



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2024년07월05일
(11) 등록번호 10-2682244
(24) 등록일자 2024년07월02일

- (51) 국제특허분류(Int. Cl.)
G06Q 10/10 (2023.01) G06F 16/22 (2019.01)
G06F 40/205 (2020.01) G06F 40/284 (2020.01)
G06N 20/00 (2019.01) G06N 3/08 (2023.01)
G06Q 10/04 (2023.01) G06Q 10/06 (2012.01)
- (52) CPC특허분류
G06Q 10/10 (2023.01)
G06F 16/2272 (2019.01)
- (21) 출원번호 10-2023-0054731(분할)
- (22) 출원일자 2023년04월26일
심사청구일자 2023년04월26일
- (65) 공개번호 10-2024-0020166
- (43) 공개일자 2024년02월14일
- (62) 원출원 특허 10-2022-0097982
원출원일자 2022년08월05일
심사청구일자 2022년08월05일
- (56) 선행기술조사문헌
KR1020210108293 A*
KR1020200135607 A*
KR102026304 B1*
*는 심사관에 의하여 인용된 문헌
- (73) 특허권자
주식회사 아이이에스지
서울특별시 용산구 한강대로 366, 612호(동자동, 트윈시티남산2)
- (72) 발명자
김종웅
인천광역시 연수구 송도문화로28번길 28, 102동 401호 (송도동, 송도글로벌캠퍼스푸르지오)
- (74) 대리인
이강욱, 김성훈

전체 청구항 수 : 총 7 항

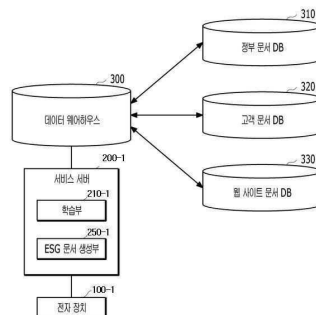
심사관 : 김은수

(54) 발명의 명칭 ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 기계학습 모델을 학습하는 방법 및 기계학습 모델로 자동완성된 ESG 문서를 생성하는 서비스 서버

(57) 요약

실시예들은 정형화 처리 대상의 ESG 문서를 파싱 처리하는 단계 - 상기 ESG 문서는 비정형 ESG 데이터를 포함함; 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하는 단계; 상기 ESG 관리 항목에 대응한 항목이 문서 내에 포함되었는지 여부에 따라 ESG 문서 내 항목에 대해 선별 상태를 지정하는 단계; 선별 상태의 지정 결과에 따라 미리 저장된 복수의 ESG 관리 항목에 대한 최종 입력 필드 값의 입력 칸을 활성화하거나 또는 비활성화하는 단계; 및 문서에 포함된 것으로 판단된, ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하는 단계를 포함한 ESG 보조 툴을 이용하여 정형화된 ESG 데이터를 기반으로, 기계학습 모델을 학습하는 방법 및 기계학습 모델로 자동완성된 ESG 문서를 생성하는 서비스 서버에 관한 것이다. .

대표도 - 도6



(52) CPC특허분류

G06F 40/205 (2020.01)

G06F 40/284 (2020.01)

G06N 20/00 (2021.08)

G06N 3/08 (2023.01)

G06Q 10/04 (2023.01)

G06Q 10/0635 (2023.01)

명세서

청구범위

청구항 1

ESG 서비스 제공 시스템 내 전자 장치(100), ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템 내 서비스 서버(200-1)에 의해 수행되는, ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 기계학습 모델을 학습하는 방법에 있어서,

상기 전자 장치(100)는 복수의 ESG 관리 항목을 미리 저장한 ESG 관리 항목 데이터베이스와 연결되고,

상기 ESG 서비스 제공 시스템과 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템은 연결되어, 상기 ESG 서비스 제공 시스템이 제공하는 ESG 서비스는 ESG 문서를 자동완성하기 위한 텍스트를 추천하는 서비스를 더 포함하고,

상기 방법은,

상기 전자 장치(100)에 의해, 정형화 처리 대상의 ESG 문서를 파싱 처리하는 단계 - 상기 ESG 문서는 비정형 ESG 데이터를 포함함;

상기 전자 장치(100)에 의해, 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하는 단계;

상기 전자 장치(100)에 의해, 스크리닝 결과에 이상이 없는지 여부를 판단하는 단계;

상기 전자 장치(100)에 의해, 상기 ESG 관리 항목에 대응한 항목이 문서 내에 포함되었는지 여부에 따라 ESG 문서 내 항목에 대해 선별 상태를 지정하는 단계;

상기 전자 장치(100)에 의해, 선별 상태의 지정 결과에 따라 미리 저장된 복수의 ESG 관리 항목에 대한 최종 입력 필드 값의 입력 칸을 활성화하거나 또는 비활성화하는 단계;

상기 전자 장치(100)에 의해, 문서에 포함된 것으로 판단된, ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하는 단계; 및

상기 서비스 서버(200-1)에 의해, 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 단계;를 포함하고,

상기 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하는 단계는,

상기 ESG 관리 항목 데이터베이스에 저장된 각 ESG 관리 항목별 코드 값 및 데이터 세트 값 중 적어도 하나를 기준으로 파싱된 ESG 문서에 해당 항목을 포함하는지 여부를 판단하는 것이고,

상기 스크리닝 결과에 이상이 없는지 여부를 판단하는 단계는,

사용자의 확인 입력에 기초하거나 또는 ESG 문서 내에서 코드 값, 데이터 세트 값 중 적어도 하나에 매칭한 데이터에 대한 주변 데이터에 기초하여 스크리닝 결과의 이상 여부를 판단하는 것이며,

상기 선별 상태를 지정하는 단계는,

파싱된 ESG 문서 내에 해당 항목이 포함되어 있으면 상기 ESG 문서를 제1 선별 상태로 지정하고, 해당 항목이 포함되어 있지 않으면 상기 ESG 문서를 제2 선별 상태로 지정하는 것이며,

상기 정형화 데이터를 생성하는 단계는,

ESG 관리 항목에 대응한 항목의 내용을 획득하여 상기 정형화 데이터를 생성하기 위해, 제1 선별 상태로 지정된 ESG 관리 항목의 상기 ESG 문서 내 위치를 제공하는 단계; 제공된 위치에 기초하여 해당 ESG 관리 항목의 내용을 획득하는 단계; 및 선별된 ESG 관리 항목의 목록 및 획득된 각각의 ESG 관리 항목의 내용을 테이블 변환 처리하여 정형 데이터를 생성하는 단계를 포함하고,

상기 해당 ESG 관리 항목의 내용을 획득하는 단계는,

사용자의 내용 입력에 기초하거나 또는 제공된 위치에 대한 주변 데이터에 기초하여 최종 필드 입력 값의 내용을 획득하는 것이고,

상기 기계학습 모델은 인코더 및 디코더를 포함하고, 입력 텍스트의 말뭉치와 잔여 텍스트의 말뭉치 간의 연관 관계를 추론하여 입력 텍스트의 말뭉치의 그 다음 순번으로 배치될 텍스트의 말뭉치를 예측하도록 학습된 것이고

상기 기계학습 모델은 복수의 트레이닝 샘플로 이루어진 트레이닝 데이터를 이용하여 학습되는 것으로서, 각 트레이닝 샘플은 동일한 샘플 텍스트의 말뭉치에 대응한 토큰 세트로부터 획득된 트레이닝 데이터, 레이블 데이터를 각각 포함하고, 상기 트레이닝 데이터는 학습을 위해 샘플 텍스트의 말뭉치로부터 획득된 토큰 세트에 포함된 토큰을 말뭉치 상의 순서대로 배열한 토큰 배열을 포함하고,

상기 인코더는 입력 토큰의 특징을 추출하도록 구성되고 - 상기 입력 토큰은 트레이닝 데이터에 포함된 토큰 배열이고,

상기 디코더는 입력 토큰의 특징에 기초하여 트레이닝 데이터의 토큰 배열 중에서 미리 설정된 순번의 텍스트 토큰으로부터 맨 끝의 텍스트 토큰을 포함한 잔여 배열 토큰을 복원하도록 구성되는 것이고,

상기 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 단계;는,

상기 서비스 서버(200-1)의 학습부(210-1)에 의해, 상기 트레이닝 데이터 세트 내 트레이닝 데이터를 상기 학습부(210-1)에 구현된 상기 기계학습 모델에 입력하여 해당 트레이닝 샘플에 대한 예측 값을 산출하는 단계 - 상기 기계학습 모델 내 인코더는 상기 트레이닝 데이터로 입력된 입력 토큰의 특징을 추출하고, 상기 기계학습 모델 내 디코더는 상기 입력 토큰의 특징에 기초하여 복원된 잔여 배열 토큰의 복원 결과를 예측 값으로 출력함; 및

상기 서비스 서버(200-1)의 학습부(210-1)에 의해, 해당 트레이닝 샘플에 대해 산출된 예측 값과 상기 해당 트레이닝 샘플에 포함된 레이블 데이터의 값인 실제 값 간의 오차를 감소하도록 상기 기계학습 모델의 파라미터를 조정하는 단계;를 포함하는 것

을 특징으로 하는,

방법.

