



- (51) 국제특허분류:
G06F 11/08 (2006.01)
- (21) 국제출원번호: PCT/KR2014/011091
- (22) 국제출원일: 2014년 11월 18일 (18.11.2014)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2014-0065157 2014년 5월 29일 (29.05.2014) KR
- (71) 출원인: 삼성에스디에스 주식회사 (SAMSUNG SDS CO., LTD.) [KR/KR]; 138-240 서울시 송파구 올림픽로 35길 125 삼성 SDS타워 West Campus, Seoul (KR).
- (72) 발명자: 남궁영환 (NAMKOONG, Young-Hwan); 135-880 서울시 강남구 삼성로 512, 6층 CSP연구소(삼성동, 삼성동빌딩), Seoul (KR). 김다운 (KIM, Da-Woon); 135-880 서울시 강남구 삼성로 512, 6층 CSP연구소(삼성동, 삼성동빌딩), Seoul (KR). 홍정훈 (HONG, Jeong-Hoon); 135-880 서울시 강남구 삼성로 512, 6층 CSP연구소(삼성동, 삼성동빌딩), Seoul (KR).
- (74) 대리인: 두호특허법인 (DOOHO IP LAW FIRM); 06628 서울시 서초구 강남대로 51길 1, 7층(서초동, 대현블루타워), 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, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

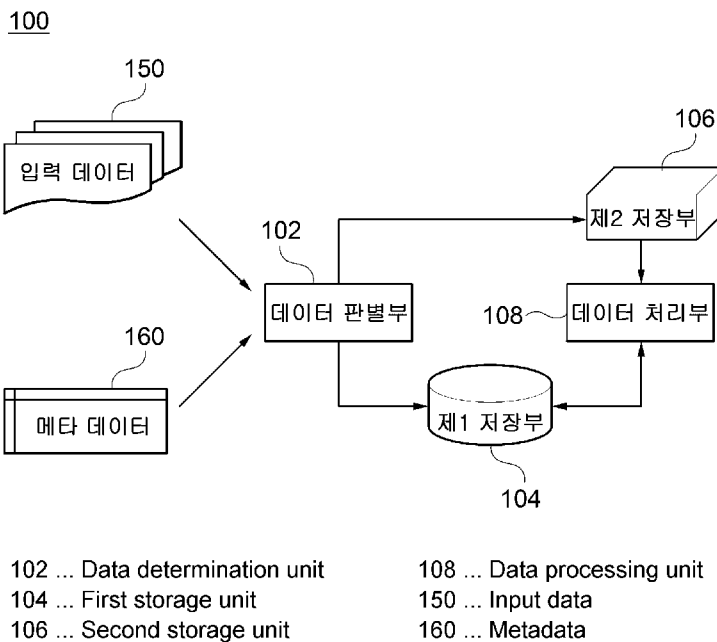
(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, 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 제 21 조(3))

(54) Title: SYSTEM AND METHOD FOR PROCESSING DATA

(54) 발명의 명칭: 데이터 처리 시스템 및 방법



(57) Abstract: Disclosed are a system and a method for processing data. A system for processing data according to an embodiment of the present invention comprises: a data determination unit for receiving input data and metadata about the structure of the input data and determining whether the input data is normal data or abnormal data on the basis of the metadata; a first storage unit for storing the normal data; a second storage unit for storing the abnormal data; and a data processing unit for modifying the abnormal data stored in the second storage unit so as to have the same structure as the normal data, and storing the modified abnormal data in the first storage unit.

(57) 요약서: 데이터 처리 시스템 및 방법이 개시된다. 본 발명의 일 실시예에 따른 데이터 처리 시스템은, 입력 데이터, 및 상기 입력 데이터의 구조에 관한 메타 데이터를 입력받고, 상기 메타 데이터를 기반으로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 데이터 판별부; 상기 정상 데이터가 저장되는 제 1 저장부; 상기 비정상 데이터가 저장되는 제 2 저장부; 및 상기 정상 데이터와 동일한 구조를 갖도록 상기 제 2 저장부에 저장된 상기 비정상 데이터를 수정하고, 수정

된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는 데이터 처리부를 포함한다.

명세서

발명의 명칭: 데이터 처리 시스템 및 방법

기술분야

- [1] 본 발명의 실시예들은 데이터 처리 시스템 및 방법에 관한 것으로서, 보다 상세하게는 대용량 데이터 적재 시 데이터 오류를 자동으로 처리하기 위한 데이터 처리 시스템 및 방법에 관한 것이다.

배경기술

- [2] 대용량의 로그 데이터 분석 초기 단계에서, 데이터의 구조화, 적재 및 저장은 매우 중요한 작업이다. 데이터는 분석을 위한 최소한의 데이터 구조 정보에 맞도록 적재 또는 저장이 이루어져야 한다. 그러나, 대부분의 기존 로그 데이터는 이러한 데이터 구조 정보에 맞지 않는 비정상 데이터를 다수 포함하고 있다. 특히, 대용량 데이터의 경우, 이러한 비정상 데이터로 인해 데이터 적재 과정에서 심각한 오류가 발생되거나, 또는 잘못된 형태로 적재되어 부정확한 데이터 분석 결과가 도출될 수 있다. 종래 기술은 로그 데이터의 적재 시 오류 부분을 자동으로 수정하거나 구조화하는 처리 기능이 없었으며, 오류 수정 또는 구조화 과정이 모두 수작업으로 이루어졌다.
- [3] 구체적으로, 종래 기술의 경우, 데이터 입력 도중 오류가 발생되면 데이터의 적재 작업이 중지되고 오류 원인을 찾아 수정한 후 처음부터 데이터 적재를 다시 수행하였다. 만약, 비정상 데이터가 여러 번 발견될 경우 이러한 작업은 매번 반복적으로 수행될 수 밖에 없었으며, 특히, 대용량 데이터 적재 시 이러한 번거로움으로 인한 시간 및 비용이 낭비되는 문제점이 있었다.

발명의 상세한 설명

기술적 과제

- [4] 본 발명의 실시예들은 대용량 데이터를 효율적으로 저장하고 데이터에 포함된 오류를 자동으로 처리하기 위한 데이터 처리 시스템 및 방법을 제공하기 위한 것이다.

과제 해결 수단

- [5] 본 발명의 예시적인 실시예에 따르면, 입력 데이터, 및 상기 입력 데이터의 구조에 관한 메타 데이터를 입력받고, 상기 메타 데이터를 기반으로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 데이터 판별부; 상기 정상 데이터가 저장되는 제 1 저장부; 상기 비정상 데이터가 저장되는 제 2 저장부; 및 상기 정상 데이터와 동일한 구조를 갖도록 상기 제 2 저장부에 저장된 상기 비정상 데이터를 수정하고, 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는 데이터 처리부를 포함하는, 데이터 처리 시스템이 제공된다.
- [6] 상기 데이터 판별부는, 상기 입력 데이터의 레코드(record) 단위로 상기 입력

- 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다.
- [7] 상기 메타 데이터는, 레코드별 데이터 필드(data field)의 개수에 관한 정보, 상기 데이터 필드 각각을 구분하는 구분자의 개수에 관한 정보, 및 상기 데이터 필드별 데이터 패턴에 관한 정보 중 하나 이상을 포함할 수 있다.
- [8] 상기 데이터 판별부는, 상기 메타 데이터에 포함된 데이터 필드의 개수 또는 상기 구분자의 개수 중 하나와 상기 입력 데이터의 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수 중 하나를 비교하거나, 또는 상기 입력 데이터의 레코드에 포함된 데이터 필드별로 상기 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 상기 입력 데이터의 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다.
- [9] 상기 데이터 처리부는, 상기 정상 데이터 내의 특정 데이터 패턴을 추출하고, 상기 비정상 데이터 내의 데이터 필드 중 상기 특정 데이터 패턴과 매칭되는 데이터 필드를 기준으로 상기 비정상 데이터를 복수 개의 가상 레코드로 구분하며, 상기 가상 레코드 단위로 상기 비정상 데이터를 수정할 수 있다.
- [10] 상기 데이터 처리부는, 상기 정상 데이터 내의 데이터 필드별로 하나 이상의 특정 데이터 값을 추출하고, 상기 가상 레코드 내의 데이터 필드 중 상기 특정 데이터 값이 존재하는 데이터 필드를 제외한 나머지 데이터 필드에 포함된 데이터를 수정할 수 있다.
- [11] 상기 데이터 처리부는, 상기 나머지 데이터 필드에 포함된 데이터 내 특정 문자열 또는 문자를 삭제하거나, 또는 기설정된 문자열 또는 문자로 치환할 수 있다.
- [12] 상기 데이터 처리부는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상 데이터인지 여부를 판별하고, 상기 가상 레코드가 정상 데이터로 판별된 경우 상기 가상 레코드를 상기 데이터 처리부 내 데이터베이스에 저장하며, 가상 레코드 모두가 상기 데이터베이스에 저장된 경우 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장할 수 있다.
- [13] 상기 데이터 처리부는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상 데이터인지 여부를 판별하고, 상기 가상 레코드가 정상 데이터가 아닌 것으로 판별된 경우 현재까지 수정한 가상 레코드 모두를 재차 수정할 수 있다.
- [14] 본 발명의 다른 예시적인 실시예에 따르면, 데이터 판별부가, 입력 데이터, 및 상기 입력 데이터의 구조에 관한 메타 데이터를 입력받는 단계; 상기 데이터 판별부가, 상기 메타 데이터를 기반으로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 단계; 상기 데이터 판별부가, 상기 정상 데이터를 제 1 저장부에 저장하는 단계; 상기 데이터 판별부가, 상기 비정상 데이터를 제 2 저장부에 저장하는 단계; 데이터 처리부가, 상기 정상 데이터와 동일한 구조를 갖도록 상기 제 2 저장부에 저장된 상기 비정상 데이터를 수정하는 단계; 및 상기 데이터 처리부가, 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는 단계를 포함하는, 데이터 처리 방법이 제공된다.

- [15] 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 단계는, 상기 입력 데이터의 레코드(record) 단위로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다.
- [16] 상기 메타 데이터는, 레코드별 데이터 필드(data field)의 개수에 관한 정보, 상기 데이터 필드 각각을 구분하는 구분자의 개수에 관한 정보, 및 상기 데이터 필드별 데이터 패턴에 관한 정보 중 하나 이상을 포함할 수 있다.
- [17] 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 단계는, 상기 메타 데이터에 포함된 데이터 필드의 개수 또는 상기 구분자의 개수 중 하나와 상기 입력 데이터의 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수 중 하나를 비교하거나, 또는 상기 입력 데이터의 레코드에 포함된 데이터 필드별로 상기 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 상기 입력 데이터의 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다.
- [18] 상기 비정상 데이터를 수정하는 단계는, 상기 정상 데이터 내의 특정 데이터 패턴을 추출하는 단계; 상기 비정상 데이터 내의 데이터 필드 중 상기 특정 데이터 패턴과 매칭되는 데이터 필드를 기준으로 상기 비정상 데이터를 복수 개의 가상 레코드로 구분하는 단계; 및 상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계를 포함할 수 있다.
- [19] 상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계는, 상기 정상 데이터 내의 데이터 필드별로 하나 이상의 특정 데이터 값을 추출하는 단계; 및 상기 가상 레코드 내의 데이터 필드 중 상기 특정 데이터 값이 존재하는 데이터 필드를 제외한 나머지 데이터 필드에 포함된 데이터를 수정하는 단계를 포함할 수 있다.
- [20] 상기 나머지 데이터 필드에 포함된 데이터를 수정하는 단계는, 상기 나머지 데이터 필드에 포함된 데이터 내 특정 문자열 또는 문자를 삭제하거나, 또는 기설정된 문자열 또는 문자로 치환할 수 있다.
- [21] 상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상 데이터인지 여부를 판별하는 단계; 상기 가상 레코드가 정상 데이터로 판별된 경우 상기 가상 레코드를 상기 데이터 처리부 내 데이터베이스에 저장하는 단계; 및 가상 레코드 모두가 상기 데이터베이스에 저장된 경우 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는 단계를 포함할 수 있다.
- [22] 상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상 데이터인지 여부를 판별하는 단계; 및 상기 가상 레코드가 정상 데이터가 아닌 것으로 판별된 경우 현재까지 수정한 가상 레코드 모두를 재차 수정하는 단계를 포함할 수 있다.

