서비스 사용자의 성별, 나이, 서비스 페이지 방문 횟수, 서비스 내 주문 건수, 회원 개월 수 중 적어도 일부를 데이터의 종류로서 정의한 정보를 포함할 수 있다. 구체적으로, 데이터 필드 정보는 각 데이터가 속한 열(column) 별로 정의될 수 있다.
- [72] 한편, 메타 정보는 특정 데이터에 포함되는 데이터 중 적어도 일부의 용도를 정의하는 데이터 용도 정보를 추가로 포함할 수 있고, 이때 데이터 비식별화 방식 정보는 '각 데이터의 종류 및 데이터 용도 정보'에 대응하는 데이터 처리 방식을 정의한 정보를 포함할 수 있다.
- [73] 한편, 데이터 비식별화 방식 정보는 데이터의 처리 방식을 다음과 같이 정의한 정보를 포함할 수 있다. 다만, 데이터의 처리 방식이 아래 열거된 방식에 한정되는 것은 아니다. 또한 데이터 비식별화 방식 정보는 전자 장치(110)의 관리자에 의해 정의될 수도 있고, 전자 장치(110) 자체적으로 각 데이터의 종류에 따라 정의한 것일 수도 있다.
- [74] (1) 제1 방식: 데이터 값을 이진 변환(binary convert)
- [75] (2) 제2 방식: 데이터 값을 각 구간별 대푯값으로 변환
- [76] (3) 제3 방식: 데이터 값을 근사값으로 변환(예컨대, truncation)
- [77] (4) 제4 방식: 데이터 값을 그대로 유지(bypass)
- [78] 각각의 방식을 보다 구체적으로 서술하면, 제1 방식은 두 개의 데이터 값을 갖는 데이터를 데이터 값에 따라 0 또는 1로 변환하거나, 셋 이상의 데이터 값을 갖는 데이터를 임계 값(threshold)을 기준으로 0 또는 1로 변환하는 방식일 수 있다. 또한 제2 방식은 데이터 값 구간을 복수의 분위로 나눈 뒤, 각각의 데이터 값을 해당 데이터 값이 속하는 분위 구간의 중간 값, 평균값, 최소값 또는 최대값

등의 대푯값으로 변환하는 방식일 수 있다. 예컨대, 제2 방식의 일 예시인 50분위 중위 값 방식에 따르면, 데이터 값 구간을 50개로 나누어 각 데이터 값이 속하는 분위 구간의 중위 값으로 해당 데이터 값이 대체될 수 있다. 한편 제3 방식은 데이터 값의 특정 자릿수 이하를 버림/올림 하거나, 데이터 값에 특정 함수를 적용하여 근사값을 구하는 방식일 수 있다.

[79] 단계 S240에서, 전자 장치(110)는 산출된 처리 기준에 기반하여 특징 데이터의 적어도 일부를 처리함으로써 특징 데이터에 대응하는 제2 비식별 데이터를 생성한다. 즉 다시 말하면, 전자 장치(110)는 처리 기준이 적용되는 데이터 필드에 속한 특징 데이터를 대상으로, 특징 데이터를 처리 기준에 따라 변환하여 제2 비식별 데이터를 생성할 수 있다.

[80] 다만, 특징 데이터들이 속하는 모든 데이터 필드 중 일부에 대해서는 처리 기준이 산출되지 않을 수 있으므로, 전자 장치(110)는 처리 기준의 존재 여부에 따라서 제2 비식별 데이터의 생성 방식을 달리할 수 있다. 예컨대, 전자 장치(110)는 처리 기준이 존재하는 데이터 필드에 속하는 특징 데이터가 처리 대상인 경우, 처리 기준에 따라 제2 비식별 데이터를 생성할 수 있고, 처리 기준이 존재하지 않는 데이터 필드에 속하는 특징 데이터가 처리 대상인 경우, 특징 데이터의 값을 이진 변환함으로써 제2 비식별 데이터를 생성할 수 있다.

[81] 단계 S250에서, 전자 장치(110)는 제1 비식별 데이터를 키(key)로, 제2 비식별 데이터 중 적어도 일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별 데이터를 생성한다. 예컨대, 최종 비식별 데이터에 테이블 A, B, C, D, E가 포함되는 경우, 해당 테이블 5개는 사용자 X에 대응하는 제1 비식별 데이터를 공통된 키로 갖되, 제2 비식별 데이터의 조합을 달리할 수 있다. 다른 예로는, 테이블 A, B, C, D, E 중 A, B는 사용자 X에 대응하는 제1 비식별 데이터를 공통된 키로 갖되 제2 비식별 데이터의 조합을 달리하고, C, D, E는 사용자 Y에 대응하는 제1 비식별 데이터를 공통된 키고 갖되 제2 비식별 데이터의 조합을 달리할 수도 있다. 이로써, 전자 장치(110)의 관리자는 최종 비식별 데이터 중 특정 사용자에게 대응하는 제1 비식별 데이터를 키로 하여, 해당 키와 연관된 값(제2 비식별 데이터)을 일괄적으로 탐색할 수 있게 된다.

[82] 일 실시 예에 따르면, 전자 장치(110)가 생성하는 최종 비식별 데이터는 생성 시점으로 구별되는 스냅샷(snapshot) 데이터일 수 있다. 구체적으로, 최종 비식별 데이터는 사전 설정된 데이터 관독 주기마다 확인된 하나 이상의 사용자 식별 데이터와 각각의 사용자 식별 데이터에 대응하는 하나 이상의 특징 데이터가 매칭된 하나 이상의 테이블을 포함할 수 있다. 이 경우, 서로 다른 시점에 생성된 스냅샷 데이터끼리는 공통된 키로 연관되지 않게 된다.

[83] 이하에서는, 도 3을 참조하여 도 2의 데이터 처리 방법을 보다 직관적으로 설명하기로 한다.

[84] 도 3은 일 실시 예에 따라 데이터가 처리되는 파이프라인의 예시도이다. 사각형으로 도시된 프로세스는 전자 장치(110)의 동작을 나타낸 것이며, 원통

도형으로 도시된 정보는 처리 대상이 되는 데이터나, 그 데이터의 처리에 이용되는 데이터를 나타낸 것이다. 다만, 각 데이터가 별개의 원통 도형으로 도시된 것은 설명의 편의를 돕기 위한 것이지 각 데이터가 저장되는 데이터베이스가 별개임을 의미하는 것은 아니다.