청구항 2

제1항에 있어서,

상기 전자 장치는 다른 ESG 문서의 비정형 ESG 데이터를 정형화하여 저장한 ESG 문서 데이터베이스와 더 연결되고,

상기 ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하는 단계는,

기존 ESG 문서 내 데이터를 이용하겠다는 사용자 입력을 수신한 경우, 상기 다른 ESG 문서 내 데이터를 해당 항목에 대한 내용으로 획득하는 단계를 더 포함하는 것을 특징으로 하는,

방법.

청구항 3

제1항에 있어서, 상기 전자 장치는 인터페이스 제공부 및 출력부를 더 포함하고, 상기 방법은,

상기 인터페이스 제공부에 의해, 이미 저장된 ESG 관리 항목을 관리하는 단계를 더 포함하고 - 상기 ESG 관리 항목의 관리는 수정, 삭제 또는 생성하는 것이고,

상기 ESG 관리 항목을 관리하는 단계는,

이미 저장된 ESG 관리 항목이 저장되었는지 확인하는 단계;

저장된 경우, ESG 관리 항목에 대한 수정, 삭제 또는 생성 요청을 수신하는 단계; 및

상기 수정, 삭제 또는 생성 요청을 수신할 경우, 저장된 ESG 관리 항목 세트를 항목 레벨 구조로 나타낸 보조 틀 인터페이스 화면을 상기 출력부를 통해 표시하는 단계;를 포함하는 것을 특징으로 하는, 방법.

청구항 4

제3항에 있어서, 상기 ESG 관리 항목을 관리하는 단계는, 요청 대상 항목이 최하위 레벨 항목이 아닐 경우 대상 항목의 내용에 대한 입력 칸을 비활성화하는 단계; 상기 요청 대상 항목이 최하위 레벨 항목일 경우 대상 항목의 내용에 대한 입력 칸을 활성화하는 단계; 활성화된 입력 칸을 통해 수정 또는 생성할 대상 항목의 내용을 입력받는 단계; 및 입력된 항목의 내용을 ESG 관리 항목 데이터베이스에 저장하는 단계를 더 포함하는 것을 특징으로 하는, 방법.

청구항 5

제4항에 있어서, 상기 활성화된 입력 칸을 통해 대상 항목의 내용을 입력받는 단계는, 상기 대상 항목의 내용을 입력받기 위한 보조 틀 인터페이스 화면을 제공하는 단계를 포함하고, 상기 보조 틀 인터페이스 화면은 활성화된 입력 칸을 포함한 제1 서브 영역 및 항목 구조 레벨을 포함한 제2 서브 영역을 포함하고, 상기 제2 서브 영역에는 대분류별 항목 세트를 각각의 동일한 항목 레벨별로 분류하여 표시되는 것을 특징으로 하는, 방법.

청구항 6

청구항 제1항 내지 제5항 중 어느 하나의 항에 따른 ESG 보조 틀을 이용하여 정형화된 ESG 데이터로 기계학습 모델을 학습하는 방법을 수행하기 위한 프로그램이 기록된, 컴퓨터 판독가능한 기록매체.

청구항 7

기계학습 모델로 자동완성된 ESG 문서를 생성하는 서비스 서버(200-1)에 있어서, 상기 서비스 서버(200-1)를 포함하는 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템은 ESG 보조 틀을 이용하여 정형화된 ESG 데이터로 이루어진 ESG 데이터베이스를 구축하기 위한, 전자 장치(100)를 포함한 ESG 서비스 제공 시스템과 연결되어, 상기 ESG 서비스 제공 시스템이 제공하는 ESG 서비스는 ESG 문서를 자동완성하기 위한 텍스트를 추천하는 서비스를 더 포함하고, 상기 전자 장치(100)는 복수의 ESG 관리 항목을 미리 저장한 ESG 관리 항목 데이터베이스와 연결되고, 상기 전자 장치(100)는, 비정형 ESG 데이터를 포함한 ESG 문서 내에서 상기 ESG 관리 항목 데이터베이스에 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응한 항목 및 해당 항목의 내용을 획득하여 상기 ESG 문서에 대한 정형화된 ESG 데이터를 생성하도록 구성된 정형 데이터 처리부; 및 ESG 문서의 비정형 데이터를 정형화된 데이터로 변경하기 위한 사용자 입력을 수신하는 보조 틀 인터페이스 화면을 출력부를 통해 제공하는 인터페이스 제공부를 포함하고, 상기 정형 데이터 처리부는, 정형화 처리 대상의 ESG 문서를 파싱 처리하고 - 상기 ESG 문서는 비정형 ESG 데이터를 포함함 -, 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하며, 스크리닝 결과에 이상이 없는지 여부를 판단하고, 상기 ESG 관리 항목에 대응한 항목이 문서 내에 포함되었는지 여부에 따라 ESG 문서 내 항목에 대해 선별 상태를 지정하고, 선별 상태의 지정 결과에 따라 미리 저장된 복수의

ESG 관리 항목에 대한 최종 입력 필드 값의 입력 칸을 활성화하거나 또는 비활성화하며, 그리고 문서에 포함된 것으로 판단된, ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하도록 구성되고,

상기 정형 데이터 처리부는, 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하기 위해, 상기 ESG 관리 항목 데이터베이스에 저장된 각 ESG 관리 항목별 코드 값 및 데이터 세트 값 중 적어도 하나를 기준으로 파싱된 ESG 문서에 해당 항목을 포함하는지 여부를 판단하고,

상기 정형 데이터 처리부는, 스크리닝 결과에 이상이 없는지 여부를 판단하기 위해, 사용자의 확인 입력에 기초하거나 또는 ESG 문서 내에서 코드 값, 데이터 세트 값 중 적어도 하나에 매칭한 데이터에 대한 주변 데이터에 기초하여 스크리닝 결과의 이상 여부를 판단하는 것이며,

상기 정형 데이터 처리부는, 선별 상태를 지정하기 위해, 파싱된 ESG 문서 내에 해당 항목이 포함되어 있으면 상기 ESG 문서를 제1 선별 상태로 지정하고, 해당 항목이 포함되어 있지 않으면 상기 ESG 문서를 제2 선별 상태로 지정하는 것이며,

상기 정형 데이터 처리부는, 상기 정형화 데이터를 생성하기 위해, ESG 관리 항목에 대응한 항목의 내용을 획득하여 상기 정형화 데이터를 생성하기 위해, 제1 선별 상태로 지정된 ESG 관리 항목의 상기 ESG 문서 내 위치를 제공하는 단계; 제공된 위치에 기초하여 해당 ESG 관리 항목의 내용을 획득하는 단계; 및 선별된 ESG 관리 항목의 목록 및 획득된 각각의 ESG 관리 항목의 내용을 테이블 변환 처리하여 정형 데이터를 생성하고,

상기 정형 데이터 처리부는, 상기 해당 ESG 관리 항목의 내용을 획득하기 위해, 사용자의 내용 입력에 기초하거나 또는 제공된 위치에 대한 주변 데이터에 기초하여 '최종 필드 입력 값'의 내용을 획득하며,

상기 서비스 서버(200-1)는 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 것이고,

상기 기계학습 모델은 인코더 및 디코더를 포함하고, 입력 텍스트의 말뭉치와 잔여 텍스트의 말뭉치 간의 연관 관계를 추론하여 입력 텍스트의 말뭉치의 그 다음 순번으로 배치될 텍스트의 말뭉치를 예측하도록 학습된 것이고

상기 기계학습 모델은 복수의 트레이닝 샘플로 이루어진 트레이닝 데이터를 이용하여 학습되는 것으로서, 각 트레이닝 샘플은 동일한 샘플 텍스트의 말뭉치에 대응한 토큰 세트로부터 획득된 트레이닝 데이터, 레이블 데이터를 각각 포함하고, 상기 트레이닝 데이터는 학습을 위해 샘플 텍스트의 말뭉치로부터 획득된 토큰 세트에 포함된 토큰을 말뭉치 상의 순서대로 배열한 토큰 배열을 포함하고,

상기 인코더는 입력 토큰의 특징을 추출하도록 구성되고 - 상기 입력 토큰은 트레이닝 데이터에 포함된 토큰 배열이고,

상기 디코더는 입력 토큰의 특징에 기초하여 트레이닝 데이터의 토큰 배열 중에서 미리 설정된 순번의 텍스트 토큰으로부터 맨 끝의 텍스트 토큰을 포함한 잔여 배열 토큰을 복원하도록 구성되는 것이고,

상기 서비스 서버(200-1)는 학습부(210-1)를 포함하는 것으로서, 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하기 위해, 상기 서비스 서버(200-1)의 학습부(210-1)에 의해, 상기 트레이닝 데이터 세트 내 트레이닝 데이터를 상기 학습부(210-1)에 구현된 상기 기계학습 모델에 입력하여 해당 트레이닝 샘플에 대한 예측 값을 산출하고 - 상기 기계학습 모델 내 인코더는 상기 트레이닝 데이터로 입력된 입력 토큰의 특징을 추출하고, 상기 기계학습 모델 내 디코더는 상기 입력 토큰의 특징에 기초하여 복원된 잔여 배열 토큰의 복원 결과를 예측 값으로 출력함 - , 그리고 상기 서비스 서버(200-1)의 학습부(210-1)에 의해, 해당 트레이닝 샘플에 대해 산출된 예측 값과 상기 해당 트레이닝 샘플에 포함된 레이블 데이터의 값인 실제 값 간의 오차를 감소하도록 상기 기계학습 모델의 파라미터를 조정하는 것

을 특징으로 하는,

서비스 서버.