발명의 효과

- [23] 본 발명의 실시예들에 따르면, 데이터 입력 과정에서 비정상 데이터를 제 2 저장부에 임시 저장하면서 정상 데이터에 대한 입력을 계속 진행하도록 함으로써, 비정상 데이터의 입력 오류에 따른 데이터 처리 시간이 지연되는 것 방지할 수 있다. 이에 따라, 비정상 데이터로 인해 데이터 저장 및 적재 작업이 중단되는 것을 방지할 수 있으며, 기존에 저장된 정상 데이터를 무효화시킬 필요가 없게 된다. 또한, 불필요한 저장, 적재 반복 작업을 개선할 수 있다.
- [24] 또한, 본 발명의 실시예들에 따르면, 정상 데이터의 데이터 패턴 및 데이터 값을 이용하여 비정상 데이터 내의 오류 지점을 효율적으로 탐색할 수 있다.

도면의 간단한 설명

- [25] 도 1은 본 발명의 일 실시예에 따른 데이터 처리 시스템의 상세 구성을 나타낸 블록도
- [26] 도 2는 본 발명의 일 실시예에 따른 입력 데이터의 예시도
- [27] 도 3은 본 발명의 일 실시예에 따른 정상 데이터와 비정상 데이터의 예시도
- [28] 도 4는 본 발명의 일 실시예에 따른 데이터 처리 과정을 개략적으로 나타낸 도면
- [29] 도 5는 본 발명의 일 실시예에 따른 비정상 데이터가 정상 데이터와 동일한 구조를 갖도록 수정된 결과를 나타낸 도면
- [30] 도 6 본 발명의 일 실시예에 따른 입력 데이터가 저장 및 적재되는 최종 결과를 나타낸 도면
- [31] 도 7은 본 발명의 일 실시예에 따른 데이터 처리 방법을 설명하기 위한 순서도
- [32] 도 8은 도 7의 S706의 예시적인 실시예를 설명하기 위한 순서도
- [33] 도 9는 도 7의 S706의 다른 예시적인 실시예를 설명하기 위한 순서도
- [34] 도 10은 도 7의 S714의 예시적인 실시예를 설명하기 위한 순서도

발명의 실시를 위한 형태

- [35] 이하, 도면을 참조하여 본 발명의 구체적인 실시예를 설명하기로 한다. 그러나 이는 예시적 실시예에 불과하며 본 발명은 이에 한정되지 않는다.
- [36] 본 발명을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [37] 본 발명의 기술적 사상은 청구범위에 의해 결정되며, 이하 실시예는 진보적인 본 발명의 기술적 사상을 본 발명이 속하는 기술분야에서 통상의 지식을 가진자에게 효율적으로 설명하기 위한 일 수단일 뿐이다.
- [38] 도 1은 본 발명의 일 실시예에 따른 데이터 처리 시스템(100)의 상세 구성을 나타낸 블록도이며, 도 2는 본 발명의 일 실시예에 따른 입력 데이터(150)의

예시이고, 도 3은 본 발명의 일 실시예에 따른 정상 데이터와 비정상 데이터의 예시이다.

- [39] 도 1에 도시된 바와 같이, 본 발명의 일 실시예에 따른 데이터 처리 시스템(100)은 데이터 판별부(102), 제 1 저장부(104), 제 2 저장부(106) 및 데이터 처리부(108)를 포함한다.
- [40] 데이터 판별부(102)는 입력 데이터(150), 및 입력 데이터(150)의 구조에 관한 메타 데이터(160)를 입력받고, 메타 데이터(160)를 기반으로 입력 데이터(150)가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별한다. 여기서, 입력 데이터(150)는 분석 대상이 되는 데이터로서 예를 들어, 특정 시스템의 로그 데이터, 장애 발생 상황에 관한 데이터 등이 될 수 있다. 데이터 판별부(102)는 예를 들어, 데이터 처리 시스템(100) 내부의 관리 서버(미도시), 데이터 처리 시스템(100) 외부의 대상 시스템(미도시) 등으로부터 입력 데이터(150)를 입력 받을 수 있다.
- [41] 도 2 및 도 3을 참조하면, 입력 데이터(150)는 레코드(record) 단위로 이루어질 수 있으며, 입력 데이터(150)의 레코드 각각은 복수 개의 데이터 필드(data field)를 가질 수 있다. 여기서, 데이터 필드 각각은 예를 들어, 작성 시간(time), 장애 내용(content), 진행 상태(process), 처리 결과(result) 등에 해당하는 데이터를 포함할 수 있다. 각 데이터 필드에는 예를 들어, 작성 시간: "20131129153000", 장애 내용: "H/W Fault", 진행 상태: "종료", 처리 결과: "모니터링 강화 / 디스크 성능 분석을 통한 빈번한 media error 원인 분석 진행" 등과 같은 데이터가 포함될 수 있다. 데이터 필드 각각의 데이터들은 구분자(예를 들어, 콤마 등)를 통해 구분될 수 있다. 데이터 판별부(102)는 입력 데이터(150)를 레코드 단위로 순차적으로 입력받을 수 있다. 또한, 데이터 판별부(102)는 입력 데이터(150)의 레코드 단위로 입력 데이터(150)가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다.
- [42] 여기서, 정상 데이터는 기설정된 데이터의 구조에 맞는 데이터를 의미하며, 비정상 데이터는 기설정된 데이터의 구조에 맞지 않는 데이터를 의미한다. 도 3의 (a)는 정상 데이터의 예시를 나타낸 도면이며, 도 3의 (b)는 비정상 데이터의 예시를 나타낸 도면이다. 도 3에 도시된 바와 같이, 정상 데이터는 기설정된 데이터의 구조와 부합하므로 정상적인 형태로 저장 및 적재될 수 있다. 그러나, 비정상 데이터(도 2에 도시된 ①, ② 및 도 3의 b에 도시된 ①, ②)는 데이터 내에 포함된 특수 문자(예를 들어, ",") , 개행 문자(예를 들어, "\n" 등에 의해 기설정된 데이터 구조와 맞지 않는 데이터 구조를 가지게 되므로 비정상적인 형태로 저장 및 적재될 수 있다. 구체적으로, 도 2 및 도 3의 ①의 경우, 데이터 "Push 요청으로 인한 CPU 사용량 과다로 CPU RUNQUEUE 초과"와 데이터 "EBD 등록 요청" 사이에 구분자 ","가 사용되어 의도치 않은 부분에서 열(데이터 필드)이 나누어지는 현상이 발생하였다. 도 2 및 도 3의 ②의 경우, 데이터 "○ 발생원인 : 서비스 사용량 증가에 따른 테이블스페이스 임계치 초과사용현상" 과 데이터 "○

조치내용 : 테이블스페이스에 30GB raw device add" 사이에 개행 문자 "/n"이 사용되어 의도치 않은 부분에서 행(레코드)이 나누어지는 현상이 발생하였다. 도 3의 (b)에서 볼 수 있는 바와 같이, 이러한 비정상 데이터는 정상 데이터와 다른 구조를 가지게 되며 비정상적인 형태로 저장 및 적재될 수 있다.

- [43] 이와 같이 입력 데이터(150)가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하기 위해, 데이터 판별부(102)는 입력 데이터(150)뿐 아니라, 입력 데이터(150)의 구조에 관한 메타 데이터(160)를 함께 입력 받을 수 있다. 일반적으로, 메타 데이터는 데이터에 관한 구조화된 데이터 즉, 다른 데이터를 설명해 주는 데이터로서 대량의 정보 가운데에서 찾고 있는 정보를 효율적으로 찾아내서 이용하기 위해 일정한 규칙에 따라 콘텐츠(여기서는, 입력 데이터(150))에 부여되는 데이터를 의미한다. 본 발명의 실시예들에 따른 메타 데이터(160)는 입력 데이터(150)의 구조에 관한 정보를 포함한다. 구체적으로, 메타 데이터(160)는 입력 데이터(150)의 레코드별 데이터 필드의 개수에 관한 정보, 입력 데이터(150)의 레코드에 포함된 데이터 필드 각각을 구분하는 구분자의 개수에 관한 정보, 및 입력 데이터(150)의 데이터 필드별 데이터 패턴에 관한 정보 중 하나 이상을 포함할 수 있다. 즉, 메타 데이터(160)를 통해 입력 데이터(150)의 레코드별로 몇 개의 데이터 필드가 존재하여야 하는지, 입력 데이터(150)의 레코드별로 몇 개의 구분자가 존재하여야 하는지, 입력 데이터(150)의 데이터 필드별로 데이터가 어떠한 패턴을 가져야 하는지 등을 알 수 있다.

- [44] 데이터 판별부(102)는 메타 데이터(160)에 포함된 데이터 필드의 개수 또는 구분자의 개수 중 하나와 입력 데이터(150)의 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수 중 하나를 비교함으로써, 입력 데이터(150)의 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다. 예를 들어, 메타 데이터(160)에 레코드별 데이터 필드의 개수가 4개로 설정되어 있다고 가정하였을 때, 데이터 판별부(102)는 입력 데이터(150)의 레코드에 포함된 데이터 필드의 개수를 계산한 후 메타 데이터(160)에 포함된 데이터 필드의 개수(4개)와 비교함으로써, 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다. 만약, 데이터 판별부(102)가 계산한 입력 데이터(150)의 레코드에 포함된 데이터 필드의 개수가 메타 데이터(160)에 포함된 데이터 필드의 개수와 일치하는 경우, 데이터 판별부(102)는 입력 데이터(150)의 레코드를 정상 데이터로 판별할 수 있다. 그러나, 데이터 판별부(102)가 계산한 입력 데이터(150)의 레코드에 포함된 데이터 필드의 개수가 메타 데이터(160)에 포함된 데이터 필드의 개수와 일치하지 않는 경우, 데이터 판별부(102)는 입력 데이터(150)의 레코드를 비정상 데이터로 판별할 수 있다. 또한, 데이터 판별부(102)는 입력 데이터(150)의 레코드에 포함된 구분자의 개수를 계산한 후 메타 데이터(160)에 포함된 구분자의 개수와 비교함으로써, 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수도

있다.

- [45] 나아가, 데이터 판별부(102)는 입력 데이터(150)의 레코드에 포함된 데이터 필드별로 메타 데이터(160)에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 입력 데이터(150)의 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수도 있다. 상술한 바와 같이, 메타 데이터(160)에는 데이터 필드별 데이터 패턴에 관한 정보가 포함되어 있을 수 있다. 예를 들어, 정상 데이터의 레코드 시작 부분("작성 시간"에 관한 데이터 필드)은 항상 "YYYYMMDDhhmmss"와 같이 14개의 숫자를 포함하는 숫자열 형태의 패턴을 가질 수 있다(여기서, Y는 연도(year), M은 월(month), D는 일(day), h는 시(hour), m은 분(minute), s는 초(second)를 의미함). 또한, "장애 내용", "진행 상태" 및 "처리 결과"에 관한 데이터 필드는 문자열 형태의 패턴을 가질 수 있으며, 특히 "진행 상태"에 관한 데이터 필드는 2개("종료") 또는 3개("진행중")의 문자를 포함하는 문자열 형태의 패턴을 가질 수 있다. 이와 같이, 메타 데이터(160)에는 데이터 필드별 데이터 패턴에 관한 정보가 포함되어 있다. 데이터 판별부(102)는 입력 데이터(150)의 레코드에 포함된 데이터 필드별로 메타 데이터(160)에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단한 후, 데이터 필드별로 모든 패턴이 매칭되는 경우 해당 레코드를 정상 데이터로 판별할 수 있다. 그러나, 입력 데이터(150)의 레코드에 포함된 데이터 필드 중 하나라도 메타 데이터(160)에 포함된 데이터 필드별 데이터 패턴과 매칭되지 않는 경우, 데이터 판별부(102)는 해당 레코드를 비정상 데이터로 판별할 수 있다. 이와 같은 과정을 통해, 데이터 판별부(102)는 입력 데이터(150)의 레코드 단위로 해당 입력 데이터(150)가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다. 다만, 여기에서 설명한 입력 데이터(150)에 대한 정상 데이터 및 비정상 데이터의 판별 방법은 일 실시예에 불과하며, 데이터 판별부(102)가 반드시 위와 같은 방법으로 입력 데이터(150)를 정상 데이터 또는 비정상 데이터로 판별하여야 하는 것은 아니다. 데이터 판별부(102)는 다양한 방법을 통해 입력 데이터(150)를 정상 데이터 또는 비정상 데이터로 판별할 수 있다.