- [85] 일 실시 예에 따르면, 전자 장치(110)는 서비스 사용자를 식별 가능한 사용자 식별 데이터 및 특징 데이터를 확인하고, 메타 정보에 기반하여 특징 데이터에 특화된 처리 기준을 산출할 수 있다. 이때, 전자 장치(110)는 특징 데이터의 분포를 추가적으로 고려하여 특징 데이터에 특화된 처리 기준을 산출할 수 있다. 다만, 이러한 특징 데이터의 분포는 전자 장치(110)가 특징 데이터를 확인하면서 함께 확인한 정보일 수도 있고, 사전에 메타 정보에 포함된 정보일 수도 있다.
- [86] 일 실시 예에 따르면, 전자 장치(110)는 산출된 처리 기준에 따라 특징 데이터의 적어도 일부를 처리하여 제2 비식별 데이터를 생성할 수 있고, 해싱 기법을 통해 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성할 수 있다. 이때, 도 3에는 제1 비식별 데이터 생성 프로세스가 처리 기준 산출 프로세스 및 제2 비식별 데이터 생성 프로세스보다 후순위로 수행되는 것처럼 도시되어 있으나, 이는 설명의 편의를 위한 것으로, 실시 예에 따라서 제1 비식별 데이터 생성 프로세스는 전자 장치(110)가 사용자 식별 데이터를 확인한 이후라면 다른 프로세스와의 순서 선후를 따지지 않고 수행될 수 있다.
- [87] 도 4 및 5는 추가적인 실시 예에 따른 데이터 처리 방법을 설명하기 위한 흐름도이다. 본 개시에 따른 데이터 처리 방법은 데이터베이스(120)에 저장된 개인 정보 또는 그 사본을 비식별화하기 위함이지만, 탈퇴한 사용자의 개인 정보는 법적으로 규제된 약관 사항을 고려할 때 장기간 보관하는 것에 어려움이 있으므로, 미리 1차적으로 비식별화 과정을 거쳐 저장되어야 한다.
- [88] 이러한 취지에서, 도 4는 탈퇴 상태인 서비스 사용자의 사용자 식별 데이터를 처리하는 방법을 나타낸 것이다. 구체적으로는 사용자 식별 데이터를 비식별화하여 제1 비식별 데이터를 생성하는 과정을 나타낸 것으로, 단계 S410은 단계 S210 이전에 수행되고 단계 S420은 단계 S210 이후에 수행될 수 있으나, 각 단계의 수행 순서가 이에 한정되는 것은 아니다.
- [89] 단계 S410에서, 전자 장치(110)는 탈퇴 상태인 서비스 사용자의 사용자 식별 데이터에 해시 함수를 1차 적용한다. 이때 적용되는 해시 함수의 종류에는 특별한 제약은 없으나, 대표적으로 SHA 기반의 해시 함수가 적용될 수 있다.
- [90] 단계 S420에서, 전자 장치(110)는 단계 S410을 통한 1차 적용 결과와 부가 데이터를 기반으로 해시 함수를 2차 적용하여 탈퇴 상태인 서비스 사용자의 제1 비식별 데이터를 생성한다. 구체적으로, 전자 장치(110)는 1차 적용 결과와 부가 데이터를 병합하여 해시 함수에 입력함으로써 탈퇴 상태인 서비스 사용자의 제1 비식별 데이터를 생성할 수 있다. 이때 적용되는 해시 함수의 종류 역시 특별한 제약은 없으며, 대표적으로 SHA 기반의 해시 함수가 적용될 수 있다.
- [91] 일 실시 예에 따르면, 부가 데이터는 UUID를 포함할 수 있으며, UUID는 도 2를

참조하여 상술한 5가지 버전의 UUID를 포함할 수 있다. 다만, 해시 함수 및 UUID의 종류는 본 개시에 기술된 것으로 한정되지 않음에 유의해야 한다.

- [92] 한편, 도 5는 탈퇴 상태인 서비스 사용자의 특징 데이터를 처리하는 방법을 나타낸 것이다. 탈퇴한 사용자의 특징 데이터는 사전에 비식별화되어 도 2의 단계 S210을 거쳐야 하므로, 도 5는 이러한 취지에서 탈퇴 상태인 서비스 사용자의 특징 데이터에 대한 전처리 과정을 나타낸다. 이 경우, 도 2의 단계 S210에서의 '특징 데이터'는 (1)현재 가입 상태인 서비스 사용자의 특징 데이터 및 (2)탈퇴 상태인 서비스 사용자의 특징 데이터에 대응하는 제2 비식별 데이터를 포함할 수 있다.
- [93] 단계 S510에서, 전자 장치(110)는 탈퇴 상태인 서비스 사용자의 특징 데이터를 확인한다.
- [94] 일 실시 예에 따르면, 전자 장치(110)는 사용자의 특징 데이터와 병합된 사용자 상태 정보를 참조하여 사용자가 탈퇴 상태인지, 가입 상태인지 파악할 수 있다. 다른 실시 예에 따르면, 현재 가입 상태인 사용자의 특징 데이터가 저장된 데이터베이스와 탈퇴 상태인 사용자의 특징 데이터가 저장된 데이터베이스가 별도로 존재하거나, 하나의 데이터베이스 내에서 현재 가입 상태인 사용자의 특징 데이터의 저장 위치와 탈퇴 상태인 사용자의 특징 데이터의 저장 위치가 다를 수 있고, 전자 장치(110)는 두 특징 데이터의 저장 매체 또는 저장된 위치에 기반하여 사용자가 탈퇴 상태인지, 가입 상태인지 파악할 수도 있다. 다만, 전자 장치(110)가 사용자의 상태를 파악하는 방식은 이에 한정되지 않는다.
- [95] 단계 S520에서, 전자 장치(110)는 데이터의 처리 방식을 정의한 메타 정보에 기반하여 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준을 산출한다.
- [96] 단계 S530에서, 전자 장치(110)는 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준에 기반하여 탈퇴 상태인 서비스 사용자의 특징 데이터의 적어도 일부를 처리함으로써 탈퇴 상태인 서비스 사용자의 제2 비식별 데이터를 생성한다.
- [97] 한편, 전자 장치(110)는 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준을 이용하지 않고, 현재 가입 상태인 서비스 사용자의 특징 데이터를 처리하기 위해 사전 산출된 처리 기준을 탈퇴 상태인 서비스 사용자의 특징 데이터 처리 시 이용할 수도 있다. 이 경우, 상술한 단계 S520은 생략될 수 있으며, 단계 S530에서 전자 장치(110)는 사전 산출된 처리 기준에 기반하여 탈퇴 상태인 서비스 사용자의 제2 비식별 데이터를 생성할 수 있다.
- [98] 이상의 도시된 흐름도 도 2, 4, 5에서는 상기 방법을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 부가되어 수행될 수 있다.
- [99] 도 6은 추가적인 실시 예에 따라 데이터가 처리되는 파이프라인의 예시도이다.

구체적으로, 도 6 상단에는 탈퇴 상태인 사용자의 데이터를 1차적으로 비식별화 처리하는 과정이 도시되어 있다.

[100] 일 실시 예에 따르면, 전자 장치(110)는 탈퇴 상태인 사용자의 사용자 식별 데이터 및 특징 데이터를 확인하고, 탈퇴 상태인 사용자의 특징 데이터에 특화된 처리 기준을 산출하여, 해당 처리 기준에 따라 탈퇴 상태인 사용자의 특징 데이터에 대응하는 제2 비식별 데이터를 생성할 수 있다.

[101] 다른 실시 예에 따르면, 전자 장치(110)는 탈퇴 상태인 사용자의 사용자 식별 데이터 및 특징 데이터를 확인하고, 도 6에 도시된 것과는 달리 별도의 처리 기준을 산출하지 않고, 현재 가입 상태인 사용자의 특징 데이터를 처리하기 위해 사전 산출된 처리 기준을 이용하여 탈퇴 상태인 사용자의 특징 데이터에 대응하는 제2 비식별 데이터를 생성할 수 있다.