발명의 설명

기술분야

[0001] 본 출원의 실시 예들은 ESG 보조 툴을 이용하여 통일되지 않은 ESG 문서 내 다양한 비정형 포맷의 ESG 데이터를 정형화된 포맷의 ESG 데이터로 정리하고, 이 정형화된 ESG 데이터로 기계학습 모델을 학습하는 방법 및 기계학습 모델로 자동완성된 ESG 문서를 생성하는 서비스 서버에 관한 것이다.

배경기술

[0002] 최근 기업들이 리스크 관리에 보다 많은 신경을 기울이면서, 기업들을 다방면으로 평가하고 이와 같은 평가 결과를 투자, 합병, 생산 라인에 대한 리스크 관리 등에 응용하고 있다.

[0003] 일반적으로 기업들은 정량적으로 획득할 수 있는 재무적인 데이터들을 토대로 기업들을 평가하는 것이 일반적이었으나, 최근 들어서 비재무적인 데이터를 토대로 기업을 평가하고 리스크를 분석하는 방법론이 부각되고 있다. 기업들이 공개하는 재무적 데이터에는 해당 기업에게 불리한 내용이 반영되지 않으며, 기업이 제공하는 재무 관련 보고서에 대한 신뢰성 또한 의문시되는 실정이다. 비재무 데이터를 통한 기업 분석이 필요한 이유를 살펴보면, 사람들의 인터넷을 통한 SNS 활동이 활발해지면서, 기업 또는 제품에 대한 특정한 소문이 퍼지거나 하는 사건 등에 의해 기업 평판이 흔들리는 경우도 있으며 기업 오너의 범죄나 건강 관련 소문들에 의해 해당 기업이 가지는 리스크가 높아지는 경우도 있으나 이와 같은 사건들은 재무적 데이터를 통해 분석하기 어려운 것이 현실이다. 따라서, 재무적인 데이터뿐 아니라 비재무적인 데이터를 통해 기업을 분석함으로써 보다 정밀한 기업 평가가 가능해질 수 있다.

[0004] 이와 같은 흐름에서, 비재무 데이터들을 ESG(Environmnet, Social, Governance)의 세 주제로 나누어서 분석하는 방법론이 부각되고 있다.

[0005] 최근에는 기후변화 위기 등으로 인해 기업의 사회적 책임이 더욱 강조되어 ESG 평가 방식이 더욱 주목을 받고 있으며, 이는 범 글로벌 차원의 ESG 규제로 확인된다.

[0006] 이러한 급변하는 대외환경 변화에 따라 기업들은 다양한 이해관계자들의 공시 요구에 대응하여, ESG 평가 방식을 만족하기 위해 자신들의 ESG 경영활동에 대한 내용을 담은 데이터를 공시하고 있다. 이러한 데이터는 지속가능성 보고서(Sustainability report, SR)로 흔히 지칭되는, PDF 파일 형태의 문서로 발간되어 공시되고 있다.

[0007] 이러한 기업들의 ESG 데이터 공개는 해가 갈수록 증가하고 있으며, 이를 통해 축적되는 기업들의 ESG 데이터가 많아지고 있다. 향후 기업의 ESG 경영이 갈수록 중요해질 것이기 때문에, 이렇게 축적된 ESG 데이터를 빅데이터 관점에서 활용하고자 하는 수요가 점차 증가하고 있다.

[0008] 하지만 기업들이 공시 및 발간하는 지속가능성 보고서와 같은 ESG 문서는 대부분 PDF 파일 형태의 비정형 데이터이다. 이러한 ESG 문서 내 비정형 데이터를 빅데이터 관점에서 활용하기 위해서는, 특히, 최근 발전 중인 AI 기술에 ESG 문서를 활용하기 위해서는 정형화된 포맷의 데이터베이스로 정리 및 저장할 필요가 있다.

선행기술문헌

특허문헌

- [0009] (특허문헌 0001) 특허공개공보 제10-2019-0059449호 (2019.05.31.)
- (특허문헌 0002) 특허등록공보 제10-1942468호 (2019.02.08.)

발명의 내용

해결하려는 과제

[0010] 상술한 문제를 해결하기 위해, 본 출원의 실시 예들은 기업들의 정형화되지 않은 ESG 데이터를 정형화하여 저장 관리 가능한, ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 이루어진 ESG 데이터베이스를 구축하는 방법 및 이를 수행하는 ESG 서비스 제공 시스템을 제공하고자 한다.

[0011] 또한, 본 출원의 실시 예들은 비정형 ESG 데이터를 정형 데이터로 정확하게 변환하기 위한 인터페이스 화면을 제공 가능한 ESG 보조 툴을 이용하여 정형화된, ESG 데이터로 기계학습 모델을 학습하는 방법 및 기계학습 모

텔로 자동완성된 ESG 문서를 생성하는 서비스 서버를 제공하고자 한다.