- [46] 이후, 데이터 판별부(102)는 위와 같은 과정을 통해 판별한 정상 데이터를 제 1 저장부(104)에 저장하며, 비정상 데이터를 제 2 저장부(106)에 저장한다. 즉, 데이터 판별부(102)는 데이터 입력 단계에서 입력 데이터(150)를 정상 데이터와 비정상 데이터로 분류한 후, 정상 데이터와 비정상 데이터를 별도의 저장부에 각각 저장할 수 있다.

- [47] 제 1 저장부(104)는 정상 데이터가 저장되는 장소로서, 데이터 판별부(102)로부터 정상 데이터를 수신하여 이를 순차적으로 저장 및 적재할 수 있다. 이때, 제 1 저장부(104)는 정상 데이터를 레코드 단위별로 저장 및 적재할 수 있다. 또한, 후술할 바와 같이, 제 1 저장부(104)는 데이터 처리부(108)로부터 수정된 비정상 데이터를 수신하여 이를 순차적으로 저장 및 적재할 수도 있다.

- [48] 제 2 저장부(106)는 비정상 데이터가 저장되는 장소로서, 데이터 판별부(102)로부터 비정상 데이터를 수신하여 이를 순차적으로 저장 및 적재할 수 있다. 데이터 처리부(108)는 제 2 저장부(106)에 저장된 비정상 데이터를 수정하여 정상 데이터로 변환한 후, 이를 다시 제 1 저장부(104)에 저장할 수 있다. 즉, 제 2 저장부(106)는 비정상 데이터를 임시 보관하는 역할을 한다.
- [49] 데이터 처리부(108)는 정상 데이터와 동일한 구조를 갖도록 제 2 저장부(106)에 저장된 비정상 데이터를 수정하고, 수정된 비정상 데이터를 제 1 저장부(104)에 저장한다.
- [50] 먼저, 데이터 처리부(108)는 제 1 저장부(104)에 저장된 정상 데이터의 데이터 필드 내 특정 데이터 패턴을 추출한다. 상술한 바와 같이, 입력 데이터의 데이터 필드 각각에는 특정 패턴을 갖는 데이터가 포함되어 있다. 예를 들어, 입력 데이터의 레코드 시작 부분("작성 시간"에 관한 데이터 필드)은 항상 "YYYYMMDDhhmmss"와 같이 14개의 숫자를 포함하는 숫자열 형태의 패턴을 가질 수 있다. 데이터 처리부(108)는 정상 데이터 내에서 임의의 특정 데이터 패턴을 추출할 수 있다.
- [51] 데이터 처리부(108)는 추출된 데이터 패턴을 이용하여 비정상 데이터 내의 데이터 필드를 매칭시킬 수 있다. 예를 들어, 데이터 처리부(108)는 비정상 데이터를 읽어들이는 과정에서 패턴 "YYYYMMDDhhmmss"와 매칭되는 부분의 데이터가 발견될 때까지 비정상 데이터를 읽어들이 수 있다. 데이터 처리부(108)는 예를 들어, 숫자열 또는 문자열 매칭 알고리즘 등을 이용하여 추출된 패턴과 비정상 데이터 내 데이터 필드를 매칭시킬 수 있다. 데이터 처리부(108)는 비정상 데이터 내의 데이터 필드 중, 추출된 특정 데이터 패턴과 매칭되는 데이터 필드를 기준으로 비정상 데이터를 복수 개의 가상 레코드(record)로 구분할 수 있다. 예를 들어, 비정상 데이터 내에 "YYYYMMDDhhmmss" 패턴이 5번 발견되는 경우, "YYYYMMDDhhmmss" 패턴이 발견된 데이터 필드를 기준으로 해당 비정상 데이터를 5개의 가상 레코드로 구분할 수 있다. 정상 데이터와 달리 비정상 데이터 내에는 구분자(", " 등의 특수문자가 의도치 않게 포함되어 있을 수 있으므로, 비정상 데이터의 레코드 내 데이터 필드는 정상 데이터의 레코드 내 데이터 필드와 매칭되지 않을 수 있다. 이에 따라, 본 발명의 실시예들은 데이터 처리부(108)가 정상 데이터에서 추출된 데이터 패턴을 이용하여 비정상 데이터를 복수 개의 가상 레코드로 구분할 수 있도록 하였다. 즉, 데이터 처리부(108)는 비정상 데이터 내 추출된 특정 패턴이 발견된 지점(데이터 필드)을 기준으로 하여 비정상 데이터를 복수 개의 가상 레코드로 구분할 수 있으며, 가상 레코드 단위로 비정상 데이터를 순차적으로 수정할 수 있다.
- [52] 다음으로, 데이터 처리부(108)는 각 가상 레코드를 수정하기 위해 정상 데이터의 특정 데이터 값을 이용할 수 있다. 데이터 처리부(108)는 정상 데이터 내의 데이터 필드별로 하나 이상의 특정 데이터 값을 추출할 수 있다. 정상

데이터 내의 데이터 필드 각각에는 서로 다른 개수의 데이터 값이 존재할 수 있다. 예를 들어, 제 1 데이터 필드는 "국적"에 관한 데이터로서 "한국", "미국" 및 "일본"을 포함하는 3개의 데이터를 포함할 수 있다. 또한, 제 2 데이터 필드는 "성별"에 관한 데이터로서 "남자" 및 "여자"를 포함하는 2개의 데이터를 포함할 수 있다. 또한, 제 3 데이터 필드는 "처리 결과"에 관한 데이터로서 "URL 모니터링 변경 조치", "해당 부품 여유분 확보"... 등을 포함하는 20개의 데이터를 포함할 수 있다.

- [53] 데이터 처리부(108)는 정상 데이터 내의 데이터 필드별로 데이터 값(예를 들어, 제 1 데이터 필드에서는 "한국", "미국" 및 "일본"/제 2 데이터 필드에서는 "남자" 및 "여자"/제 3 데이터 필드에서는 "URL 모니터링 변경 조치" 및 "해당 부품 여유분 확보" 등)를 추출하고, 추출된 데이터 값과 가상 레코드 내의 데이터 필드를 각각 매칭시킬 수 있다. 예를 들어, 정상 데이터의 제 1 데이터 필드 내의 "한국", "미국" 및 "일본" 중 어느 하나가 비정상 데이터 내의 데이터 필드에 포함되어 있는지 여부를 판별할 수 있으며, "한국", "미국" 및 "일본" 중 어느 하나를 포함하는 비정상 데이터 내의 데이터 필드는 정상 데이터와 매칭되는 것으로 볼 수 있다. 이후, 데이터 처리부(108)는 매칭된 데이터 필드는 정상인 것으로 간주하고 매칭되지 않은 데이터 필드에 대해서만 수정 작업을 진행할 수 있다. 즉, 데이터 처리부(108)는 가상 레코드 내의 데이터 필드 중 추출된 특정 데이터 값이 존재하는 데이터 필드를 제외한 나머지 데이터 필드를 수정 대상으로 하여 수정 작업을 진행할 수 있다. 예를 들어, 하나의 가상 레코드 내에 10개의 데이터 필드가 존재한다고 가정하였을 때, 1~5, 7~8, 10번째 데이터 필드는 정상 데이터에서 추출된 특정 데이터 값이 존재하지만 6, 9번째 데이터 필드는 정상 데이터에서 추출된 특정 데이터 값이 존재하지 않을 수 있다. 이 경우, 데이터 처리부(108)는 정상 데이터에서 추출된 특정 데이터 값이 존재하지 않는 6, 9번째 데이터 필드만을 수정 대상으로 하여 수정 작업을 진행할 수 있다. 여기서, 데이터 처리부(108)가 정상 데이터의 데이터 필드에 포함된 모든 데이터 값을 이용하여 가상 레코드의 데이터 필드별 수정 대상 여부를 판별하는 것으로 설명하였으나, 이는 하나의 실시예에 불과하며, 데이터 처리부(108)는 정상 데이터의 데이터 필드에 포함된 데이터 중 일부 데이터 값만을 이용하여 가상 레코드의 데이터 필드별 수정 대상 여부를 판별할 수도 있다. 예를 들어, 데이터 처리부(108)는 정상 데이터의 데이터 필드별로 출현 횟수가 많은 하나 이상의 특정 데이터 값을 이용하여 가상 레코드의 데이터 필드별 수정 대상 여부를 판별할 수 있다.

- [54] 또한, 정상 데이터에서 추출된 특정 데이터 값을 가상 레코드 내 데이터 필드에 매칭시키는 과정에서, 데이터 처리부(108)는 정상 데이터의 데이터 필드별 데이터 개수를 고려하여 데이터 필드별 수정 대상 여부를 판별할 수 있다. 예를 들어, 정상 데이터의 제 1 데이터 필드에 포함된 데이터의 개수가 2개이고, 제 2 데이터 필드에 포함된 데이터의 개수가 5개이고, 제 3 데이터 필드에 포함된

데이터의 개수가 4개라고 가정하였을 때, 데이터 처리부(108)는 먼저 데이터의 개수가 가장 적은 제 1 데이터 필드에 포함된 데이터가 가상 레코드에 포함되어 있는지 여부를 판별할 수 있다. 이후, 데이터 처리부(108)는 데이터의 개수가 많은 제 3 데이터 필드, 제 2 데이터 필드에 포함된 데이터가 가상 레코드에 포함되어 있는지 여부를 순차적으로 판별할 수 있다. 데이터의 개수가 적은 데이터 필드의 경우, 가상 레코드 내 데이터 필드와의 매칭이 상대적으로 수월할 수 있다. 데이터의 개수가 많은 데이터 필드는 상대적으로 오류 발생 가능성이 높으며 매칭에 있어서도 상대적으로 많은 시간이 소요될 수 있다. 이에 따라, 데이터 처리부(108)는 데이터의 개수가 적은 데이터 필드별로 가상 레코드와의 매칭 작업을 순차적으로 수행할 수 있다.

[55] 이와 같은 과정을 통해, 데이터 처리부(108)는 가상 레코드 내에 포함된 데이터 필드 중 수정 대상이 되는 데이터 필드를 결정할 수 있다. 본 발명의 실시예들에 따르면, 정상 데이터의 데이터 패턴 및 데이터 값을 이용하여 비정상 데이터 내의 오류 지점을 효율적으로 탐색할 수 있다.

[56] 다음으로, 데이터 처리부(108)는 수정 대상으로 판별된 데이터 필드에 포함된 데이터를 수정할 수 있다. 해당 데이터 필드 내에는 의도치 않은 특수 문자(예를 들어, ",", " "), 개행 문자(예를 들어, "\n" 등이 포함되어 있을 수 있다. 데이터 처리부(108)는 수정 대상으로 판별된 데이터 필드에 포함된 데이터 내 특정 문자열(숫자열을 포함함) 또는 문자(숫자를 포함함)를 삭제하거나, 또는 기설정된 문자열 또는 문자로 치환함으로써, 데이터를 수정할 수 있다. 데이터 처리부(108)는 일반적으로 널리 알려진 텍스트 변환 기법을 이용하여 문제가 되는 특수 문자, 기호, 대소문자 등을 다른 문자, 기호 등으로 변환할 수 있다. 데이터 처리부(108)는 다양한 데이터 변환 기법을 이용하여 수정 대상으로 판별된 데이터 필드에 포함된 데이터를 수정할 수 있으며, 특정 데이터 변환 기법에 한정되어 데이터 변환을 수행하는 것은 아니다.