[102] 한편 도 6에서 볼 수 있듯이, 탈퇴 상태인 사용자의 데이터는 1차적인 비식별화 과정을 거쳐 다시 도 3의 파이프라인 상의 도입 프로세스에서 '특징 데이터'로서 이용될 수 있다.

[103] 또한 도 6의 하단에 도시된 대로, 전자 장치(110)는 외부로부터의 탈퇴 관련 데이터 제거 명령에 따라, 또는 자체 판단에 따라 가입 상태인 사용자의 데이터, 사용자 식별 데이터, 특징 데이터 또는 제2 비식별 데이터 중에서 탈퇴한 사용자에 대한 데이터를 제거할 수 있다. 구체적으로, 가입 상태인 사용자 중에서 일부가 서비스를 탈퇴하는 경우, 전자 장치(110)는 가입 상태인 사용자의 데이터에서 새로이 서비스를 탈퇴한 사용자의 데이터를 제거할 수 있다. 또한, 이미 처리 과정에 있는 사용자 식별 데이터, 특징 데이터 또는 제2 비식별 데이터 등에서도 새로이 서비스를 탈퇴한 사용자의 사용자 식별 데이터를 기반으로 하여 탈퇴한 사용자의 데이터를 제거할 수 있다.

[104] 도 7a 내지 7e는 일 실시 예에 따라 처리되는 데이터 구조를 나타낸 예시도이다.

[105] 먼저 도 7a에는 비식별화 처리를 거치지 않은 데이터의 구조가 도시되어 있다.

각 데이터의 생성 연/월에 대한 정보, 서비스 사용자를 식별 가능한 시리얼 번호로서의 사용자 식별 데이터, 사용자의 성별을 나타내는 이진수(설명의 편의상 0을 여성, 1을 남성에 해당하는 이진수로 전제한다), 사용자의 나이, 사용자가 회원 가입 후 경과된 개월 수가 나타나 있다. 도시된 대로, 시리얼 번호, 각 사용자의 나이, 회원 개월 수 등의 민감한 개인 정보가 그대로 드러나 있다.

[106] 한편 도 7b에는 일부 데이터에 대해 비식별화 처리가 진행된 데이터의 구조가 도시되어 있다. 즉, 개인 정보 중 시리얼 번호에 대해 해시 함수가 적용되어 일정한 비트 수의 제1 비식별 데이터로 변환되었다. 주목할 것은, 아래 두 행의 제1 비식별 데이터는 그 값이 동일하므로 동일한 서비스 사용자에 대한 사용자 식별 데이터가 변환된 것임을 유추할 수 있으며, 상단 행의 제1 비식별 데이터는 별도 사용자의 사용자 식별 데이터가 변환된 것임을 알 수 있다.

[107] 한편 도 7c에는 비식별화 처리된 각 테이블이 어떤 상태의 사용자에 대한 데이터를 비식별화 처리한 것인지 나타내는 데이터 구조가 도시되어 있다.

도시된 대로, 각 테이블은 모두 현재 가입 상태인 사용자의 데이터에 대해서는 비식별화 처리가 되었고, 탈퇴 상태인 사용자의 데이터에 대해서는 비식별화 처리를 수행하지 않았음을 알 수 있다. 즉 일 실시 예에 따르면, 전자 장치(110)는 가입 상태인 사용자와 탈퇴 상태인 사용자 각각에 대한 비식별화 처리 여부를 나타내는 정보를 참조하여, 비식별화 처리 대상인 상태의 사용자의 데이터에 대해서만 본 개시에 따른 비식별화 처리를 수행할 수 있다. 또한 다른 실시 예에 따르면, 전자 장치(110)는 일련의 비식별화 처리 과정을 수행한 이후, 도 7c에 도시된 대로 비식별화 처리 여부를 나타내는 데이터 구조를 저장하고, 이를 참조하여 관리자에게 리포트를 전송하거나, 알림을 송출하거나, 비식별화 처리되지 않은(false 상태에 해당) 데이터에 대한 비식별화 처리 명령을 요청할 수도 있다.

[108] 한편 도 7d에는 각 테이블의 비식별화 처리에 적용된 처리 기준을 산출하는 데 이용된 데이터 필드 정보 및 데이터 비식별화 방식 정보가 도시되어 있다.

Quantile200은 200분위 중위 값 방식으로서, 1행의 테이블 내 order_cate2_fashion_r3m 필드에 속한 데이터, 2행의 테이블 내 order_cate1_fashion_r12m 필드에 속한 데이터, 3행의 테이블 내 gmv_payltr_r12m 필드에 속한 데이터에 대해서 200분위 중위 값 방식이 적용되었음을 나타낸다.

[109] 한편 도 7e에는 데이터 필드 정보 및 데이터 비식별화 방식 정보에 기반하여 산출된, 특정 데이터에 특화된 처리 기준이 도시되어 있다. 구체적으로, 동일한 sb_rocketpay.csm_AA_category2_sample 테이블의 order_cate2_fashion_kids_r3m 필드에 속한 데이터, order_cate1_special1_r12m 필드에 속한 데이터, order_cate2_daily_supply_r12m 필드에 속한 데이터에 대해서 200분위 중위 값 방식(quantile200)이 적용되어, 산출된 분위 값이 5월에 기재되어 있다. 즉, 전자 장치(110)는 동일한 테이블 내에서도 메타 정보(데이터 필드 정보, 데이터 비식별화 방식 정보)를 고려하여 각 필드 별로 특화된 처리 기준을 산출할 수 있다.

[110] 도 8은 일 실시 예에 따라 데이터를 처리하는 전자 장치를 설명하기 위한 블록도이다.

[111] 전자 장치(110)는 일 실시 예에 따라, 트랜시버(111), 프로세서(113) 및 메모리(115)를 포함할 수 있다. 일 실시 예에서, 전자 장치(110)는 트랜시버(111)를 통하여 데이터베이스(120)와 연결되고, 데이터를 교환할 수 있다.

[112] 프로세서(113)는 도 1 내지 도 7e를 통하여 전술한 적어도 하나의 방법을 수행할 수 있다. 메모리(115)는 도 1 내지 도 7e를 통하여 전술한 적어도 하나의 방법을 수행하기 위한 정보를 저장할 수 있다. 메모리(115)는 휘발성 메모리 또는 비휘발성 메모리일 수 있다.

[113] 프로세서(113)는 프로그램을 실행하고, 정보를 제공하기 위해 전자 장치(110)를 제어할 수 있다. 프로세서(113)에 의하여 실행되는 프로그램의

코드는 메모리(115)에 저장될 수 있다.