과제의 해결 수단

- [0012] 본 출원의 일 측면에 따른 ESG 보조 툴을 이용하여 정형화된, ESG 데이터로 기계학습 모델을 학습하는 방법은 전자 장치 및 서비스 서버에 의해 수행될 수 있다. 상기 전자 장치는 복수의 ESG 관리 항목을 미리 저장한 ESG 관리 항목 데이터베이스와 연결된 것이다. 상기 방법은, 전자 장치(100)에 의해, 정형화 처리 대상의 ESG 문서를 파싱 처리하는 단계 - 상기 ESG 문서는 비정형 ESG 데이터를 포함함; 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하는 단계; 전자 장치(100)에 의해, 스크리닝 결과에 이상이 없는지 여부를 판단하는 단계; 전자 장치(100)에 의해, 상기 ESG 관리 항목에 대응한 항목이 문서 내에 포함되었는지 여부에 따라 ESG 문서 내 항목에 대해 선별 상태를 지정하는 단계; 전자 장치(100)에 의해, 선별 상태의 지정 결과에 따라 미리 저장된 복수의 ESG 관리 항목에 대한 최종 입력 필드 값의 입력 칸을 활성화하거나 또는 비활성화하는 단계; 문서에 포함된 것으로 판단된, ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하는 단계; 및 서비스 서버(200-1)에 의해, 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 단계;를 포함할 수 있다. 여기서, 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 단계;는, 상기 서비스 서버(200-1)의 학습부(210-1)에 의해, 상기 트레이닝 데이터 세트 내 트레이닝 데이터를 상기 학습부(210-1)에 구현된 상기 기계학습 모델에 입력하여 해당 트레이닝 샘플에 대한 예측 값을 산출하는 단계 - 상기 기계학습 모델 내 인코더는 상기 트레이닝 데이터로 입력된 입력 토큰의 특징을 추출하고, 상기 기계학습 모델 내 디코더는 상기 입력 토큰의 특징에 기초하여 복원된 잔여 배열 토큰의 복원 결과를 예측 값으로 출력함; 및 상기 서비스 서버(200-1)의 학습부(210-1)에 의해, 해당 트레이닝 샘플에 대해 산출된 예측 값과 상기 해당 트레이닝 샘플에 포함된 레이블 데이터의 값인 실제 값 간의 오차를 감소하도록 상기 기계학습 모델의 파라미터를 조정하는 단계;를 포함할 수 있다.
- [0013] 상기 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하는 단계는, 상기 ESG 관리 항목 데이터베이스에 저장된 각 ESG 관리 항목별 코드 값 및 데이터 세트 값 중 적어도 하나를 기준으로 파싱된 ESG 문서에 해당 항목을 포함하는지 여부를 판단하는 것이다.
- [0014] 상기 스크리닝 결과에 이상이 없는지 여부를 판단하는 단계는, 사용자의 확인 입력에 기초하거나 또는 ESG 문서 내에서 코드 값, 데이터 세트 값 중 적어도 하나에 매칭한 데이터에 대한 주변 데이터에 기초하여 스크리닝 결과의 이상 여부를 판단하는 것이다.
- [0015] 상기 선별 상태를 지정하는 단계는, 파싱된 ESG 문서 내에 해당 항목이 포함되어 있으면 상기 ESG 문서를 제1 선별 상태로 지정하고, 해당 항목이 포함되어 있지 않으면 상기 ESG 문서를 제2 선별 상태로 지정하는 것이다.
- [0016] 상기 정형화 데이터를 생성하는 단계는, ESG 관리 항목에 대응한 항목의 내용을 획득하여 상기 정형화 데이터를 생성하기 위해, 제1 선별 상태로 지정된 ESG 관리 항목의 상기 ESG 문서 내 위치를 제공하는 단계; 제공된 위치에 기초하여 해당 ESG 관리 항목의 내용을 획득하는 단계; 및 선별된 ESG 관리 항목의 목록 및 획득된 각각의 ESG 관리 항목의 내용을 테이블 변환 처리하여 정형 데이터를 생성하는 단계를 포함한다. 상기 해당 ESG 관리 항목의 내용을 획득하는 단계는, 사용자의 내용 입력에 기초하거나 또는 제공된 위치에 대한 주변 데이터에 기초하여 최종 필드 입력 값의 내용을 획득하는 것이다.
- [0017] 상기 기계학습 모델은 입력 텍스트의 말뭉치와 잔여 텍스트의 말뭉치 간의 연관 관계를 추론하여 입력 텍스트의 말뭉치의 그 다음 순번으로 배치될 텍스트의 말뭉치를 예측하도록 학습된, 인공 신경망 구조를 포함한 모델인 것이다.
- [0018] 일 실시 예에서, 상기 전자 장치는 다른 ESG 문서의 비정형 ESG 데이터를 정형화하여 저장한 ESG 문서 데이터베이스와 더 연결될 수 있다. 상기 ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하는 단계는, 기존 ESG 문서 내 데이터를 이용하겠다는 사용자 입력을 수신한 경우, 상기 다른 ESG 문서 내 데이터를 해당 항목에 대한 내용으로 획득하는 단계를 더 포함할 수 있다.
- [0019] 일 실시 예에서, 상기 전자 장치는 출력부를 더 포함할 수 있다. 상기 방법은, 이미 저장된 ESG 관리 항목을 관리하는 단계를 더 포함하고 - 상기 ESG 관리 항목의 관리는 수정, 삭제 또는 생성하는 것이다. 상기 ESG 관리 항목을 관리하는 단계는, 이미 저장된 ESG 관리 항목이 저장되었는지 확인하는 단계; 저장된 경우, ESG 관리 항목에 대한 수정, 삭제 또는 생성 요청을 수신하는 단계; 상기 수정, 삭제 또는 생성 요청을 수신할 경우, 저장

된 ESG 관리 항목 세트를 항목 레벨 구조로 나타낸 보조 툴 인터페이스 화면을 상기 출력부를 통해 표시하는 단계;를 포함할 수 있다.

[0020] 일 실시 예에서, 상기 ESG 관리 항목을 관리하는 단계는, 요청 대상 항목이 최하위 레벨 항목이 아닐 경우 대상 항목의 내용에 대한 입력 칸을 비활성화하는 단계; 상기 요청 대상 항목이 최하위 레벨 항목일 경우 대상 항목의 내용에 대한 입력 칸을 활성화하는 단계; 활성화된 입력 칸을 통해 수정 또는 생성할 대상 항목의 내용을 입력받는 단계; 및 입력된 항목의 내용을 ESG 관리 항목 데이터베이스에 저장하는 단계를 더 포함할 수 있다.

[0021] 일 실시 예에서, 상기 활성화된 입력 칸을 통해 대상 항목의 내용을 입력받는 단계는, 상기 대상 항목의 내용을 입력받기 위한 보조 툴 인터페이스 화면을 제공하는 단계를 포함할 수 있다. 상기 보조 툴 인터페이스 화면은 활성화된 입력 칸을 포함한 제1 서브 영역 및 항목 구조 레벨을 포함한 제2 서브 영역을 포함하고, 상기 제2 서브 영역에는 대분류별 항목 세트를 각각의 동일한 항목 레벨별로 분류하여 표시된다.

[0022] 본 출원의 다른 일 측면에 따른 컴퓨터 판독가능한 기록매체는 상술한 실시 예들에 따른 ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 기계학습 모델을 학습하는 방법을 수행하기 위한 프로그램을 기록할 수 있다.

[0023] 본 출원의 또 다른 일 측면에 따른 기계학습 모델로 자동완성된 ESG 문서를 생성하는 서비스 서버는, ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 이루어진 ESG 데이터베이스를 구축하기 위한 전자 장치와 연결된다. 상기 전자 장치는 복수의 ESG 관리 항목을 미리 저장한 ESG 관리 항목 데이터베이스와 연결된다.

[0024] 상기 전자 장치는, 비정형 ESG 데이터를 포함한 ESG 문서 내에서 상기 ESG 관리 항목 데이터베이스에 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응한 항목 및 해당 항목의 내용을 획득하여 상기 ESG 문서에 대한 정형화된 ESG 데이터를 생성하도록 구성된 정형 데이터 처리부; 및 ESG 문서의 비정형 데이터를 정형화된 데이터로 변경하기 위한 사용자 입력을 수신하는 보조 툴 인터페이스 화면을 출력부를 통해 제공하는 인터페이스 제공부를 포함한다. 상기 정형 데이터 처리부는, 정형화 처리 대상의 ESG 문서를 파싱 처리하고 - 상기 ESG 문서는 비정형 ESG 데이터를 포함함 -, 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하며, 스크리닝 결과에 이상이 없는지 여부를 판단하고, 상기 ESG 관리 항목에 대응한 항목이 문서 내에 포함되었는지 여부에 따라 ESG 문서 내 항목에 대해 선별 상태를 지정하고, 선별 상태의 지정 결과에 따라 미리 저장된 복수의 ESG 관리 항목에 대한 최종 입력 필드 값의 입력 칸을 활성화하거나 또는 비활성화하며, 그리고 문서에 포함된 것으로 판단된, ESG 관리 항목에 대응한 항목의 내용을 획득하여 정형화 데이터를 생성하도록 구성되고, 또한, 상기 정형 데이터 처리부는, 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 파싱된 ESG 문서가 포함하는지 스크리닝하기 위해, 상기 ESG 관리 항목 데이터베이스에 저장된 각 ESG 관리 항목별 코드 값 및 데이터 세트 값 중 적어도 하나를 기준으로 파싱된 ESG 문서에 해당 항목을 포함하는지 여부를 판단한다. 또한, 상기 정형 데이터 처리부는, 스크리닝 결과에 이상이 없는지 여부를 판단하기 위해, 사용자의 확인 입력에 기초하거나 또는 ESG 문서 내에서 코드 값, 데이터 세트 값 중 적어도 하나에 매칭한 데이터에 대한 주변 데이터에 기초하여 스크리닝 결과의 이상 여부를 판단하는 것이다. 상기 정형 데이터 처리부는, 선별 상태를 지정하기 위해, 파싱된 ESG 문서 내에 해당 항목이 포함되어 있으면 상기 ESG 문서를 제1 선별 상태로 지정하고, 해당 항목이 포함되어 있지 않으면 상기 ESG 문서를 제2 선별 상태로 지정하는 것이다. 상기 정형 데이터 처리부는, 상기 정형화 데이터를 생성하기 위해, ESG 관리 항목에 대응한 항목의 내용을 획득하여 상기 정형화 데이터를 생성하기 위해, 제1 선별 상태로 지정된 ESG 관리 항목의 상기 ESG 문서 내 위치를 제공하는 단계; 제공된 위치에 기초하여 해당 ESG 관리 항목의 내용을 획득하는 단계; 및 선별된 ESG 관리 항목의 목록 및 획득된 각각의 ESG 관리 항목의 내용을 테이블 변환 처리하여 정형 데이터를 생성한다. 상기 정형 데이터 처리부는, 상기 해당 ESG 관리 항목의 내용을 획득하기 위해, 사용자의 내용 입력에 기초하거나 또는 제공된 위치에 대한 주변 데이터에 기초하여 '최종 필드 입력 값'의 내용을 획득한다.

[0025] 상기 서비스 서버는 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하도록 구성될 수 있다.