[57] 데이터 처리부(108)는 위와 같은 과정을 통해 가상 레코드를 수정할 수 있으며, 하나의 가상 레코드를 수정할 때마다 수정된 가상 레코드가 정상 데이터인지 여부를 판별할 수 있다. 만약, 수정된 가상 레코드가 정상 데이터로 판별될 경우, 데이터 처리부(108)는 정상 데이터로 판별된 가상 레코드를 내부 데이터베이스(미도시)에 저장할 수 있다. 데이터 처리부(108)는 나머지 가상 레코드에 대해서도 위와 같은 동일한 과정을 통해 수정 작업을 진행한 후 정상 데이터로 판별된 가상 레코드를 순차적으로 내부 데이터베이스에 저장할 수 있다. 이때, 수정된 가상 레코드가 정상 데이터가 아닌 것으로 판별될 경우, 데이터 처리부(108)는 현재까지 수정한 가상 레코드 모두를 재차 수정할 수 있다. 즉, 데이터 처리부(108)는 백트래킹(backtracking) 과정을 통해 오류 부분을 추적하여 해당 부분부터 다시 위의 수정 작업을 진행할 수 있다. 데이터의 수정 작업 과정에서 미처 발견되지 않은 오류로 인해 잘못된 결과가 생성될 수 있기 때문에, 데이터 처리부(108)는 수정된 가상 레코드를 한번 더 검증함으로써

비정상 데이터를 정상 데이터로 완전하게 변환할 수 있다.

- [58] 여기서, 데이터 처리부(108)는 상술한 메타 데이터를 기반으로 수정된 가상 레코드가 정상 데이터인지 여부를 판별할 수 있다. 예를 들어, 데이터 처리부(108)는 메타 데이터에 포함된 데이터 필드의 개수 또는 구분자의 개수를 가상 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수와 비교하거나, 또는 가상 레코드에 포함된 데이터 필드별로 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 가상 레코드가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다. 다만, 가상 레코드가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 방법이 이에 한정되는 것은 아니며, 데이터 처리부(108)는 다양한 방법을 통해 가상 레코드가 정상 데이터인지 여부를 판별할 수 있다.
- [59] 한편, 수정된 가상 레코드 모두가 데이터 처리부(108) 내 데이터베이스에 저장되는 경우, 즉 수정된 가상 레코드 모두가 정상 데이터로 판별되는 경우, 데이터 처리부(108)는 수정된 비정상 데이터를 제 1 저장부(104)에 저장할 수 있다. 이때, 데이터 처리부(108)는 수정된 비정상 데이터를 레코드(또는 가상 레코드) 단위로 제 1 저장부(104)에 순차적으로 저장할 수 있다.
- [60] 도 4는 본 발명의 일 실시예에 따른 데이터 입력 과정을 개략적으로 나타낸 도면이다. 도 4를 참조하면, 본 발명의 실시예들에 따른 데이터 처리 시스템(100)은 비정상 데이터가 발견되는 경우 해당 비정상 데이터를 제 2 저장부(106)에 임시 저장한 후, 데이터 입력을 계속하여 진행한다. 이 과정에서 정상 데이터는 제 1 저장부(104)에 저장되며, 정상 데이터의 입력이 완료되면 제 2 저장부(106)에 저장된 비정상 데이터의 수정 작업이 진행된다. 즉, 본 발명의 실시예들에 따르면, 데이터 입력 과정에서 비정상 데이터를 제 2 저장부(106)에 임시 저장하면서 정상 데이터에 대한 입력을 계속 진행하도록 함으로써, 비정상 데이터의 입력 오류에 따른 데이터 처리 시간의 지연을 미연에 방지할 수 있다. 이에 따라, 비정상 데이터로 인해 데이터 저장 및 적재 작업이 중단되는 것을 방지할 수 있으며, 기존에 저장된 정상 데이터를 무효화시킬 필요가 없게 된다. 또한, 불필요한 저장, 적재 반복 작업을 개선할 수 있다.
- [61] 도 5는 본 발명의 일 실시예에 따른 비정상 데이터가 정상 데이터와 동일한 구조를 갖도록 수정된 결과를 나타낸 도면이다.
- [62] 도 5의 (a)는 상술한 비정상 데이터의 예시를 나타낸 도면이다. 상술한 바와 같이, 첫 번째 레코드(①)는, 특정 데이터 필드 내에 구분자인 ","가 포함됨에 따라 데이터 필드가 의도치 않게 구분되는 현상이 발생하여 기설정된 데이터 구조에 맞지 않는 비정상 데이터로 판별되었다. 두 번째 레코드(②)는, 특정 데이터 필드 내에 행을 나누는 개행 문자가 포함됨에 따라 행이 잘못된 부분에서 나누어지는 현상이 발생하여 기설정된 데이터 구조에 맞지 않는 비정상 데이터로 판별되었다.
- [63] 도 5의 (b)는 (a)의 비정상 데이터를 수정하여 정상 데이터로 변환한 결과를

나타낸 도면이다. 첫 번째 레코드(①)의 경우 데이터 "Push 요청으로 인한 CPU 사용량 과다로 CPU RUNQUEUE 초과" 와 데이터 "EBD 등록 요청" 사이의 구분자 ","를 다른 특수 문자 "."로 치환하였으며, 두 번째 레코드(②)의 경우 데이터 "○ 발생원인 : 서비스 사용량 증가에 따른 테이블스페이스 임계치 초과사용현상"과 데이터 "○ 조치내용 : 테이블스페이스에 30GB raw device add" 사이의 개행 문자 "\n"을 공백 문자 " "로 치환하였다. 다만, 이와 같은 수정 방식은 하나의 실시예에 불과하며, 데이터 처리부(108)는 다양한 방식을 활용하여 수정 대상이 되는 레코드 내 데이터를 수정할 수 있다.

- [64] 도 6은 본 발명의 일 실시예에 따른 입력 데이터가 저장 및 적재되는 최종 결과를 나타낸 도면이다. 상술한 과정을 통해 수정된 비정상 데이터는 제 1 저장부(104)에 저장될 수 있다. 도 6을 참조하면, 기존의 비정상 데이터는 데이터 처리부(108)에서의 수정 과정을 거쳐 정상 데이터로 변환되었다. 즉, 기존의 비정상 데이터가 정상 데이터와 동일한 구조를 갖도록 변환되었음을 확인할 수 있다.
- [65] 도 7은 본 발명의 일 실시예에 따른 데이터 처리 방법을 설명하기 위한 순서도이다. 먼저, 데이터 처리 시스템(100)의 데이터 판별부(102)는 입력 데이터, 입력 데이터의 구조에 관한 메타 데이터를 입력받는다(702). 여기서, 입력 데이터(150)는 분석 대상이 되는 데이터로서 예를 들어, 특정 시스템의 로그 데이터, 장애 발생 상황에 관한 데이터 등이 될 수 있다. 메타 데이터(160)는 입력 데이터(150)의 구조에 관한 정보로서, 레코드별 데이터 필드의 개수에 관한 정보, 데이터 필드 각각을 구분하는 구분자의 개수에 관한 정보, 및 데이터 필드별 데이터 패턴에 관한 정보 중 하나 이상을 포함할 수 있다.
- [66] 다음으로, 데이터 판별부(102)는 입력 데이터를 읽어들인 후(S704), 메타 데이터를 기반으로 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별한다(S706). 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 과정은 도 8 및 도 9를 참조하여 구체적으로 설명하기로 한다.
- [67] 다음으로, 데이터 판별부(102)는 정상 데이터를 제 1 저장부(104)에 저장하고(S708), 비정상 데이터를 제 2 저장부(106)에 저장한다(S710). 즉, 데이터 판별부(102)는 데이터 입력 단계에서 입력 데이터(150)를 정상 데이터와 비정상 데이터로 분류한 후, 정상 데이터와 비정상 데이터를 별도의 저장부에 각각 저장할 수 있다. 여기서, 데이터 판별부(102)가 정상 데이터를 제 1 저장부(104)에 저장한 후 비정상 데이터를 제 2 저장부(106)에 저장하는 것으로 설명하였으나, 이에 한정되는 것은 아니며 데이터 판별부(102)는 비정상 데이터를 제 2 저장부(106)에 저장한 후 정상 데이터를 제 1 저장부(104)에 저장할 수도 있으며, 정상 데이터와 비정상 데이터를 제 1 저장부(104)와 제 2 저장부(106)에 동시에 저장할 수도 있다.
- [68] 데이터 판별부(102)는 입력 데이터를 순차적으로 입력 받고, 입력 데이터가

- 마지막 입력 데이터인지 여부를 판별한다(S712). 입력 데이터가 마지막 데이터가 아닐 경우, S704 단계로 돌아가 상술한 단계를 반복한다.
- [69] 입력 데이터가 마지막 데이터일 경우, 데이터 처리부(108)는 정상 데이터와 동일한 구조를 갖도록 제 2 저장부(106)에 저장된 비정상 데이터를 수정한다(S714). 데이터 처리부(108)가 비정상 데이터를 수정하는 과정은 도 10을 참조하여 구체적으로 설명하기로 한다.
- [70] 도 8은 도 7의 S706의 예시적인 실시예를 설명하기 위한 순서도이다. 먼저, 데이터 판별부(102)는 입력 데이터 내 구분자의 수를 계산한다(S802). 구분자는 입력 데이터의 데이터 필드 각각을 구분하기 위한 문자, 기호, 부호 등을 의미하며, 예를 들어 콤마(",")일 수 있다.
- [71] 다음으로, 데이터 판별부(102)는 입력 데이터에 포함된 구분자의 개수와 메타 데이터에 포함된 데이터 필드의 개수를 비교한다(S804).
- [72] 만약, 입력 데이터에 포함된 구분자의 개수와 메타 데이터에 포함된 데이터 필드의 개수가 일치하는 경우(입력 데이터의 데이터 필드 후단에도 구분자가 존재하는 것으로 가정), 데이터 판별부(102)는 입력 데이터를 정상 데이터로 판별한다(S806). 다만, 이에 한정되는 것은 아니며, 데이터 판별부(102)는 입력 데이터에 포함된 구분자의 개수가 메타 데이터에 포함된 데이터 필드의 개수보다 하나 작은 경우(구분자가 입력 데이터의 데이터 필드 사이에만 존재하는 것으로 가정), 데이터 판별부(102)는 입력 데이터를 정상 데이터로 판별할 수도 있다.
- [73] 만약, 입력 데이터에 포함된 구분자의 개수와 메타 데이터에 포함된 데이터 필드의 개수가 일치하지 않는 경우, 입력 데이터를 비정상 데이터로 판별한다(S808). 한편, 여기서는, 데이터 판별부(102)가 입력 데이터에 포함된 구분자의 개수와 메타 데이터에 포함된 데이터 필드의 개수를 비교함으로써 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판단하는 것으로 설명하였으나, 이는 하나의 실시예에 불과하다. 데이터 판별부(102)는 예를 들어, 입력 데이터에 포함된 데이터 필드의 개수와 메타 데이터에 포함된 데이터 필드의 개수를 비교함으로써 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판단할 수도 있다.
- [74] 도 9는 도 7의 S706의 다른 예시적인 실시예를 설명하기 위한 순서도이다. 먼저, 데이터 판별부(102)는 입력 데이터의 레코드에 포함된 데이터 필드별로 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단한다(S902, S904).
- [75] 만약, 입력 데이터의 레코드에 포함된 데이터 필드 중 메타 데이터에 포함된 데이터 패턴과 매칭되는 데이터 필드가 존재하는 경우, 데이터 판별부(102)는 매칭되는 데이터 필드를 제외한 입력 데이터의 데이터 필드를 새로운 비교 대상으로 설정한다(S906).
- [76] 이 과정에서, 데이터 판별부(102)는 남은 비교 대상이 존재하는지 여부를

판단한다(S908). 만약, 남은 비교 대상이 존재할 경우, 데이터 판별부(102)는 기 매칭된 데이터 필드를 제외한 입력 데이터의 데이터 필드 중 메타 데이터에 포함된 데이터 패턴이 매칭되는 데이터 필드가 존재하는지 여부를 판단한다. 즉, S904 단계부터 다시 반복 수행한다.