- [114] 프로세서(113)는 트랜시버(111) 및 메모리(115)와 연결되어, 서비스 사용자를 식별 가능한 사용자 식별 데이터 및 사용자 식별 데이터에 대응하는 서비스 사용자의 특징 데이터를 확인하고, 해싱(Hashing) 기법을 통해 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성하고, 데이터의 처리 방식을 정의한 메타 정보에 기반하여 특징 데이터에 특화된 처리 기준을 산출하고, 처리 기준에 기반하여 특징 데이터의 적어도 일부를 처리함으로써 특징 데이터에 대응하는 제2 비식별 데이터를 생성하고, 제1 비식별 데이터를 키(key)로, 제2 비식별 데이터 중 적어도 일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별 데이터를 생성할 수 있다.
- [115] 도 8에 도시된 전자 장치(110)는 본 실시 예와 관련된 구성요소들만이 도시되어 있다. 따라서, 도 8에 도시된 구성요소들 외에 다른 범용적인 구성요소들이 더 포함될 수 있음을 본 실시 예와 관련된 기술분야에서 통상의 지식을 가진 자라면 이해할 수 있다.
- [116] 전술한 실시 예들에 따른 장치는 프로세서, 프로그램 데이터를 저장하고 실행하는 메모리, 디스크 드라이브와 같은 영구 저장부(permanent storage), 외부 장치와 통신하는 통신 포트, 터치 패널, 키(key), 버튼 등과 같은 사용자 인터페이스 장치 등을 포함할 수 있다. 소프트웨어 모듈 또는 알고리즘으로 구현되는 방법들은 상기 프로세서상에서 실행 가능한 컴퓨터가 읽을 수 있는 코드들 또는 프로그램 명령들로서 컴퓨터가 읽을 수 있는 기록 매체 상에 저장될 수 있다. 여기서 컴퓨터가 읽을 수 있는 기록 매체로 마그네틱 저장 매체(예컨대, ROM(read-only memory), RAM(random-access memory), 플로피 디스크, 하드 디스크 등) 및 광학적 판독 매체(예컨대, 시디롬(CD-ROM), 디브이디(DVD: Digital Versatile Disc)) 등이 있다. 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템들에 분산되어, 분산 방식으로 컴퓨터가 판독 가능한 코드가 저장되고 실행될 수 있다. 매체는 컴퓨터에 의해 판독가능하며, 메모리에 저장되고, 프로세서에서 실행될 수 있다.
- [117] 본 실시 예는 기능적인 블록 구성들 및 다양한 처리 단계들로 나타내어질 수 있다. 이러한 기능 블록들은 특정 기능들을 실행하는 다양한 개수의 하드웨어 또는/및 소프트웨어 구성들로 구현될 수 있다. 예를 들어, 실시 예는 하나 이상의 마이크로프로세서들의 제어 또는 다른 제어 장치들에 의해서 다양한 기능들을 실행할 수 있는, 메모리, 프로세싱, 로직(logic), 룩 업 테이블(look-up table) 등과 같은 직접 회로 구성들을 채용할 수 있다. 구성 요소들이 소프트웨어 프로그래밍 또는 소프트웨어 요소들로 실행될 수 있는 것과 유사하게, 본 실시 예는 데이터 구조, 프로세스들, 루틴들 또는 다른 프로그래밍 구성들의 조합으로 구현되는 다양한 알고리즘을 포함하여, C, C++, 자바(Java), 어셈블러(assembler) 등과 같은 프로그래밍 또는 스크립팅 언어로 구현될 수 있다. 기능적인 측면들은 하나 이상의 프로세서들에서 실행되는 알고리즘으로 구현될 수 있다. 또한, 본 실시

예는 전자적인 환경 설정, 신호 처리, 메시지 처리, 및/또는 데이터 처리 등을 위하여 종래 기술을 채용할 수 있다. “매커니즘”, “요소”, “수단”, “구성”과 같은 용어는 넓게 사용될 수 있으며, 기계적이고 물리적인 구성들로서 한정되는 것은 아니다. 상기 용어는 프로세서 등과 연계하여 소프트웨어의 일련의 처리들(routines)의 의미를 포함할 수 있다.

- [118] 전술한 실시 예들은 일 예시일 뿐 후술하는 청구항들의 범위 내에서 다른 실시 예들이 구현될 수 있다.

청구범위

- [청구항 1] 전자 장치에서의 데이터 비식별화를 위한 처리 방법에 있어서,
서비스 사용자를 식별 가능한 사용자 식별 데이터 및 상기 사용자 식별
데이터에 대응하는 서비스 사용자의 특징 데이터를 확인하는 단계;
해싱(Hashing) 기법을 통해 상기 사용자 식별 데이터에 대응하는 제1
비식별 데이터를 생성하는 단계;
데이터의 처리 방식을 정의한 메타 정보에 기반하여 상기 특징 데이터에
특화된 처리 기준을 산출하는 단계;
상기 처리 기준에 기반하여 상기 특징 데이터의 적어도 일부를
처리함으로써 상기 특징 데이터에 대응하는 제2 비식별 데이터를
생성하는 단계; 및
상기 제1 비식별 데이터를 키(key)로, 상기 제2 비식별 데이터 중 적어도
일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별
데이터를 생성하는 단계를 포함하는, 데이터 처리 방법.
- [청구항 2] 제1항에 있어서,
상기 메타 정보는,
상기 특징 데이터에 포함되는 각 데이터의 종류를 정의하는 데이터 필드
정보 및 상기 각 데이터의 종류에 대응하는 데이터 처리 방식을 정의하는
데이터 비식별화 방식 정보를 포함하는, 데이터 처리 방법.
- [청구항 3] 제2항에 있어서,
상기 데이터 필드 정보는,
상기 서비스 사용자의 성별, 나이, 서비스 페이지 방문 횟수, 서비스 내
주문 건수, 회원 개월 수 중 적어도 일부를 데이터의 종류로서 정의하는,
데이터 처리 방법.
- [청구항 4] 제2항에 있어서,
상기 메타 정보는,
상기 특징 데이터에 포함되는 데이터 중 적어도 일부의 용도를 정의하는
데이터 용도 정보를 추가로 포함하고,
상기 데이터 비식별화 방식 정보는,
상기 각 데이터의 종류 및 상기 데이터 용도 정보에 대응하는 데이터 처리
방식을 정의하는, 데이터 처리 방법.
- [청구항 5] 제2항에 있어서,
상기 데이터 비식별화 방식 정보는,
데이터 값을 이진 변환(binary convert)하는 제1 방식, 데이터 값을 각
구간별 대푯값으로 변환하는 제2 방식, 데이터 값을 근사값으로 변환하는
제3 방식, 데이터 값을 그대로 유지하는 제4 방식 중 적어도 하나를
데이터의 처리 방식으로서 정의하는, 데이터 처리 방법.