발명의 효과

[0026] 본 발명의 일 측면에 따른 ESG 서비스 제공 시스템은 고유한 보조 툴 인터페이스 화면을 제공하는 ESG 보조 툴을 이용하여 기존 수작업의 비효율을 줄이고, 완전 자동화가 어려운 제약조건의 문제점을 극복할 수 있다. 상기 보조 툴 인터페이스 화면은 ESG 문서 내 데이터의 분류, 저장 및 관리를 효율적으로 수행하는 효과를 제공한다.

[0027] 그 결과, 상기 ESG 서비스 제공 시스템은 기업의 지속가능성 보고서로부터 ESG 데이터를 수집 변환하여 저장 관리하고 이에 기초한 기업의 ESG 경영현황을 파악할 수 있으며, 특히 비정형 ESG 데이터가 상기 보조 툴의 프로

세스를 따라 과싱 및 정형화되어 데이터베이스로 구축됨으로써, 빅데이터 관점에서 분석 및 가공 처리에 따라 적절하게 ESG 데이터를 활용할 수 있다.

[0028] 또한, 상기 ESG 서비스 제공 시스템은 정형화된 ESG 데이터를 학습 데이터로 제공해, ESG 문서를 자동완성할 수 있는 기계학습 모델을 학습할 수 있다.

[0029] 본 발명의 효과들은 이상에서 언급한 효과들로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 청구범위의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0030] 본 발명 또는 종래 기술의 실시 예의 기술적 해결책을 보다 명확하게 설명하기 위해, 실시 예에 대한 설명에서 필요한 도면이 아래에서 간단히 소개된다. 아래의 도면들은 본 명세서의 실시 예를 설명하기 목적일 뿐 한정된 목적이 아니라는 것으로 이해되어야 한다. 또한, 설명의 명료성을 위해 아래의 도면들에서 과장, 생략 등 다양한 변형이 적용된 일부 요소들이 도시될 수 있다.

도 1은, 본 출원의 일 측면에 따른, ESG 서비스 제공 시스템의 개략도이다.

도 2a 및 도 2b는, 본 출원의 다른 일 측면에 따른, ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 이루어진 ESG 데이터베이스를 구축하는 방법의 흐름도이다.

도 3은, 본 출원의 일 실시 예에 따른, 항목의 최종 입력 필드 값의 입력 칸을 포함한 보조 툴 인터페이스 화면을 도시한다.

도 4는, 본 출원의 일 실시 예에 따른, ESG 관리 항목을 관리하는 과정의 흐름도이다.

도 5는, 본 출원의 일 실시 예에 따른, 항목 레벨 구조를 표시한 보조 툴 인터페이스 화면을 도시한다.

도 6은, 본 출원의 일 측면에 따른, ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템의 개략도이다.

도 7은, 본 출원의 일 실시 예에 따른, ESG 자동완성 모델을 생성하는 과정의 개략도이다.

도 8은, 본 출원의 일 실시 예에 따른, ESG 자동완성 모델에서 출력 텍스트를 예측하는 동작의 개략도이다.

도 9는, 본 출원의 다른 일 측면에 따른, ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하는 방법의 흐름도이다.

도 10은, 본 출원의 일 실시 예에 따른, 추천 텍스트의 말뭉치의 출력 결과를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0031] 이하에서, 도면을 참조하여 본 출원의 실시 예들에 대하여 상세히 살펴본다.

[0032] 그러나, 이는 본 개시(disclosure)를 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 개시의 실시 예의 다양한 변경(modification), 균등물(equivalent), 및/또는 대체물(alternative)을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다.

[0033] 본 명세서에서, "가진다," "가질 수 있다," "포함한다," 또는 "포함할 수 있다" 등의 표현은 해당 특징(예: 수치, 기능, 동작, 단계, 부품, 요소 및/또는 성분 등의 구성요소)의 존재를 가리키며, 추가적인 특징의 존재나 부가를 제외시키는 것이 아니다.

[0034] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다.

[0035] 다양한 실시 예에서 사용된 "제 1", "제 2", "첫째" 또는 "둘째" 등의 표현들은 다양한 구성요소들을, 순서 및/또는 중요도에 상관없이 수식할 수 있고, 해당 구성요소들을 한정하지 않는다. 상기 표현들은 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 수 있다. 예를 들면, 제1 구성요소와 제2 구성요소는, 순서 또는 중요도와 무관하게, 서로 다른 구성요소를 나타낼 수 있다.

- [0036] 본 명세서에서 사용된 표현 "~하도록 구성된(또는 설정된)(configured to)"은 상황에 따라, 예를 들면, "~에 적합한(suitable for)," "~하는 능력을 가지는(having the capacity to)," "~하도록 설계된(designed to)," "~하도록 변경된(adapted to)," "~하도록 만들어진(made to)," 또는 "~를 할 수 있는(capable of)"과 바꾸어 사용될 수 있다. 용어 "~하도록 구성(또는 설정)된"은 하드웨어적으로 "특별히 설계된(specifically designed to)"것만을 반드시 의미하지 않을 수 있다. 대신, 어떤 상황에서는, "~하도록 구성된 장치"라는 표현은, 그 장치가 다른 장치 또는 부품들과 함께 "~할 수 있는" 것을 의미할 수 있다. 예를 들면, 문구 "A, B, 및 C를 수행하도록 구성(또는 설정)된 프로세서"는 해당 동작을 수행하기 위한 전용 프로세서(예: 임베디드 프로세서), 또는 메모리 장치에 저장된 하나 이상의 소프트웨어 프로그램들을 실행함으로써, 해당 동작들을 수행할 수 있는 범용 프로세서(generic-purpose processor)(예: CPU 또는 application processor)를 의미할 수 있다.
- [0038] 도 1은, 본 출원의 일 측면에 따른, ESG 서비스 제공 시스템의 개략도이다.
- [0039] 실시예들에 따른 ESG 서비스 제공 시스템(1)은 전적으로 하드웨어이거나, 전적으로 소프트웨어이거나, 또는 부분적으로 하드웨어이고 부분적으로 소프트웨어인 측면을 가질 수 있다. 예컨대, 시스템은 데이터 처리 능력이 구비된 하드웨어 및 이를 구동시키기 위한 운용 소프트웨어를 통칭할 수 있다. 본 명세서에서 "부(unit)", "시스템" 및 "장치" 등의 용어는 하드웨어 및 해당 하드웨어에 의해 구동되는 소프트웨어의 조합을 지칭하는 것으로 의도된다. 예를 들어, 하드웨어는 CPU(Central Processing Unit), GPU(Graphic Processing Unit) 또는 다른 프로세서(processor)를 포함하는 데이터 처리 기기일 수 있다. 또한, 소프트웨어는 실행중인 프로세스, 객체(object), 실행파일(executable), 실행 스레드(thread of execution), 프로그램(program) 등을 지칭할 수 있다.
- [0040] 도 1을 참조하면, 상기 ESG 서비스 제공 시스템(1)이 제공하는 ESG 서비스는 ESG 문서 내 비정형 형식의 ESG 데이터를 정형 형식으로 변경하는 정형화 보조 툴 제공 서비스를 포함한다. 상기 ESG 서비스 제공 시스템(1)은 이 서비스를 제공하기 위해, 데이터베이스 서버(200)와 연결된 전자 장치(100)를 포함한다.
- [0041] 상기 데이터베이스 서버(200)는 통신부(210) 및 하나 이상의 저장소를 포함한다. 특정 실시 예들에서, 상기 데이터베이스 서버(200)는 ESG 문서 DB(220), ESG 관리 항목 DB(230) 및 파싱 텍스트 DB(240) 중 하나 이상의 저장소를 포함할 수도 있다.
- [0042] 통신부(210)는 전자 장치(100)와 같은 외부 장치들과 데이터를 유/무선의 전기 통신으로 송수신하도록 구성된다. 상기 데이터베이스 서버(200)는 통신부(210)를 통해 하나 이상의 저장소에 데이터를 저장하거나 저장된 데이터를 외부 장치로 제공할 수도 있다.
- [0043] 상기 통신부(210)는 송신기(transmitter), 수신기(receiver), 송수신기(transceiver), 통신부(communication unit), 통신 모델(communication model) 또는 통신 회로(communication circuit)로 지칭할 수 있다. 통신부(210)는 유선 접속 시스템 및 무선 접속 시스템들인 IEEE(institute of electrical and electronics engineers) 802.xx 시스템, IEEE Wi-Fi 시스템, 3GPP(3rd generation partnership project) 시스템, 3GPP LTE(long term evolution) 시스템, 3GPP 5G NR(new radio) 시스템, 3GPP2 시스템, 블루투스(Bluetooth) 등 다양한 무선 통신 규격 중 적어도 하나를 지원할 수 있다.
- [0044] ESG 문서 DB(220)는 아래에서 서술할, 정형 데이터 처리부(140)에 의해 획득된, ESG 문서를 정형화 처리한 데이터를 저장한다. 일부 실시 예들에서, ESG 문서 DB(220)는 정형 데이터 처리부(140)에 의해 테이블 변환 처리된 각각의 ESG 문서별 항목 및 그 값을 저장할 수도 있다.
- [0045] ESG 문서 DB(220)에 저장되는 테이블 변환 처리 결과 데이터는 아래에서 정형 데이터 처리부(140)를 참조해 보다 상세히 서술한다.
- [0046] ESG 관리 항목 DB(230)는 비정형 ESG 데이터를 정형화 데이터로 변환하는데 요구되는 관리 항목을 저장한다.
- [0047] 일부 실시 예들에서, ESG 관리 항목 DB(230)는 하나 이상의 대분류에 대한 항목 세트를 저장할 수도 있다. 상기 대분류는 보고서의 작성 기준, ESG 주제 폴로 이루어진다. 일부 실시 예들에서, 상기 대분류는 지속가능성 보고서의 공시 기준인 GRI(Global Reporting Initiative) 또는 미국 지속가능회계기준위원회, SASB(Sustainability Accounting Standards Board), TCFD(Task Force on Climate-Related Financial Disclosures), UN SDGs(UN Sustainable Development Goals), 산업통상부의 ESG 가이드라인(또는 K-ESG로 지칭됨)가 제공하는 산업별로 구분한 중요성 주제, 기준을 포함할 수도 있다.
- [0048] 예를 들어, 상기 ESG 관리 항목 DB(230)는 GRI, SASB, TCFD, UN-SDGs, K-ESG, 및/또는 기타 기준으로 이루어진