- [77] 만약, 남은 비교 대상이 존재하지 않을 경우, 데이터 판별부(102)는 입력 데이터를 정상 데이터로 판별한다(S910).
- [78] 한편, S904 단계에서, 입력 데이터의 레코드에 포함된 데이터 필드 중 메타 데이터에 포함된 데이터 패턴이 매칭되는 데이터 필드가 존재하지 않을 경우, 데이터 판별부(102)는 입력 데이터를 비정상 데이터로 판별한다(S912).
- [79] 도 10은 도 7의 S714의 예시적인 실시예를 설명하기 위한 순서도이다. 먼저, 데이터 처리부(108)는 제 2 저장부(106)에 저장된 비정상 데이터를 획득하고, 제 1 저장부(104)에 저장된 정상 데이터 내의 특정 데이터 패턴 및 데이터 필드별 특정 데이터 값을 추출한다(S1002). 상술한 바와 같이, 입력 데이터의 데이터 필드 각각에는 특정 패턴을 갖는 데이터가 포함되어 있다. 예를 들어, 입력 데이터의 레코드 시작 부분("작성 시간"에 관한 데이터 필드)은 항상 "YYYYMMDDhhmmss" 와 같이 14개의 숫자를 포함하는 숫자열 형태의 패턴을 가질 수 있다. 데이터 처리부(108)는 정상 데이터 내에서 임의의 특정 데이터 패턴을 추출할 수 있다. 또한, 정상 데이터 내의 데이터 필드 각각에는 서로 다른 개수의 데이터 값이 존재할 수 있으며, 데이터 처리부(108)는 정상 데이터 내의 데이터 필드별로 하나 이상의 특정 데이터 값(예를 들어, "한국", "미국", "일본"/"남자", "여자"/"URL 모니터링 변경 조치", "해당 부품 여유분 확보" 등)를 추출할 수 있다.
- [80] 다음으로, 데이터 처리부(108)는 비정상 데이터의 일부분을 읽어들인 후(S1004), 비정상 데이터 내의 데이터 필드 중 추출된 특정 데이터 패턴과 매칭되는 데이터 필드가 있는지 검색한다(S1006).
- [81] 만약, 추출된 특정 데이터 패턴과 매칭되는 데이터 필드가 발견되지 않을 경우, 데이터 처리부(108)는 매칭되는 데이터 필드가 발견될 때까지 비정상 데이터의 범위를 계속 추가하여 검색한다(S1008).
- [82] 만약, 추출된 특정 데이터 패턴과 매칭되는 데이터 필드가 발견되는 경우, 데이터 처리부(108)는 추출된 특정 데이터 패턴과 매칭되는 다음 데이터 필드까지의 데이터를 하나의 가상 레코드로 설정한다(S1010). 예를 들어, 비정상 데이터 내에 "YYYYMMDDhhmmss" 패턴이 5번 발견되는 경우, "YYYYMMDDhhmmss" 패턴이 발견된 데이터 필드를 기준으로 해당 비정상 데이터를 5개의 가상 레코드로 설정할 수 있다. 데이터 처리부(108)는 가상 레코드 단위로 비정상 데이터를 순차적으로 수정할 수 있다.
- [83] 다음으로, 데이터 처리부(108)는 정상 데이터의 데이터 필드별 데이터 개수를 고려하여 가상 레코드 내의 데이터 필드 중 정상 데이터에서 추출된 특정 데이터 값이 존재하는 데이터 필드를 검색한다(S1012, S1014). 예를 들어, 정상 데이터의

제 1 데이터 필드에 포함된 데이터의 개수가 2개이고, 제 2 데이터 필드에 포함된 데이터의 개수가 5개이고, 제 3 데이터 필드에 포함된 데이터의 개수가 4개라고 가정하였을 때, 데이터 처리부(108)는 먼저 데이터의 개수가 가장 적은 제 1 데이터 필드에 포함된 데이터가 가상 레코드에 포함되어 있는지 여부를 판별할 수 있다. 이후, 데이터 처리부(108)는 데이터의 개수가 많은 제 3 데이터 필드, 제 2 데이터 필드에 포함된 데이터가 가상 레코드에 포함되어 있는지 여부를 순차적으로 판별할 수 있다. 데이터의 개수가 적은 데이터 필드의 경우, 가상 레코드 내 데이터 필드와의 매칭이 상대적으로 수월할 수 있다. 데이터의 개수가 많은 데이터 필드는 상대적으로 오류 발생 가능성이 높으며 매칭에 있어서도 상대적으로 많은 시간이 소요될 수 있다. 이에 따라, 데이터 처리부(108)는 데이터의 개수가 적은 데이터 필드별로 가상 레코드와의 매칭 작업을 순차적으로 수행할 수 있다.

- [84] 다음으로, 데이터 처리부(108)는 수정 대상으로 판별된 데이터 필드에 포함된 데이터를 수정한다(S1016). 데이터 처리부(108)는 수정 대상으로 판별된 데이터 필드에 포함된 데이터 내 특정 문자열 또는 문자를 삭제하거나, 또는 기설정된 문자열 또는 문자로 치환함으로써, 데이터를 수정할 수 있다. 데이터 처리부(108)는 일반적으로 널리 알려진 텍스트 변환 기법을 이용하여 문제가 되는 특수 문자, 기호, 대소문자 등을 다른 문자, 기호 등으로 변환할 수 있다.
- [85] 다음으로, 데이터 처리부(108)는 수정된 가상 레코드가 정상 데이터인지 여부를 판별한다(S1018). 즉, 데이터 처리부(108)는 하나의 가상 레코드를 수정할 때마다 수정된 가상 레코드가 정상 데이터인지 여부를 판별할 수 있다. 데이터 처리부(108)는 상술한 메타 데이터를 기반으로 수정된 가상 레코드가 정상 데이터인지 여부를 판별할 수 있다. 예를 들어, 데이터 처리부(108)는 메타 데이터에 포함된 데이터 필드의 개수 또는 구분자의 개수를 가상 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수와 비교하거나, 또는 가상 레코드에 포함된 데이터 필드별로 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 가상 레코드가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별할 수 있다.
- [86] 만약, 수정된 가상 레코드가 정상 데이터로 판별될 경우, 데이터 처리부(108)는 정상 데이터로 판별된 가상 레코드를 내부 데이터베이스(미도시)에 저장한다(S1020).
- [87] 데이터 처리부(108)는 남은 가상 레코드가 존재하는지 여부를 판단한 후(S1022), 남은 가상 레코드가 존재할 경우 S1004 단계로 돌아가 나머지 가상 레코드에 대해서도 위와 같은 동일한 과정을 통해 수정 작업을 진행한 후 정상 데이터로 판별된 가상 레코드를 순차적으로 내부 데이터베이스에 저장할 수 있다.
- [88] 한편, S1018 단계에서, 수정된 가상 레코드가 정상 데이터가 아닌 것으로 판별될 경우, 데이터 처리부(108)는 가장 최근에 수정한 가상 레코드를 읽어들이

후(S1024), 이를 이전 가상 레코드와 결합할 수 있다(S1026). 즉, 데이터 처리부(108)는 현재까지 수정한 가상 레코드 모두를 재차 수정하기 위해 S1012 단계부터 다시 진행한다. 데이터 처리부(108)는 백트래킹(backtracking) 과정을 통해 오류 부분을 추적하여 해당 부분부터 다시 위의 수정 작업을 진행할 수 있다. 데이터의 수정 작업 과정에서 미처 발견되지 않은 오류로 인해 잘못된 결과가 생성될 수 있기 때문에, 데이터 처리부(108)는 수정된 가상 레코드를 한번 더 검증함으로써 비정상 데이터를 정상 데이터로 완전하게 변환할 수 있다.

[89] 한편, S1022 단계에서 남은 가상 레코드가 존재하지 않을 경우, 데이터 처리부(108)는 수정된 비정상 데이터를 제 1 저장부(104)에 저장한다(S1028). 이때, 데이터 처리부(108)는 비정상 레코드를 레코드 단위로 제 1 저장부(104)에 순차적으로 저장할 수 있다.

[90] 도 7 내지 도 10에 도시된 방법은 데이터 처리 시스템(100)에 의해 수행될 수 있다. 도시된 순서도에서는 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. 또한 실시예에 따라 상기 방법에 도시되지 않은 하나 이상의 단계들이 상기 방법과 함께 수행될 수도 있다.

[91] 이상에서 대표적인 실시예를 통하여 본 발명에 대하여 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

[92] [부호의 설명]

[93] 100: 데이터 처리 시스템

[94] 102: 데이터 판별부

[95] 104: 제 1 저장부

[96] 106: 제 2 저장부

[97] 108: 데이터 처리부

[98] 150: 입력 데이터

[99] 160: 메타 데이터

청구범위

- [청구항 1] 입력 데이터, 및 상기 입력 데이터의 구조에 관한 메타 데이터를 입력받고, 상기 메타 데이터를 기반으로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 데이터 판별부;
상기 정상 데이터가 저장되는 제 1 저장부;
상기 비정상 데이터가 저장되는 제 2 저장부; 및
상기 정상 데이터와 동일한 구조를 갖도록 상기 제 2 저장부에 저장된 상기 비정상 데이터를 수정하고, 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는 데이터 처리부를 포함하는, 데이터 처리 시스템.
- [청구항 2] 청구항 1에 있어서,
상기 데이터 판별부는, 상기 입력 데이터의 레코드(record) 단위로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는, 데이터 처리 시스템.
- [청구항 3] 청구항 2에 있어서,
상기 메타 데이터는, 레코드별 데이터 필드(data field)의 개수에 관한 정보, 상기 데이터 필드 각각을 구분하는 구분자의 개수에 관한 정보, 및 상기 데이터 필드별 데이터 패턴에 관한 정보 중 하나 이상을 포함하는, 데이터 처리 시스템.
- [청구항 4] 청구항 3에 있어서,
상기 데이터 판별부는, 상기 메타 데이터에 포함된 데이터 필드의 개수 또는 상기 구분자의 개수 중 하나와 상기 입력 데이터의 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수 중 하나를 비교하거나, 또는 상기 입력 데이터의 레코드에 포함된 데이터 필드별로 상기 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 상기 입력 데이터의 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는, 데이터 처리 시스템.
- [청구항 5] 청구항 1에 있어서,
상기 데이터 처리부는, 상기 정상 데이터 내의 특정 데이터 패턴을 추출하고, 상기 비정상 데이터 내의 데이터 필드 중 상기 특정 데이터 패턴과 매칭되는 데이터 필드를 기준으로 상기 비정상 데이터를 복수 개의 가상 레코드로 구분하며, 상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는, 데이터 처리 시스템.
- [청구항 6] 청구항 5에 있어서,
상기 데이터 처리부는, 상기 정상 데이터 내의 데이터 필드별로

하나 이상의 특정 데이터 값을 추출하고, 상기 가상 레코드 내의 데이터 필드 중 상기 특정 데이터 값이 존재하는 데이터 필드를 제외한 나머지 데이터 필드에 포함된 데이터를 수정하는, 데이터 처리 시스템.

[청구항 7]

청구항 6에 있어서,

상기 데이터 처리부는, 상기 나머지 데이터 필드에 포함된 데이터 내 특정 문자열 또는 문자를 삭제하거나, 또는 기설정된 문자열 또는 문자로 치환하는, 데이터 처리 시스템.