- [청구항 6] 제1항에 있어서,
상기 제1 비식별 데이터는,
상기 사용자 식별 데이터와 고유한 값을 갖는 부가 데이터를 기반으로
해시 함수를 통해 생성된 데이터를 포함하는, 데이터 처리 방법.
- [청구항 7] 제6항에 있어서,
상기 부가 데이터는,
UUID(Universally Unique Identifier)를 포함하고,
상기 해시 함수는,
SHA(Secure Hash Algorithm) 기반 해시 함수를 포함하는, 데이터 처리
방법.
- [청구항 8] 제7항에 있어서,
상기 UUID는,
타임스탬프 및 MAC 주소(Media Access Control Address)를 기반으로
생성되는 제1 버전 UUID, 상기 제1 버전 UUID를 구성하는 일부 데이터가
대체된 제2 버전 UUID, 상기 서비스 사용자의 이름 및
이름공간(namespace)에 기반하여 MD5(Message-Digest algorithm 5) 기반
해시 함수를 통해 생성된 제3 버전 UUID, 무작위로 생성된 제4 버전
UUID, 상기 서비스 사용자의 이름 및 이름공간에 기반하여 SHA 기반
해시 함수를 통해 생성된 제5 버전 UUID 중 어느 하나인, 데이터 처리
방법.
- [청구항 9] 제6항에 있어서,
상기 부가 데이터는,
사용자 식별 데이터는 동일하지만 최종 비식별 데이터가 생성되는
시점이 상이한 경우, 각각의 최종 비식별 데이터의 생성에 이용되는 부가
데이터는 상이한 것을 특징으로 하는, 데이터 처리 방법.
- [청구항 10] 제1항에 있어서,
상기 제2 비식별 데이터는,
상기 처리 기준이 존재하는 데이터 필드에 속하는 특징 데이터가 처리
대상인 경우, 상기 처리 기준에 따라 생성되고, 상기 처리 기준이
존재하지 않는 데이터 필드에 속하는 특징 데이터가 처리 대상인 경우,
특징 데이터의 값을 이진 변환함으로써 생성되는, 데이터 처리 방법.
- [청구항 11] 제1항에 있어서,
상기 최종 비식별 데이터는,
생성 시점으로 구별되는 스냅샷 데이터로서, 사전 설정된 데이터 관독
주기마다 확인된 하나 이상의 사용자 식별 데이터와 각각의 사용자 식별
데이터에 대응하는 하나 이상의 특징 데이터가 매칭된 하나 이상의
테이블을 포함하는, 데이터 처리 방법.
- [청구항 12] 제1항에 있어서,

상기 특징 데이터는,
현재 가입 상태인 서비스 사용자의 특징 데이터 및 탈퇴 상태인 서비스 사용자의 특징 데이터에 대응하는 제2 비식별 데이터를 포함하는, 데이터 처리 방법.

[청구항 13] 제12항에 있어서,
탈퇴 상태인 서비스 사용자의 특징 데이터를 확인하는 단계; 및
현재 가입 상태인 서비스 사용자의 특징 데이터에 대해 사전 산출된 처리 기준에 기반하여 상기 탈퇴 상태인 서비스 사용자의 특징 데이터의 적어도 일부를 처리함으로써 상기 탈퇴 상태인 서비스 사용자의 제2 비식별 데이터를 생성하는 단계를 더 포함하는, 데이터 처리 방법.

[청구항 14] 제12항에 있어서,
탈퇴 상태인 서비스 사용자의 특징 데이터를 확인하는 단계;
데이터의 처리 방식을 정의한 메타 정보에 기반하여 상기 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준을 산출하는 단계; 및
상기 탈퇴 상태인 서비스 사용자의 특징 데이터에 특화된 처리 기준에 기반하여 상기 탈퇴 상태인 서비스 사용자의 특징 데이터의 적어도 일부를 처리함으로써 상기 탈퇴 상태인 서비스 사용자의 제2 비식별 데이터를 생성하는 단계를 더 포함하는, 데이터 처리 방법.

[청구항 15] 제1항에 있어서,
상기 제1 비식별 데이터를 생성하는 단계는,
탈퇴 상태인 서비스 사용자의 제1 비식별 데이터를 생성함에 있어서,
탈퇴 상태인 서비스 사용자의 사용자 식별 데이터에 해시 함수를 1차 적용하는 단계; 및
상기 1차 적용 결과와 상기 부가 데이터를 기반으로 해시 함수를 2차 적용하여 상기 탈퇴 상태인 서비스 사용자의 제1 비식별 데이터를 생성하는 단계를 포함하는, 데이터 처리 방법.

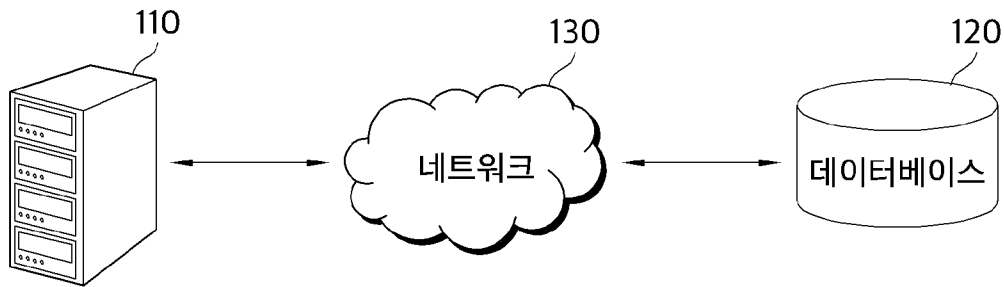
[청구항 16] 제1항의 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 비일시적 기록 매체.

[청구항 17] 데이터 비식별화를 위한 전자 장치로서,
트랜시버, 명령어를 저장하는 메모리 및 프로세서를 포함하고,
상기 프로세서는, 상기 트랜시버 및 상기 메모리와 연결되어,
서비스 사용자를 식별 가능한 사용자 식별 데이터 및 상기 사용자 식별 데이터에 대응하는 서비스 사용자의 특징 데이터를 확인하고,
해싱(Hashing) 기법을 통해 상기 사용자 식별 데이터에 대응하는 제1 비식별 데이터를 생성하고,
데이터의 처리 방식을 정의한 메타 정보에 기반하여 상기 특징 데이터에 특화된 처리 기준을 산출하고,
상기 처리 기준에 기반하여 상기 특징 데이터의 적어도 일부를