대분류를 저장할 수도 있다.

- [0049] 각 대분류에 대한 항목 세트는 해당 대분류의 기준, 주제에서 작성을 요구하는 ESG 키워드, ESG 주제 항목을 포함할 수도 있다. 각 대분류에 대한 항목 세트는 각각의 대분류가 공통으로 포함하는 주제 항목(공통 주제 항목)을 포함할 수도 있다. 또한, 각 대분류에 대한 항목 세트는 개별 대분류에 고유한 복수의 다른 항목을 포함할 수도 있다. 예를 들어, GRI 대분류의 항목 세트는 GRI200, GRI300, GRI400의 문서 항목을 포함할 수도 있다.
- [0050] 상기 ESG 관리 항목DB(230)는 항목 세트 내 각각의 항목을 해당 코드 값 및/또는 데이터 세트 값으로 저장할 수도 있다.
- [0051] 일부 실시 예들에서, 상기 대분류에 대한 항목 세트는 평가 대상 기업의 산업에 기초하여 더 분류될 수도 있다. 동일 대분류가 적용되어도 평가 대상 기업의 산업 군에 따라 항목 세트가 상이할 수도 있다. 예를 들어, SASB는 헬스케어 산업에 대해서는 '정보 보안'을 중요 주제로 선정된 반면 전력 산업에 대해서는 중요 주제로 선정하지 않았다. 따라서, SASB가 헬스케어 산업에 속하는 평가 대상 기업에 적용될 경우 '정보 보안' 키워드, 관련 항목은 대응한 비정형 ESG 데이터를 추출하는데 이용되고, SASB가 전력 산업에 속하는 평가 대상 기업에 적용될 경우 '정보 보안' 키워드, 관련 항목은 대응한 비정형 ESG 데이터를 추출하는데 이용되지 않을 수도 있다.
- [0052] 일부 실시 예들에서, 상기 ESG 문서 내 복수의 항목은 계층 구조로 지정될 수도 있다. 각각의 대분류의 항목 세트 내 각각의 항목은 최상위 레벨부터 최하위 레벨별로 추가 분류될 수도 있다. 예를 들어, GRI 대분류의 항목 세트에서 최상위 레벨(레벨 1)은 공통 주제, GRI200, GRI300, GRI400을 포함할 수도 있다. 공통 주제의 레벨 1에서 다음 레벨(레벨 2)은 조직프로필, 전략, 거버넌스를 포함할 수도 있다. 조직프로필의 레벨 2에서 다음 레벨(레벨 3)은 조직의 규모, 임직원을 포함할 수도 있다. 조직의 규모의 레벨 3에서 다음 레벨(레벨 4)은 국내 임직원 수, 해외 임직원 수, 사업장 수, 순매출액을 포함할 수도 있다.
- [0053] 상기 ESG 문서 항목 및 이를 저장하는 ESG 관리 항목 DB(230)에 대해서는 아래의 도 2a에서 보다 상세히 서술한다.
- [0054] 파싱 텍스트 DB(240)는 파싱된 ESG 문서의 텍스트를 저장한다. 일부 실시 예들에서, 상기 파싱된 ESG 문서의 텍스트는 아래에서 서술할, 정형 데이터 처리부(140)로부터 획득될 수도 있다.
- [0055] 상기 파싱된 ESG 문서의 텍스트에 대해서는 정형 데이터 처리부(140)를 참조해 아래에서 보다 상세히 서술한다.
- [0057] 상기 전자 장치(100)는 예를 들어, 랩탑 컴퓨터, 기타 컴퓨팅 장치, 태블릿, 셀룰러 폰, 스마트 폰, 스마트 워치, 스마트 글래스, 헤드 마운트 디스플레이(HMD), 기타 모바일 장치, 기타 웨어러블 장치일 수도 있다.
- [0058] 상기 전자 장치(100)는 비정형 ESG 데이터를 정형화하는 보조 툴을 포함한다. 상기 전자 장치(100)는 획득한 ESG 문서에 포함된 비정형 ESG 데이터를 추출하고, 상기 보조 툴을 이용하여 추출된 비정형 ESG 데이터를 정형화하도록 구성된다. 또한, 상기 전자 장치(100)는 정형화된 ESG 데이터를 데이터베이스 서버(200)로 공급하여 데이터베이스를 구축하는 것을 지원하도록 구성된다.
- [0059] 상기 전자 장치(100)는 데이터베이스 서버(200)와 통신하는 단말 장치로서, 데이터를 처리할 수 있는 적어도 하나의 프로세서, 데이터를 저장하는 메모리, 데이터를 송/수신하는 통신부를 포함한다.
- [0060] 특정 실시 예들에서 상기 전자 장치(100)는 도 1에 도시된 것처럼, 통신부(110), 입력부(120), 인터페이스 제공부(130), 정형 데이터 처리부(140) 및 출력부(150)를 포함할 수도 있다.
- [0061] 상기 통신부(110)는 데이터베이스 서버(200)와 같은 외부 장치들과 데이터를 유/무선의 전기 통신으로 송수신하도록 구성된다. 상기 전자 장치(100)는 통신부(110)를 통해 입력부(120), 인터페이스 제공부(130), 정형 데이터 처리부(140) 및 출력부(150)에 데이터를 제공하거나 처리 결과를 외부 장치로 제공할 수도 있다.
- [0062] 상기 통신부(110)는 통신부(210)와 유사하므로, 자세한 설명은 생략한다. 상기 통신부(110)에 대한 설명은 상기 통신부(210)와의 차이점을 위주로 서술한다.
- [0063] 통신부(110)는 비정형 ESG 데이터를 포함한 ESG 문서를 수신하도록 구성된다. 사용자는 통신부(110)를 통해 복수의 ESG 문서를 웹 페이지에서 수집하고 상기 입력부(120)에 업로드할 수도 있다.
- [0064] 상기 ESG 문서는 정부 웹 페이지, 기업 웹 페이지, 기타 웹 사이트 등에서 이미 공개되어 있는 국내외 평가 대상 기업의 ESG 지속가능 보고서, 또는 개별 기업의 ESG 담당자가 작성하여 배포한 기타 ESG 문서를 포함한다. 상기 평가 대상 기업은 사업체로서의 기업만을 지칭하지 않고 기관, 비영리 단체 등의 다양한 형태의 단체를 포

괄하여 지칭할 수 있다.