[청구항 8]

청구항 5에 있어서,

상기 데이터 처리부는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상 데이터인지 여부를 판별하고, 상기 가상 레코드가 정상 데이터로 판별된 경우 상기 가상 레코드를 상기 데이터 처리부 내 데이터베이스에 저장하며, 가상 레코드 모두가 상기 데이터베이스에 저장된 경우 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는, 데이터 처리 시스템.

[청구항 9]

청구항 5에 있어서,

상기 데이터 처리부는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상 데이터인지 여부를 판별하고, 상기 가상 레코드가 정상 데이터가 아닌 것으로 판별된 경우 현재까지 수정한 가상 레코드 모두를 재차 수정하는, 데이터 처리 시스템.

[청구항 10]

데이터 판별부가, 입력 데이터, 및 상기 입력 데이터의 구조에 관한 메타 데이터를 입력받는 단계;

상기 데이터 판별부가, 상기 메타 데이터를 기반으로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 단계;

상기 데이터 판별부가, 상기 정상 데이터를 제 1 저장부에 저장하는 단계;

상기 데이터 판별부가, 상기 비정상 데이터를 제 2 저장부에 저장하는 단계;

데이터 처리부가, 상기 정상 데이터와 동일한 구조를 갖도록 상기 제 2 저장부에 저장된 상기 비정상 데이터를 수정하는 단계; 및
상기 데이터 처리부가, 수정된 상기 비정상 데이터를 상기 제 1 저장부에 저장하는 단계를 포함하는, 데이터 처리 방법.

[청구항 11]

청구항 10에 있어서,

상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 단계는, 상기 입력 데이터의 레코드(record) 단위로 상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는, 데이터 처리 방법.

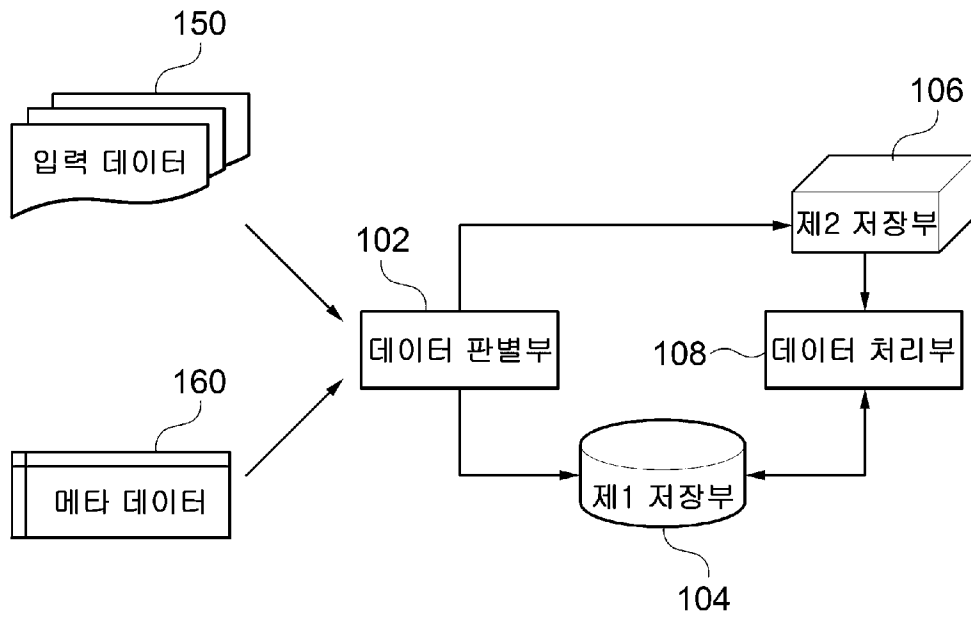
- [청구항 12] 청구항 11에 있어서,
상기 메타 데이터는, 레코드별 데이터 필드(data field)의 개수에 관한 정보, 상기 데이터 필드 각각을 구분하는 구분자의 개수에 관한 정보, 및 상기 데이터 필드별 데이터 패턴에 관한 정보 중 하나 이상을 포함하는, 데이터 처리 방법.
- [청구항 13] 청구항 12에 있어서,
상기 입력 데이터가 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는 단계는, 상기 메타 데이터에 포함된 데이터 필드의 개수 또는 상기 구분자의 개수 중 하나와 상기 입력 데이터의 레코드에 포함된 데이터 필드의 개수 또는 구분자의 개수 중 하나를 비교하거나, 또는 상기 입력 데이터의 레코드에 포함된 데이터 필드별로 상기 메타 데이터에 포함된 데이터 필드별 데이터 패턴이 매칭되는지 여부를 판단함으로써, 상기 입력 데이터의 레코드 각각이 정상 데이터인지 또는 비정상 데이터인지 여부를 판별하는, 데이터 처리 방법.
- [청구항 14] 청구항 10에 있어서,
상기 비정상 데이터를 수정하는 단계는,
상기 정상 데이터 내의 특정 데이터 패턴을 추출하는 단계;
상기 비정상 데이터 내의 데이터 필드 중 상기 특정 데이터 패턴과 매칭되는 데이터 필드를 기준으로 상기 비정상 데이터를 복수 개의 가상 레코드로 구분하는 단계; 및
상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계를 포함하는, 데이터 처리 방법.
- [청구항 15] 청구항 14에 있어서,
상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계는, 상기 정상 데이터 내의 데이터 필드별로 하나 이상의 특정 데이터 값을 추출하는 단계; 및
상기 가상 레코드 내의 데이터 필드 중 상기 특정 데이터 값이 존재하는 데이터 필드를 제외한 나머지 데이터 필드에 포함된 데이터를 수정하는 단계를 포함하는, 데이터 처리 방법.
- [청구항 16] 청구항 15에 있어서,
상기 나머지 데이터 필드에 포함된 데이터를 수정하는 단계는, 상기 나머지 데이터 필드에 포함된 데이터 내 특정 문자열 또는 문자를 삭제하거나, 또는 기설정된 문자열 또는 문자로 치환하는, 데이터 처리 방법.
- [청구항 17] 청구항 14에 있어서,
상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계는, 상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상

데이터인지 여부를 판별하는 단계;
상기 가상 레코드가 정상 데이터로 판별된 경우 상기 가상
레코드를 상기 데이터 처리부 내 데이터베이스에 저장하는 단계;
및
가상 레코드 모두가 상기 데이터베이스에 저장된 경우 수정된
상기 비정상 데이터를 상기 제 1 저장부에 저장하는 단계를
포함하는, 데이터 처리 방법.

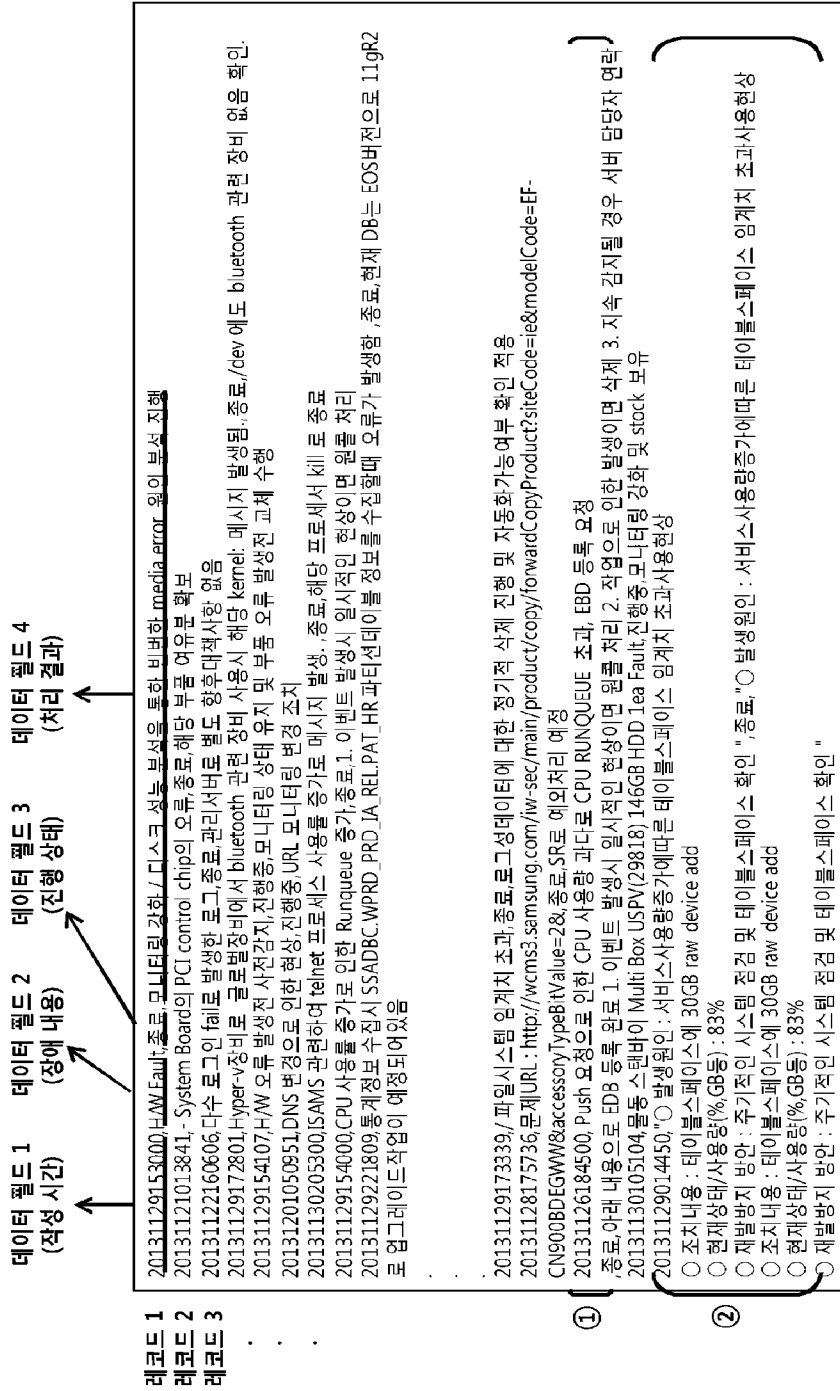
[청구항 18]

청구항 14에 있어서,
상기 가상 레코드 단위로 상기 비정상 데이터를 수정하는 단계는,
상기 가상 레코드를 수정할 때마다 상기 가상 레코드가 정상
데이터인지 여부를 판별하는 단계; 및
상기 가상 레코드가 정상 데이터가 아닌 것으로 판별된 경우
현재까지 수정한 가상 레코드 모두를 재차 수정하는 단계를
포함하는, 데이터 처리 방법.