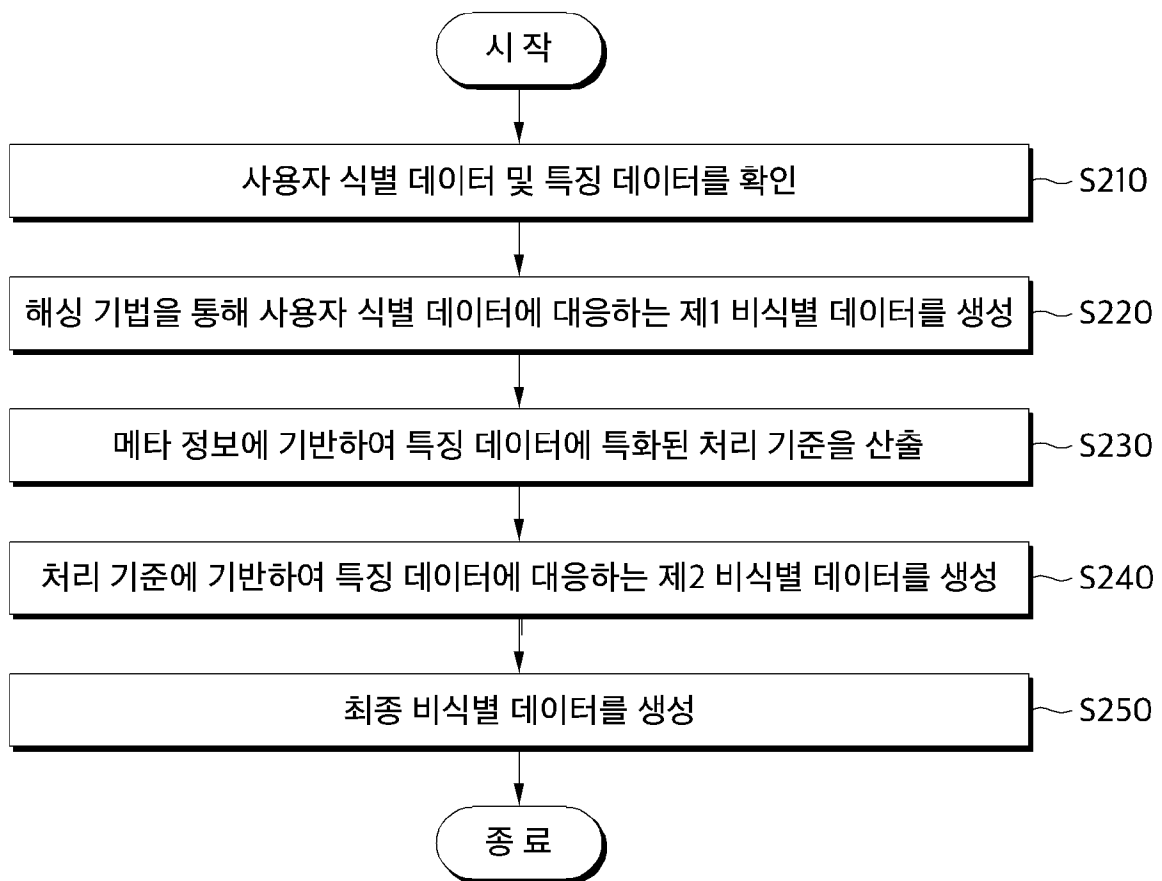
처리함으로써 상기 특징 데이터에 대응하는 제2 비식별 데이터를 생성하고,
상기 제1 비식별 데이터를 키(key)로, 상기 제2 비식별 데이터 중 적어도 일부를 값(value)으로 갖는 하나 이상의 테이블을 포함하는 최종 비식별 데이터를 생성하는, 전자 장치.

[도1]

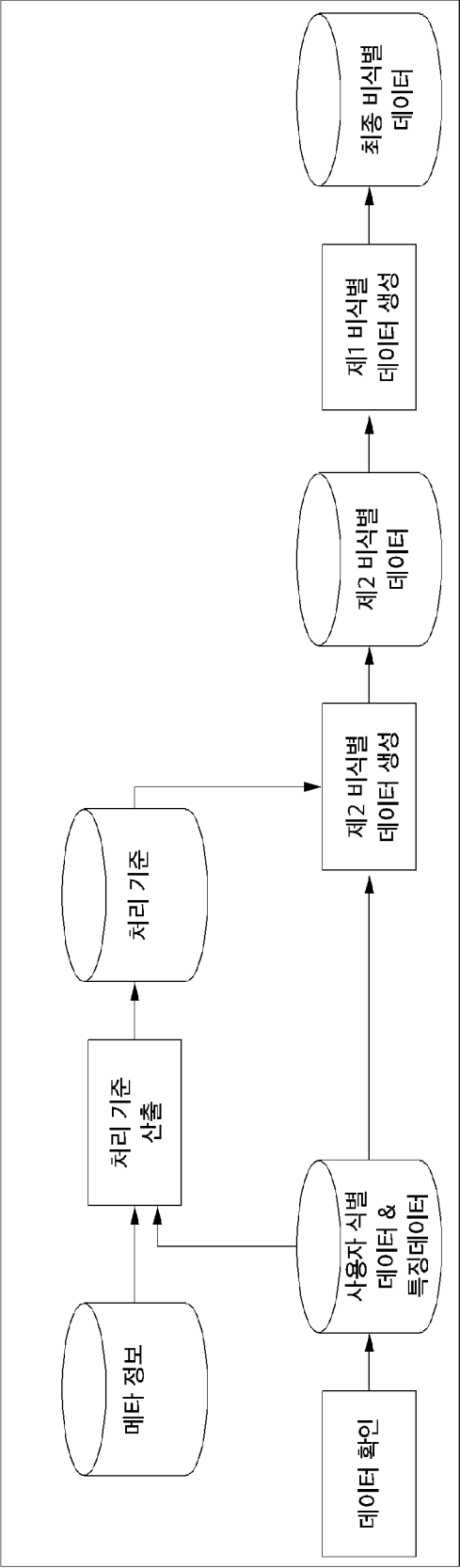
100



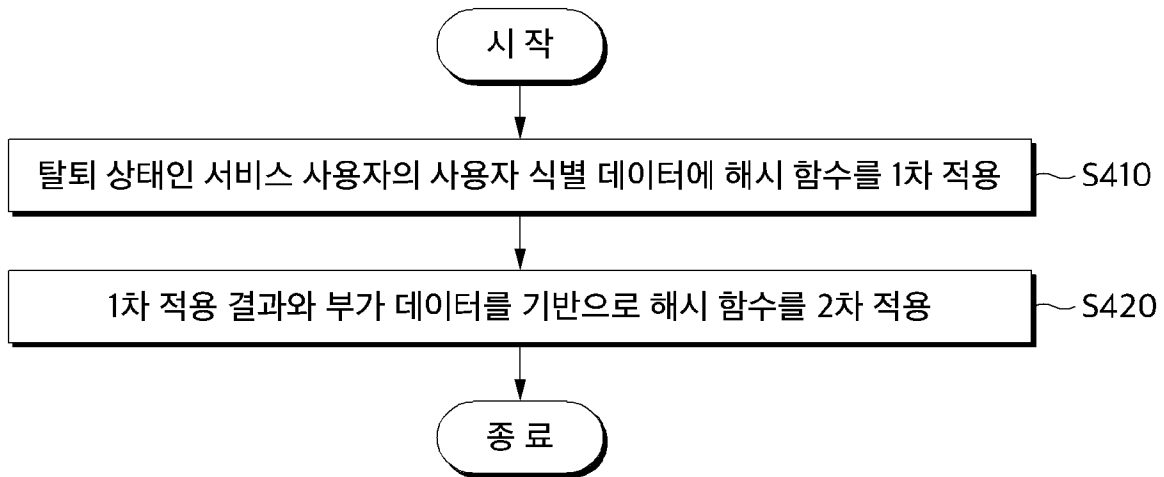
[도2]