- [0065] 상기 ESG 문서는 해당 기업의 홈페이지, 정부 포털 등에서 공개된 DB로 수집할 수도 있다.
- [0066] 일부 실시 예들에서, 상기 ESG 문서는 정부 문서 DB(310), 고객 문서 DB(320), 및 웹 사이트 문서 DB(330) 중 하나 이상의 외부 DB로부터 획득될 수도 있다.
- [0067] 상기 정부 문서 DB(310)는 정부 데이터 포털에 액세스하여 검색된 적어도 하나의 기업에 대한 ESG 문서를 수집할 수도 있다.
- [0068] 상기 고객 문서 DB(320)는 상기 시스템(1)과 관련된 고객, 예를 들어, 상기 시스템(1)에서 제공하는 서비스에 가입한 고객 기업의 데이터 포털에 액세스하여 해당 고객에 대한 ESG 문서를 수집할 수도 있다. 단일 고객에 대해 다수의 ESG 문서가 수집될 수도 있다.
- [0069] 상기 웹 사이트 문서 DB(330)는 주요 웹 사이트로 미리 지정된 하나 이상의 웹 사이트 각각에 액세스하고 액세스한 웹 사이트에서 공개된 ESG 문서를 검색하여 수집할 수도 있다.
- [0070] 상기 ESG 데이터는 문서를 생성한 평가 대상 기업의 ESG 현황을 서술하는 정보를 나타낸다. 상기 ESG 문서는 Portable Document Format (PDF) 문서일 수도 있으나, 이에 제한되진 않는다.
- [0071] 상기 비정형 ESG 데이터는 비정형 포맷으로 ESG 정보를 나타낸 데이터이다. 비정형 데이터는 데이터베이스 서버(200)에서 인식 가능한 구조를 가지지 않는 데이터이다. 상기 비정형 ESG 데이터는 구조화되지 않은 텍스트, 이미지를 포함할 수도 있다.
- [0073] *
- [0074] 입력부(120), 인터페이스 제공부(130), 정형 데이터 처리부(140)는 적어도 하나의 프로세서로 구현될 수도 있다.
- [0075] 입력부(120)는 통신부(110)로부터 입력된 비정형 ESG 데이터를 포함한 ESG 문서를 정형 데이터 처리부(140)로 공급하도록 구성된다.
- [0076] 인터페이스 제공부(130)는 ESG 문서의 비정형 데이터를 정형 데이터로 변경하기 위한 사용자의 입력을 수신하도록 구성된다. 상기 인터페이스 제공부(130)는 사용자가 ESG 항목, 코드 값, 데이터 세트 값, 답변 내용 및/또는 최종 입력 필드 값의 내용을 수정하도록 구성된다.
- [0077] 일부 실시 예들에서, 상기 인터페이스 제공부(130)는 데이터베이스 서버(200)의 ESG 관리 항목 DB(230) 내 코드 값, 데이터 세트 값을 기준으로 문서 내에서 사용자 입력을 적용할 항목의 위치를 검색하고, 해당 위치 또는 이를 포함한 페이지로 화면을 이동하도록 출력장치를 제어하도록 구성될 수도 있다.
- [0078] 일부 실시 예들에서, 상기 인터페이스 제공부(130)는 선별 처리 결과에 이상이 있을 경우 사용자가 직접 항목을 수정하도록 보조 툴 인터페이스 화면을 제공할 수도 있다. 상기 보조 툴 인터페이스 화면은 항목 등에 대한 수정 입력을 수신하도록 구성된다.
- [0079] 인터페이스 제공부(130)의 동작에 대해서는 아래의 도 2를 참조해 보다 상세히 서술한다.
- [0080] 상기 정형 데이터 처리부(140)는 ESG 문서 내에 데이터베이스 서버(200)의 ESG 관리 항목 DB(230)에 미리 저장된 복수의 ESG 관리 항목에 대응한 항목 및 상기 항목의 내용을 획득하여 상기 ESG 문서에 대한 정형화된 ESG 데이터를 생성하도록 구성된다.
- [0081] 상기 정형 데이터 처리부(140)는 상기 ESG 문서(예를 들어, 지속 가능성 보고서의 PDF 데이터)를 파싱 처리하도록 구성된다.
- [0082] 또한, 상기 정형 데이터 처리부(140)는 ESG 관리 항목 DB(230)에 미리 저장된 복수의 ESG 관리 항목을 기준으로 문서 내 대응한 항목을 검색하고, 보조 툴을 이용하여 해당 항목의 내용을 획득하도록 구성된다.
- [0083] 이러한 정형 데이터 처리부(140)의 동작에 대해서는 아래의 도 2를 참조해 보다 상세히 서술한다.
- [0084] 출력부(150)는 디스플레이 또는 스피커 중에서 하나 이상을 포함할 수 있다. 출력부(150)는 프로세서, 또는 인터페이스 제공부(130), 정형 데이터 처리부(140)의 제어에 따라서 콘텐츠를 출력할 수 있다. 디스플레이는 전자 장치(110)에서 처리되는 정보를 화상으로 표시할 수 있다. 예컨대, 디스플레이는 전자 장치(100)에서 구동되는 응용 프로그램의 실행화면 정보, 또는 이러한 실행화면 정보에 따른 UI(User Interface), GUI(Graphic User

Interface) 정보를 표시할 수 있다. 상기 디스플레이는 예를 들어, LCD, OLED, 플렉서블 디스플레이, 기타 디스플레이 구성요소 등을 포함할 수도 있다.

- [0085] 많은 실시 예들에서, 디스플레이는 터치 센서와 상호 레이어 구조를 이루거나 일체형으로 형성된 터치 패널로 구현되어, 입력 인터페이스 화면과 출력 인터페이스 화면을 동시에 제공할 수도 있다. 터치 입력은 (예컨대, 사용자의 신체, 또는 도구 등을 포함한) 포인팅 객체에 의해 입력된다.
- [0086] 출력부(150)는 인터페이스 제공부(130)에서 제공하는 보조 툴 인터페이스 화면을 표시하도록 구성된다. 상기 출력부(150)는 항목 등에 대한 정보를 수정하는 입력칸을 표시할 수도 있다.
- [0087] 또한, 상기 출력부(150)는 ESG 이미지를 출력할 수도 있다. 또한, 상기 출력부(150)는 파싱 텍스트를 출력할 수도 있다. 일부 실시 예들에서, 상기 출력부(150)는 출력부(150)의 화면 일 부분(예컨대, 일 측면)에 ESG 문서의 이미지를 배치해 출력하고, 다른 일부분 (예컨대, 상단)에 파싱 텍스트를 배치해 출력할 수도 있다.
- [0088] 상기 전자 장치(100), 데이터베이스 서버(200) 또는 시스템(1)이 다른 구성요소를 포함할 수도 있다는 것이 통상의 기술자에게 명백할 것이다. 예를 들어, 상기 데이터베이스 서버(200)는 데이터 엔트리를 위한 입력 장치, 및 인쇄 또는 다른 데이터 표시를 위한 출력 장치를 포함하는, 본 명세서에 서술된 동작에 필요한 다른 하드웨어 요소를 포함할 수도 있다. 또한, 상기 시스템(1)은 데이터베이스 서버(200)와 외부 장치(예를 들어, 사용자 단말, 또는 외부 데이터베이스 등) 사이를 연결하는 네트워크, 네트워크 인터페이스 및 프로토콜 등을 더 포함할 수 있다.
- [0090] 도 2a 및 도 2b는, 본 출원의 다른 일 측면에 다른, ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 이루어진 ESG 데이터베이스를 구축하는 방법의 흐름도이다.
- [0091] 도 2a 및 도 2b의 방법은 도 1의 ESG 서비스 제공 시스템과 같은 하나 이상의 컴퓨팅 장치에 의해 수행될 수도 있다.
- [0092] 도 2a를 참조하면, 상기 방법은, (예컨대, 통신부(110)에 의해) ESG 문서를 수신하는 단계(S100);를 포함한다.
- [0093] 단계(S100)에서 복수의 ESG 문서가 수집될 수도 있다.
- [0094] 또한, 상기 방법은, (예컨대, 정형 데이터 처리부(140)에 의해) 단계(S100)에서 수신한 ESG 문서를 파싱 처리하고, 파싱된 ESG 문서에 포함되어 있는, 미리 저장된 관리 항목에 해당하는 항목의 데이터를 획득하여 정형화된 ESG 데이터를 생성하는 단계(S200)를 포함할 수도 있다.
- [0095] 상기 단계(S200)는, 단계(S100)에서 수신한 ESG 문서를 상기 정형 데이터 처리부(140)로 업로드 하는 단계(S210); 및 상기 ESG 문서를 파싱 처리하는 단계(S220)를 포함한다.
- [0096] 상기 정형 데이터 처리부(140)는 상기 ESG 문서(예를 들어, 지속 가능성 보고서의 PDF 데이터)를 파싱 처리도록 구성된다. 일부 실시 예들에서, 상기 정형 데이터 처리부(140)는 파싱 툴(parsing tool)을 포함할 수도 있다. 상기 정형 데이터 처리부(140)는 파싱 툴을 이용하여 비정형 데이터를 파싱 처리할 수도 있다. 상기 파싱 툴은 PDFMiner일 수도 있으나, 이에 제한되진 않는다.
- [0097] 예를 들어, 상기 정형 데이터 처리부(140)는 상기 파싱 툴을 PDF 포맷의 ESG 문서에 적용하여 파싱된 PDF 데이터를 생성할 수도 있다. 상기 파싱 툴은 PDF 포맷의 ESG 문서로부터 하나 이상의 텍스트 줄을 포함한 텍스트 블록, 이미지, 및/또는 기하학 형상(geometric shapes)을 추출할 수도 있다.
- [0098] 일부 실시 예들에서, 추출된 텍스트 블록은 텍스트, 바운딩 박스, 텍스트 라인의 속성 중 적어도 하나의 속성을 가질 수도 있다. 예를 들어, 파싱된 PDF 데이터는 추출된 텍스트 블록을 포함하며, 각 텍스트 블록은 하나 이상의 텍스트 라인을 포함할 수도 있다.
- [0099] 일부 실시 예들에서, 각각의 텍스트 라인은 해당 라인의 텍스트, 및 라인을 둘러싼 바운딩 박스의 속성 중 적어도 하나의 속성을 가질 수도 있다. 상기 텍스트 블록 또는 텍스트 라인의 바운딩 박스는 각각의 텍스트 블록 또는 텍스트 라인이 일반적으로 차지하는 페이지의 영역을 정의하는, 텍스트 블록 주위의 박스이다. 상기 바운딩 박스는 직사각형 형상의 박스를 지칭할 수도 있다.
- [0100] 일부 실시 예들에서, 상기 비정형 ESG 데이터가 이미지 데이터를 포함할 경우, 상기 정형 데이터 처리부(140)는 파싱 처리 이전에 ESG 문서의 이미지 데이터를 OCR(Optical Character Recognition) 처리하여 비정형 텍스트를 인식하도록 더 구성될 수도 있다. 일부 실시 예들에서, 상기 파싱 툴은 OCR 처리 가능하도록 구성될 수도 있다.