[Fig. 1]
100



[Fig. 2]



[Fig. 3]

정상 데이터

작성 시간	장애 내용	진행 상태	처리 결과
20131129153000	H/W Fault	종료	모니터링 강화 / 디스크 성능 분석을 통한 빈번한 media error 원인 분석 진행
20131121013841	- System Board의 PCI control chip의 오류	종료	해당 부품 여유분 확보
20131122160606	다수 로그이 테일로 발생한 로그	종료	관리서버로 별도 향후대책사항 없음
20131129172801	Hyper-v장비로 글로벌장비에서 bluetooth 관련 장비 사용시 해당 kernel 메시지가 발생됨.	종료	/dev 에도 bluetooth 관련 장비 없음 확인.
20131129154107	H/W 오류 발생전 사전감지	진행중	모니터링 상태 유지 및 부품 오류 발생전 교체 수행
20131201050951	DNS 변경으로 인한 현상	진행중	URL 모니터링 변경 조치
20131130205300	ISAMS 권력하여 telnet 프로세스 사용률 증가로 메시지가 발생.	종료	해당 프로세서 클로 종료
20131129154000	CPU 사용률 증가로 인한 Runqueue 증가	종료	1. 이벤트 발생시 일시적인 현상이며 원콜 처리
20131129221809	통계정보 수집시 SSADBC.WPRD_PRD_IA_REL.PAT_HR 파티션 테이블 정보를 수집할때 오류가 발생한	종료	현재 DB는 EOS버전으로 11gR2로 업그레이드작업이 예정되어있음

(a)

데이터 필드 1

데이터 필드 2

데이터 필드 3

데이터 필드 4

비정상 데이터

작성 시간	장애 내용	진행 상태	처리 결과
20131126184500	Push 요청으로 인한 CPU 사용량 과다로 CPU RUNQUEUE 초과	EBD 등록 요청	종료
20131129014450	○ 발생원인 : 서버스 사용량증가에따른 테이블스페이스 임계치 초과사용현상		아래 내용으로 EDB 등록 완료 1. 이벤트 발생시 일시적인 현상이면 원콜 처리 2. 작업으로 인한 발생이면 삭제 3. 지속 감지될 경우 서버 담당자 연락
20131129014450	○ 조치내용 : 테이블스페이스 30GB raw device add		
20131129014450	○ 현재상태/사용량(%GB등) : 83%		
20131129014450	○ 재발방지 방안 : 주기적인 시스템 점검 및 테이블스페이스 확인		

①

(b)

[Fig. 5]

비정상 데이터

작성 시간	장애 내용	진행 상태	처리 결과	처리 결과
20131126184500	Push 요청으로 인한 CPU 사용량 과다로 CPU RUNQUEUE 초과	EBD 등록 요청	종료	아래 내용으로 EDB 등록 완료 1. 이벤트 발생시 일시적인 현상이면 원인을 처리 2. 작업으로 인한 발생이면 삭제 3. 지속 감지될 경우 서버 담당자 연락
20131129014450	발생 원인 : 서비스사용량증가에따른 테이블스페이스 임계치 초과 사용현상			
20131129014450	조치내용 : 테이블스페이스에 30GB raw device add 현재상태/사용량(%GB등) : 83%			
20131129014450	재발방지 방안 : 주기적인 시스템 점검 및 테이블스페이스 확인			

①

②

(a)

① 열의 구분 문자가 ""이었으나 데이터 안에 ""가 사용되어서 열이 나누어짐

② 행을 나누는 개행 문자가 데이터에 포함되어 이 행이 잘못된 부분에서 나누어짐



비정상 데이터 수정

비정상 데이터의 수정 및 구조화 결과

작성 시간	장애 내용	진행 상태	처리 결과
20131126184500	Push 요청으로 인한 CPU 사용량 과다로 CPU RUNQUEUE 초과, EBD 등록 요청	구분문자 수정	아래 내용으로 EDB 등록 완료 1. 이벤트 발생시 일시적인 현상이면 원인을 처리 2. 작업으로 인한 발생이면 삭제 3. 지속 감지될 경우 서버 담당자 연락
20131129014450	발생 원인 : 서비스사용량증가에따른 테이블스페이스 임계치 초과사용현상 조치내용 : 테이블스페이스에 30GB raw device add 현재상태/사용량(%GB등) : 83% 재발방지 방안 : 주기적인 시스템 점검 및 테이블스페이스 확인	개행문자 수정	종료

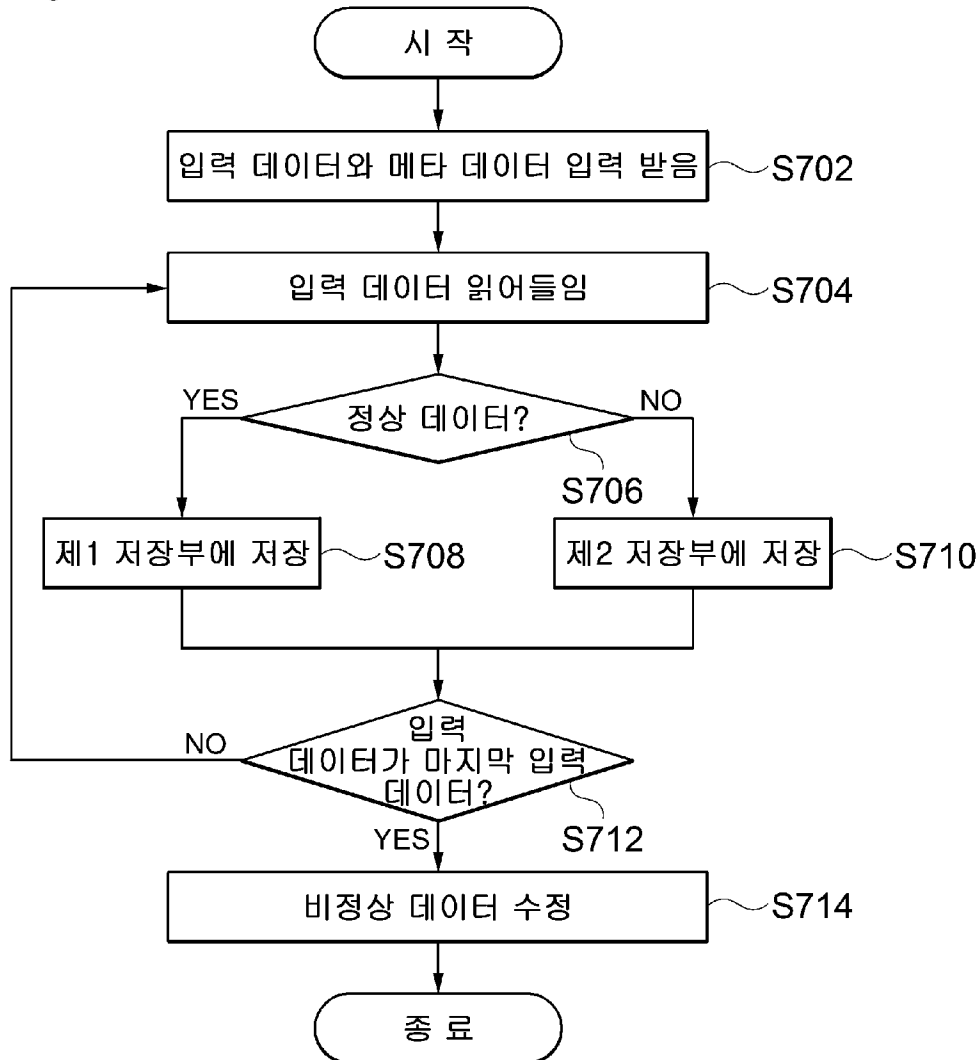
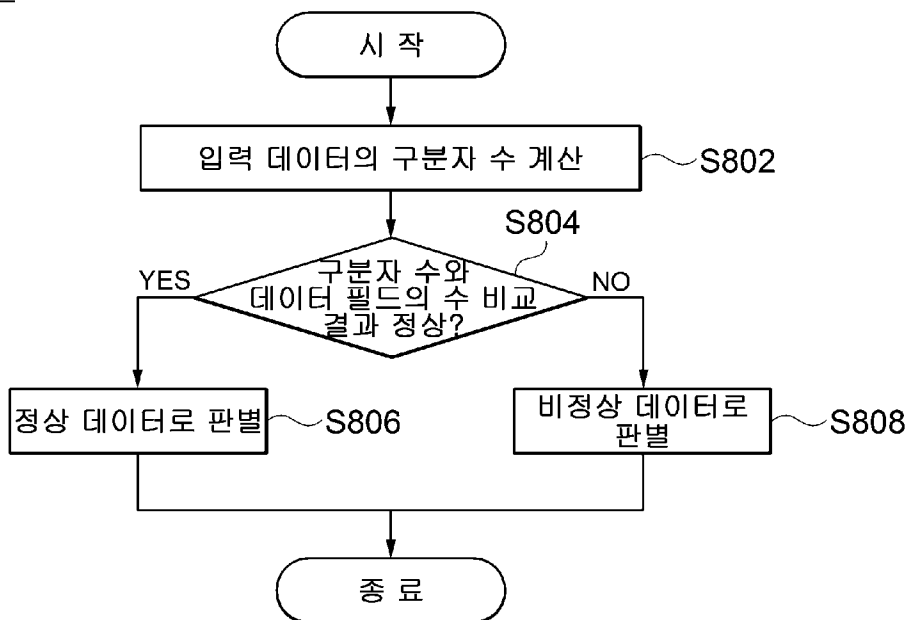
(b)

- ① : 구분자와 겹치는 문자를 사전 설정한 다른 문자로 변환 (예: ", "(콤마) → ". "(마침표))
- ② : 개행 문자를 사전 설정한 다른 문자로 변환 (예: "/n"(개행문자) → " " (공백문자))

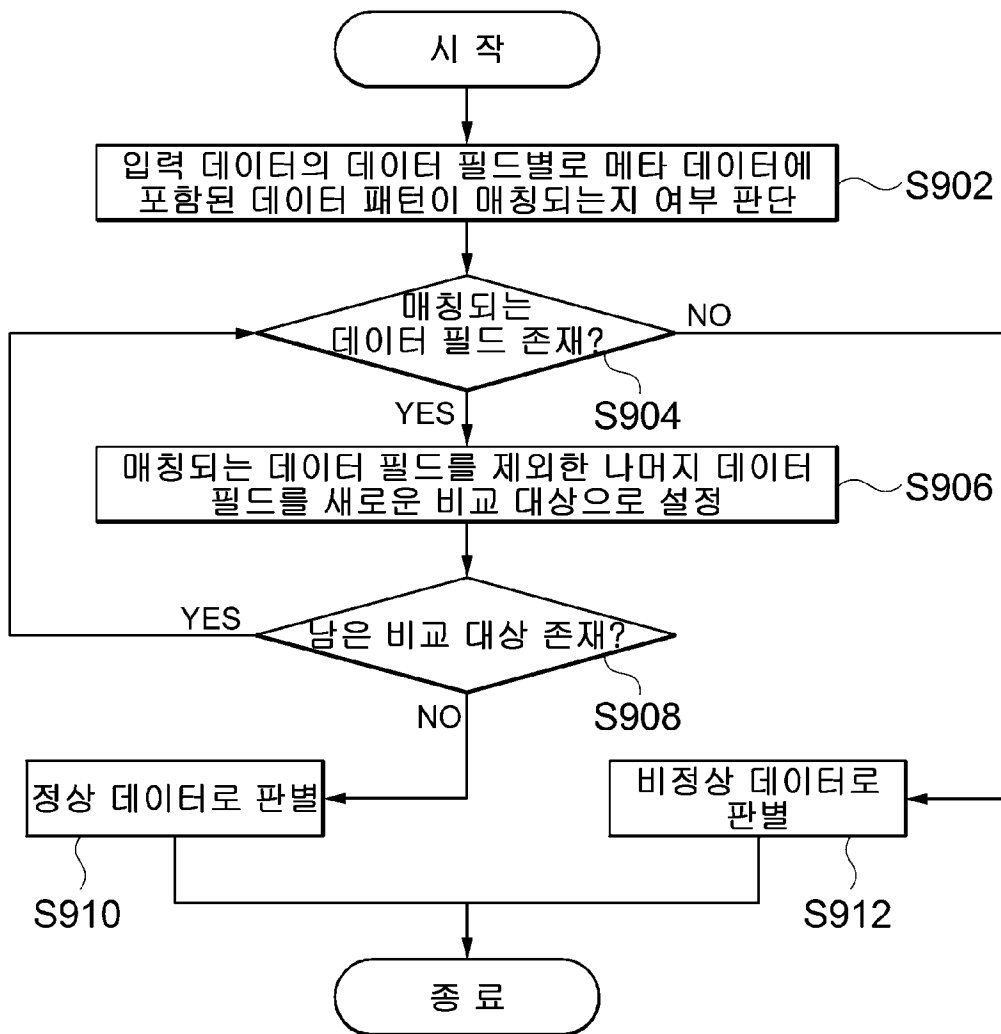
[Fig. 6]