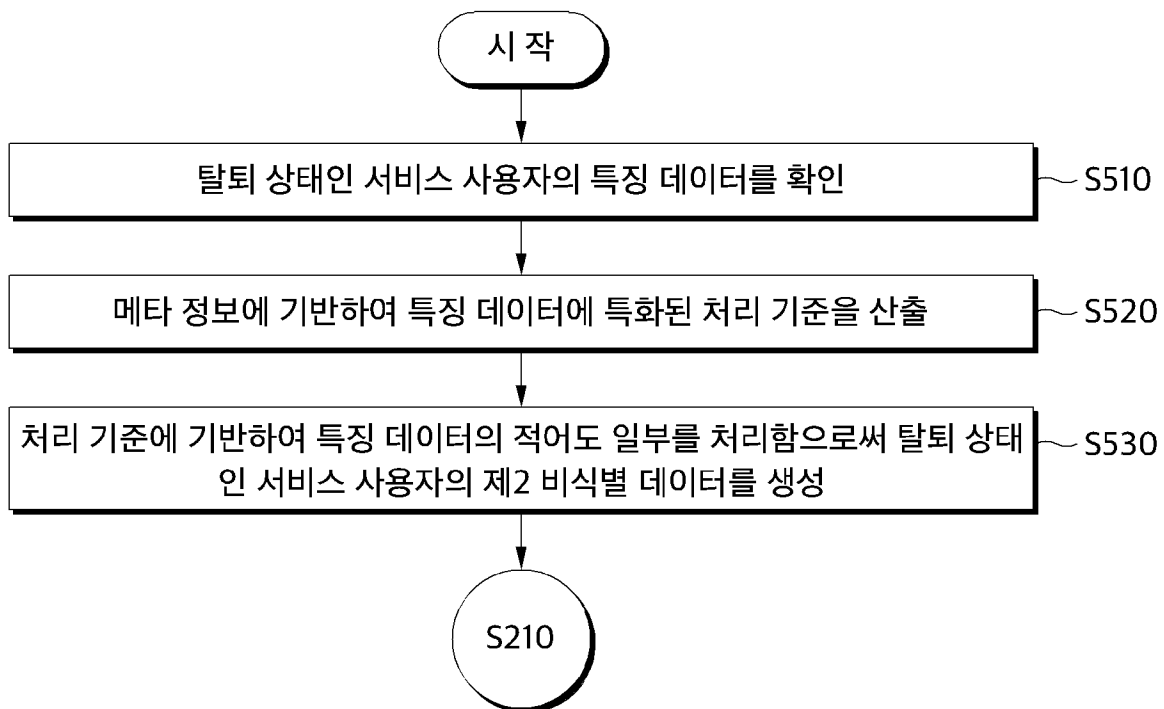
[도3]



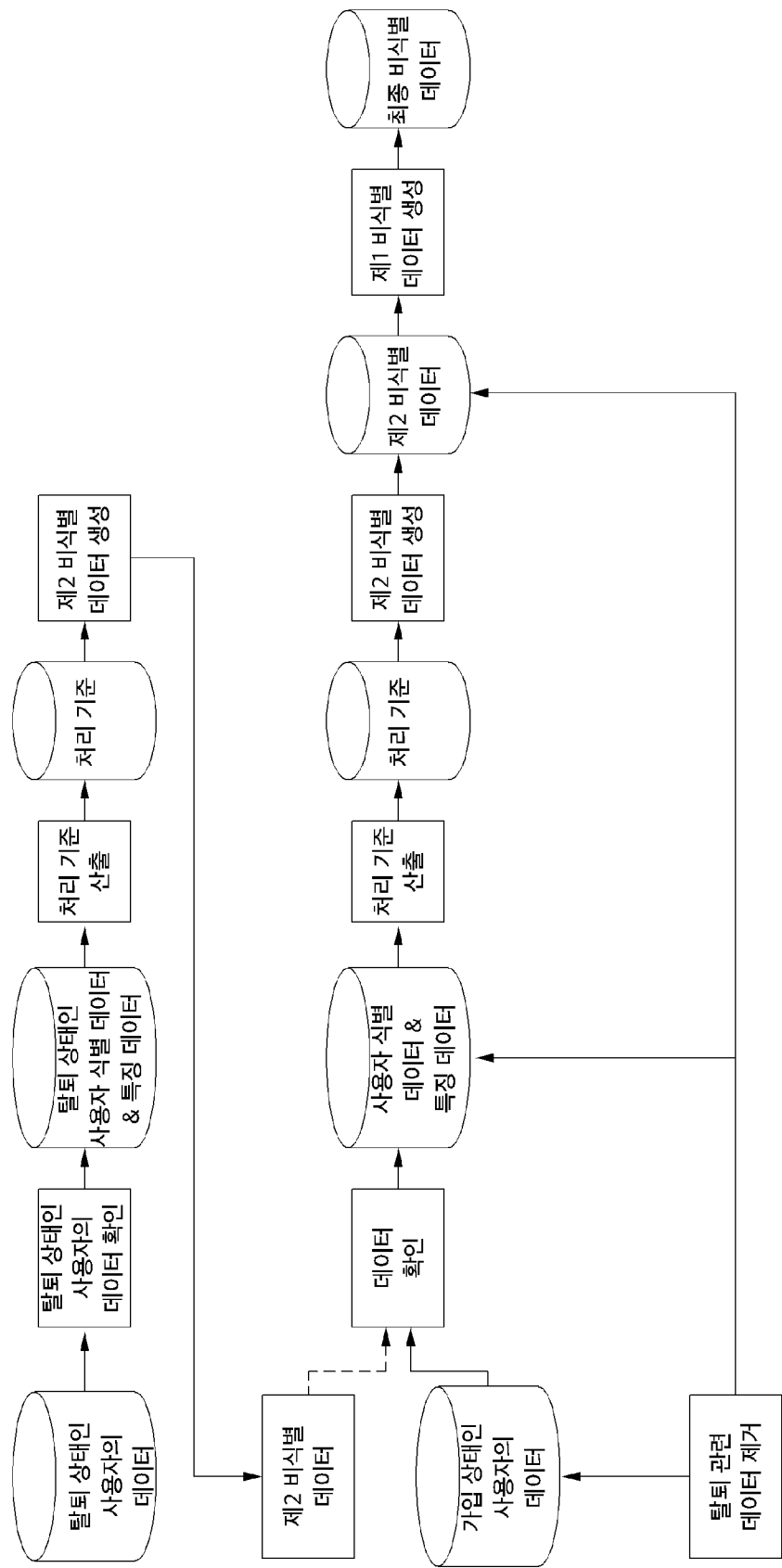
[도4]



[도5]



[도6]



[도7a]

연/월 △ 1	사용자 식별 데이터	성별	나이	회원 개월 수
202201	103400808	1	36	77
202202	103400808	1	36	78
202203	103400808	1	36	79

[도7b]

연/월 △	제1 비식별 데이터 △	성별 △	나이 △	회원 개월 수 △
202202	c14640e4517028624017f2382b4c6412802a787fa63cce4ad9d53fa370f0ec76	1	45	132
202201	d486895125ce1ec80c47e7cd2c9c5207dd6b3272215416c1c37741fda16d8805	0	38	131
202202	d486895125ce1ec80c47e7cd2c9c5207dd6b3272215416c1c37741fda16d8805	0	38	132

[도7c]

테이블 이름	△	가입 상태인 사용자 비식별화	△	탈퇴 상태인 사용자 비식별화	△
sb_rocketpay.csm_AA_pi_sample		• true		false	
sb_rocketpay.csm_AA_usage_sample		• true		false	
sb_rocketpay.csm_AA_payment_sample		• true		false	