작성 시간	장애 내용	진행 상태	처리 결과
20131129153000	H/W Fault	종료	모니터링 강화 / 디스크 성능 분석을 통한 빈번한 media error 원인 분석 진행
20131121013841	System Board의 PCI control chip의 오류	종료	해당 부품 여유분 확보
20131122160606	다수 로그인 fail로 발생한 로그	종료	관리서버로 별도 향후대책사항 없음
20131129172801	Hyper-v장비로 글로벌장치에서 bluetooth 관련 장비 사용	종료	/dev 에도 bluetooth 관련 장비 없음 확인
20131129154107	시 해당 kernel 메시지 발생됨	지행중	모니터링 상태 유지 및 부품 오류 발생전 교체 수행
20131201050951	DNS 변경으로 인한 현상	지행중	URL 모니터링 변경 조치
20131130205300	SAMS 관련하여 telnet 프로세스 사용률 증가로 메시지 발생	종료	해당 프로세서 kill 로 종료
20131129154000	CPU 사용률 증가로 인한 Runqueue 증가	종료	1. 이벤트 발생시 일시적인 현상이면 워크 처리
20131129221809	통계정보 수집시 SSADBC.WPRD_PRD_IA_REL.PAT_HR 파티션 데이터베이스 정보를 수집할때 오류가 발생함	종료	현재 DB는 EOS버전으로 11qR2로 업그레이드작업이 예정되어있음
20131129173339	파일시스템 임계치 초과	종료	로그성데이터에 대한 정기적 삭제 진행 및 자동화기능여부 확인 적용
20131128175736	문제URL : http://wcmcs3.samsung.com/iw-sec/main/product/copy/forwardCopyProduct?siteCode=ie&modelCode=EF-CN900BDEGWW&accessoryTypeBitValue=2&	종료	SR로 예외처리 예정
20131126184500	Push 요청으로 인한 CPU 사용량 과다로 CPU RUNQUEUE 초과, EBD 등록 요청	종료	아래 내용으로 EDB 등록 완료 1. 이벤트 발생시 일시적인 현상이면 워크 처리 2. 작업으로 인한 발생이면 삭제 3. 지속 감지될 경우 서버 담당자 연락
20131130105104	물동 스택바이 Multi Box USP(29818) 146GB HDD 1ea Fault	지행중	모니터링 강화 및 stock 보유
20131129014450	스택인	종료	

[Fig. 7]

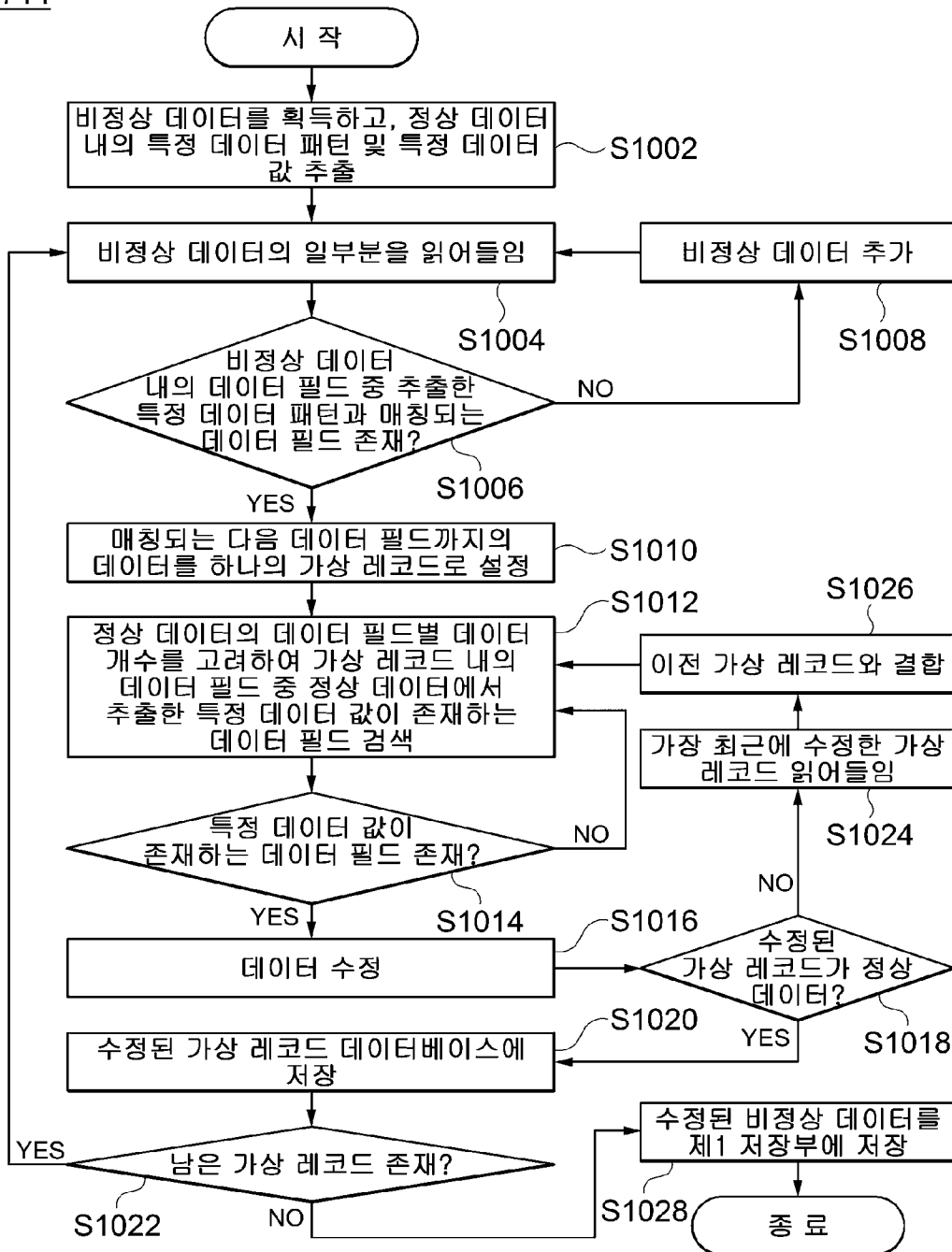
[Fig. 8]
S706

[Fig. 9]
S706



[Fig. 10]

S714



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2014/011091**A. CLASSIFICATION OF SUBJECT MATTER****G06F 11/08(2006.01)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 11/08; none ; G06F 17/30; G07F 7/10; G06F 12/00; G06F 17/40

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: input, meta data, normal data, abnormal data, distinction, structure, data processing

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 2014-051348 A2 (SAMSUNG SDS CO., LTD.) 03 April 2014 See paragraphs [0014], [0033]-[0040], [0052]-[0053], [0063]-[0066]; and figures 3, 10.	1-18
Y	KR 10-2008-0094081 A (ADVANCED MICRO DEVICES INC.) 22 October 2008 See claims 1, 4.	1-18
A	US 2013-0311456 A1 (WINKLER, Patrick et al.) 21 November 2013 See paragraphs [0004]-[0015], [0038]-[0042]; and figure 2.	1-18
A	KR 10-2011-0097921 A (AB INITIO TECHNOLOGY LLC.) 31 August 2011 See paragraphs [0004]-[0008], [0049]-[0051]; and figure 2.	1-18
A	US 2013-0117518 A1 (KOBAYASHI, Tomohiro) 09 May 2013 See paragraphs [0061]-[0068]; and figure 4.	1-18



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

"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

13 MARCH 2015 (13.03.2015)

Date of mailing of the international search report

13 MARCH 2015 (13.03.2015)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex-Daejeon, 189 Seonsa-ro, Daejeon 302-701,
Republic of Korea

Facsimile No. 82-42-472-7140

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2014/011091

Patent document cited in search report	Publication date	Patent family member	Publication date
WO 2014-051348 A2	03/04/2014	KR 10-2014-0042169 A US 2014-095523 A1 WO 2014-051348 A3	07/04/2014 03/04/2014 22/05/2014
KR 10-2008-0094081 A	22/10/2008	CN 101416132 A CN 101416132 B DE 112007000254 T5 GB 2450638 A JP 2009-524934 A KR 10-1331249 B1 TW 200811912 A TW 1420567 B US 07583833 B2 US 2007-0179663 A1 WO 2007-089494 A1	22/04/2009 13/06/2012 13/11/2008 31/12/2008 02/07/2009 20/11/2013 01/03/2008 21/12/2013 01/09/2009 02/08/2007 09/08/2007
US 2013-0311456 A1	21/11/2013	NONE	
KR 10-2011-0097921 A	31/08/2011	AU 2009-322602 A1 CA 2744881 A1 CN 102232212 A EP 2370892 A1 JP 05525541 B2 JP 2012-510687 A US 2010-0138388 A1 WO 2010-065511 A1	10/06/2010 10/06/2010 02/11/2011 05/10/2011 18/06/2014 10/05/2012 03/06/2010 10/06/2010
US 2013-0117518 A1	09/05/2013	EP 2590072 A1 EP 2590072 A4 JP 5376058 B2 WO 2012-001780 A1	08/05/2013 09/10/2013 25/12/2013 05/01/2012

A. 발명이 속하는 기술분류(국제특허분류(IPC))

G06F 11/08(2006.01)i

B. 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 11/08; none ; G06F 17/30; G07F 7/10; G06F 12/00; G06F 17/40

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: 입력, 메타 데이터, 정상 데이터, 비정상 데이터, 판별, 구조, 데이터 처리

C. 관련 문헌

카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
Y	WO 2014-051348 A2 (삼성에스디에스 주식회사) 2014.04.03 단락 [0014], [0033]-[0040], [0052]-[0053], [0063]-[0066]; 및 도면 3, 10 참조.	1-18
Y	KR 10-2008-0094081 A (어드밴스드 마이크로 디바이스즈, 인코포레이티드) 2008.10.22 청구항 제1,4항 참조.	1-18
A	US 2013-0311456 A1 (PATRICK WINKLER 외 2명) 2013.11.21 단락 [0004]-[0015], [0038]-[0042]; 및 도면 2 참조.	1-18
A	KR 10-2011-0097921 A (아브 이니티오 테크놀로지 엘엘시) 2011.08.31 단락 [0004]-[0008], [0049]-[0051]; 및 도면 2 참조.	1-18
A	US 2013-0117518 A1 (TOMOHIRO KOBAYASHI) 2013.05.09 단락 [0061]-[0068]; 및 도면 4 참조.	1-18

☐ 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

“T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

“E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

“X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

“L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

“Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

“O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

“&” 동일한 대응특허문헌에 속하는 문헌

“P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

국제조사의 실제 완료일

2015년 03월 13일 (13.03.2015)

국제조사보고서 발송일

2015년 03월 13일 (13.03.2015)

ISA/KR의 명칭 및 우편주소



대한민국 특허청
(302-701) 대전광역시 서구 청사로 189,
4동 (둔산동, 정부대전청사)

팩스 번호 ++82 42 472 7140

심사관

유재천

전화번호 +82-42-481-8647



국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
WO 2014-051348 A2	2014/04/03	KR 10-2014-0042169 A US 2014-095523 A1 WO 2014-051348 A3	2014/04/07 2014/04/03 2014/05/22
KR 10-2008-0094081 A	2008/10/22	CN 101416132 A CN 101416132 B DE 112007000254 T5 GB 2450638 A JP 2009-524934 A KR 10-1331249 B1 TW 200811912 A TW I420567 B US 07583833 B2 US 2007-0179663 A1 WO 2007-089494 A1	2009/04/22 2012/06/13 2008/11/13 2008/12/31 2009/07/02 2013/11/20 2008/03/01 2013/12/21 2009/09/01 2007/08/02 2007/08/09
US 2013-0311456 A1	2013/11/21	없음	
KR 10-2011-0097921 A	2011/08/31	AU 2009-322602 A1 CA 2744881 A1 CN 102232212 A EP 2370892 A1 JP 05525541 B2 JP 2012-510687 A US 2010-0138388 A1 WO 2010-065511 A1	2010/06/10 2010/06/10 2011/11/02 2011/10/05 2014/06/18 2012/05/10 2010/06/03 2010/06/10
US 2013-0117518 A1	2013/05/09	EP 2590072 A1 EP 2590072 A4 JP 5376058 B2 WO 2012-001780 A1	2013/05/08 2013/10/09 2013/12/25 2012/01/05