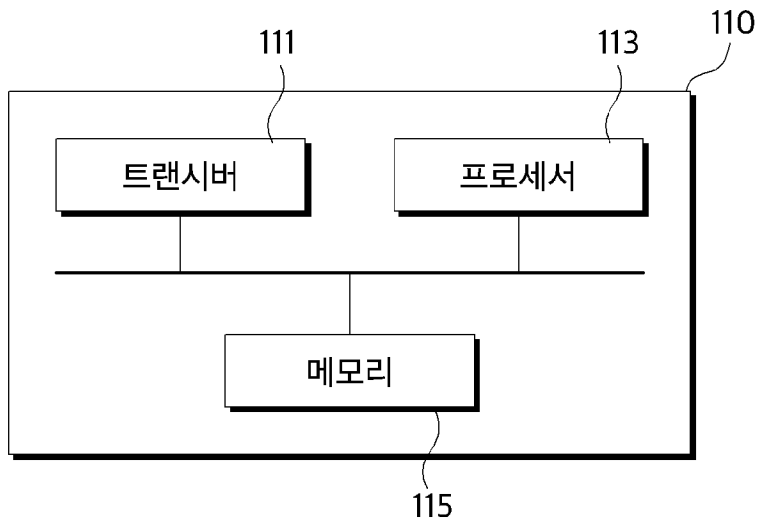
[도7d]

데이터블 이름	데이터 필드 정보	데이터 비식별화 방식 정보
sb_rocketpay.csm_AA_category2_r3m_sample	order_cate2_fashion_r3m	quantile200
sb_rocketpay.csm_AA_category2_r12m_sample	order_cate1_fashion_r12m	quantile200
sb_rocketpay.csm_payLater_usage_sample	gmv_payltr_r12m	quantile200

[도7e]

날짜

[도8]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2022/011120

A. CLASSIFICATION OF SUBJECT MATTER

G06F 21/62(2013.01)i; G06F 16/907(2019.01)i; G06Q 30/02(2012.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 21/62(2013.01); G06K 9/00(2006.01); H04L 12/24(2006.01); H04L 29/06(2006.01); H04L 9/32(2006.01);
H04W 12/02(2009.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 비식별화(de-identification), 패턴(pattern), 필드(field), 해시(hash), 키(key), 값(value), 정책(policy)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	EP 3582133 A1 (AVIDENT-IT AB) 18 December 2019 (2019-12-18) See paragraph [0023]; and claim 1.	1-8,10,16-17
A		9,11-15
Y	US 2017-0149793 A1 (SYMANTEC CORPORATION) 25 May 2017 (2017-05-25) See paragraphs [0031], [0043] and [0052].	1-8,10,16-17
Y	US 2020-0193060 A1 (PRIVACY ANALYTICS INC.) 18 June 2020 (2020-06-18) See claims 1, 10 and 12-13.	6-8
A	JP 7100563 B2 (HITACHI, LTD.) 13 July 2022 (2022-07-13) See paragraphs [0013]-[0043]; and figures 1-9.	1-17
A	US 2020-0293772 A1 (FACEBOOK, INC.) 17 September 2020 (2020-09-17) See paragraphs [0095]-[0119]; and figures 7-10.	1-17



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“D” document cited by the applicant in the international application

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

11 April 2023

Date of mailing of the international search report

12 April 2023

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon Building 4, 189 Cheongsaro, Seo-gu, Daejeon 35208

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2022/011120

Patent document cited in search report				Publication date (day/month/year)		Patent family member(s)		Publication date (day/month/year)	
EP	3582133	A1	18 December 2019	EP	3582133	B1		05 January 2022	
				SE	1850734	A1		16 December 2019	
US	2017-0149793	A1	25 May 2017	CN	108351946	A		31 July 2018	
				CN	108351946	B		08 March 2022	
				EP	3378007	A1		26 September 2018	
				EP	3378007	B1		19 January 2022	
				JP	2019-500679	A		10 January 2019	
				US	10326772	B2		18 June 2019	
				WO	2017-087074	A1		26 May 2017	
US	2020-0193060	A1	18 June 2020	US	10318763	B2		11 June 2019	
				US	10586074	B2		10 March 2020	
				US	11334685	B2		17 May 2022	
				US	2018-0173893	A1		21 June 2018	
				US	2019-0258826	A1		22 August 2019	
				US	2022-0253559	A1		11 August 2022	
JP	7100563	B2	13 July 2022	JP	2020-077256	A		21 May 2020	
				WO	2020-095662	A1		14 May 2020	
US	2020-0293772	A1	17 September 2020	CA	2963113	A1		30 September 2017	
				US	10210386	B2		19 February 2019	
				US	10706277	B2		07 July 2020	
				US	11210510	B2		28 December 2021	
				US	2017-0286765	A1		05 October 2017	
				US	2019-0180100	A1		13 June 2019	

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G06F 21/62(2013.01)i; G06F 16/907(2019.01)i; G06Q 30/02(2012.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G06F 21/62(2013.01); G06K 9/00(2006.01); H04L 12/24(2006.01); H04L 29/06(2006.01); H04L 9/32(2006.01); H04W 12/02(2009.01) 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 비식별화(de-identification), 패턴(pattern), 필드(field), 해시(hash), 키(key), 값(value), 정책(policy)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	EP 3582133 A1 (AVIDENT-IT AB) 2019.12.18 단락 [0023]; 및 청구항 1	1-8,10,16-17
A		9,11-15
Y	US 2017-0149793 A1 (SYMANTEC CORPORATION) 2017.05.25 단락 [0031], [0043], [0052]	1-8,10,16-17
Y	US 2020-0193060 A1 (PRIVACY ANALYTICS INC.) 2020.06.18 청구항 1, 10, 12-13	6-8
A	JP 7100563 B2 (HITACHI, LTD.) 2022.07.13 단락 [0013]-[0043]; 및 도면 1-9	1-17
A	US 2020-0293772 A1 (FACEBOOK, INC.) 2020.09.17 단락 [0095]-[0119]; 및 도면 7-10	1-17
☐ 추가 문헌이 C(계속)에 기재되어 있습니다. ☒ 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2023년04월11일(11.04.2023)		국제조사보고서 발송일 2023년04월12일(12.04.2023)
ISA/KR의 명칭 및 우편주소 대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578		심사관 양정록 전화번호 +82-42-481-5709

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
EP 3582133 A1	2019/12/18	EP 3582133 B1	2022/01/05
		SE 1850734 A1	2019/12/16
US 2017-0149793 A1	2017/05/25	CN 108351946 A	2018/07/31
		CN 108351946 B	2022/03/08
		EP 3378007 A1	2018/09/26
		EP 3378007 B1	2022/01/19
		JP 2019-500679 A	2019/01/10
		US 10326772 B2	2019/06/18
		WO 2017-087074 A1	2017/05/26
US 2020-0193060 A1	2020/06/18	US 10318763 B2	2019/06/11
		US 10586074 B2	2020/03/10
		US 11334685 B2	2022/05/17
		US 2018-0173893 A1	2018/06/21
		US 2019-0258826 A1	2019/08/22
		US 2022-0253559 A1	2022/08/11
JP 7100563 B2	2022/07/13	JP 2020-077256 A	2020/05/21
		WO 2020-095662 A1	2020/05/14
US 2020-0293772 A1	2020/09/17	CA 2963113 A1	2017/09/30
		US 10210386 B2	2019/02/19
		US 10706277 B2	2020/07/07
		US 11210510 B2	2021/12/28
		US 2017-0286765 A1	2017/10/05
		US 2019-0180100 A1	2019/06/13