그러면, 파싱 툴을 이용하여 OCR 처리 및 파싱 처리를 동시에 수행함으로써, 텍스트 박스 내 텍스트를 인식하고 텍스트가 인식된 PDF 페이지 내 구조(예컨대, 텍스트 속성)를 분석할 수도 있다.

- [0101] 파싱 툴에 의해 추출된 이미지는 각각의 바운딩 박스와 연관될 수도 있다.
- [0102] 파싱 툴에 의해 추출되는 기하학적 형상은, 예를 들어 테이블, 벡터 그림에 나타난 선, 곡선, 직사각형, 기타 도형 객체를 포함할 수도 있다. 파싱 툴은 PDF 페이지의 이미지, 테이블 주변의 바운딩 박스를 찾아 추출할 수도 있다.
- [0103] 이와 같이 정형 데이터 처리부(140)의 파싱 결과는 텍스트 블록, 텍스트 라인, 이미지, 및/또는 기하학적 형상이 차지하는 ESG PDF 문서의 페이지의 영역, 부분을 정의하는 다양한 바운딩 박스를 포함할 수도 있다.
- [0104] 일부 실시 예들에서, 상기 정형 데이터 처리부(140)는 파싱 결과에 포함된 파싱된 ESG 문서의 텍스트를 데이터 베이스 서버(200)로 전송하여 파싱 텍스트 DB(240)에 저장할 수도 있다.
- [0105] 또한, 상기 방법은, 미리 저장된 복수의 ESG 관리 항목 중 적어도 하나에 대응하는 항목을 단계(S220)에서 파싱된 ESG 문서가 포함하는지 스크리닝하는 단계(S230)를 포함한다.
- [0106] 일부 실시 예들에서, 상기 단계(S230)는, ESG 관리 항목 DB(230)에 저장된 각 ESG 관리 항목별 코드 값 및/또는 데이터 세트 값을 기준으로 파싱된 ESG 문서에 해당 항목을 포함하는지 여부를 판단하는 것일 수도 있다.
- [0107] 상기 일부 실시 예들에서, 정형 데이터 처리부(140)는 ESG 관리 항목 DB(230)에 저장된 코드 값 및/또는 데이터 세트 값을 기준으로 파싱된 ESG 문서에 해당 항목을 포함하는지 여부를 판단하는, 스크리닝 동작을 수행하도록 구성된다.
- [0108] 상기 ESG 관리 항목 DB(230)는 복수의 ESG 관리 항목을 저장한다. 각각의 ESG 관리 항목은 적어도 하나의 코드 값 및/또는 적어도 하나의 데이터 세트 값을 포함할 수도 있다. 그러면, 상기 ESG 관리 항목 DB(230)에 저장된 ESG 문서 항목을 구성하는 코드 값, 데이터 세트 값은 파싱된 ESG 문서가 저장된 해당 항목을 포함하고 있는지를 판단하기 위한 기준으로 이용될 수도 있다.
- [0109] 상기 정형 데이터 처리부(140)는 코드 값, 데이터 세트 값이 파싱된 ESG 문서에서 검색된 경우, 파싱된 ESG 문서가 ESG 관리 항목 DB(230)에 저장된 ESG 문서 항목 중 적어도 하나를 갖고 있는 것으로 판단한다. 검색된 코드 값, 데이터 세트 값에 대응한 항목이 상기 파싱된 ESG 문서에 포함된 것으로 판단된다.
- [0110] 그러면, 단계(S230)에서 정형 데이터 처리부(140)는 파싱된 ESG 문서가 포함하는 것으로 추정되는, ESG 관리 항목에 대응한 항목의 목록을 획득할 수도 있다.
- [0111] 일부 실시 예들에서, 상기 방법은, 스크리닝 결과에 이상이 없는지 여부를 판단하는 단계(S240)를 더 포함할 수도 있다. 상기 단계(S240)는 아래에서 서술할 단계(S250) 이전에 수행될 수도 있다.
- [0112] 여기서 이상 스크리닝 결과는 ESG 문서 내 ESG 관리 항목이 아닌데도 ESG 관리 항목으로 잘못 선별된 것을 나타낸다.
- [0113] ESG 문서 내에 미리 저장된 ESG 관리 항목에 대응한 코드 값 및/또는 데이터 세트 값에 매칭되는 코드 값 및/또는 데이터 세트 값이 검색되어도, 검색된 값이 실제로는 문서 내 ESG 관리 항목과 관련이 없는 데이터일 수도 있다. 예를 들어, GRI 기준에 따른 ESG 관리 항목은 102-1과 같이 '3자리 숫자', '불임표(hyphen, -)', '1 또는 2자리 숫자'가 순차적으로 배열된 구조를 가진다. ESG 문서 내에 GRI 기준에 따른 ESG 관리 항목이 아닌, 별도의 동일 구조의 데이터가 우연히 포함되어 있을 수가 있다. 이 경우, ESG 관리 항목과 관련이 없는 데이터가 코드 값 및/또는 데이터 세트 값에 매칭되는 결과로 검색될 수도 있고, 그러면 정형 데이터 처리부(140)는 ESG 문서가 ESG 관리 항목을 포함한 스크리닝 결과가 산출될 수도 있다.
- [0114] 일부 실시 예들에서, 상기 단계(S240)는, 사용자의 확인 입력 및/또는 ESG 문서 내에서 코드 값 및/또는 데이터 세트 값에 매칭한 데이터에 대한 주변 데이터에 기초하여 스크리닝 결과의 이상 여부를 판단할 수도 있다.
- [0115] 정형 데이터 처리부(140)는 미리 저장된 ESG 관리 항목에 대응한 코드 값 및/또는 데이터 세트 값에 매칭한 데이터의 위치를 제공하고, 제공된 위치에 대해 확인 입력이 수신되면 이상이 없는 것으로 판단할 수도 있다(S240). 정형 데이터 처리부(140)가 매칭한 데이터의 위치를 인터페이스 제공부(130)로 공급하면, 인터페이스 제공부(130)는 매칭한 데이터의 위치를 포함한 보조 툴 인터페이스 화면을 표시하여, 각 위치의 데이터가 ESG 관리 항목인지를 확인하는 사용자 입력을 유도할 수도 있다(S240).

- [0116] 또한, 정형 데이터 처리부(140)는 미리 저장된 ESG 관리 항목에 대응한 코드 값 및/또는 데이터 세트 값에 매칭한 데이터의 위치를 기준으로 일정 범위 이내에 배치된, 주변 데이터에 기초하여 매칭한 데이터가 실제로 ESG 관리 항목에 대응하는 지를 판단할 수도 있다(S240).
- [0117] 또한, 상기 단계(S240)는, 이상이 있을 경우 ESG 문서에 포함된 것으로 선별된 결과 중 적어도 일부를 수정하는 단계를 포함할 수도 있다.
- [0118] 상기 단계(S240)에서 정형 데이터 처리부(140)에서 스크리닝 결과에 이상이 있다고 판단된 경우 인터페이스 제공부(130)는 이상 스크리닝 결과를 갖는 데이터를 수정하기 위한 보조 툴 인터페이스 화면을 제공할 수도 있다. 상기 보조 툴 인터페이스 화면은 ESG 관리 항목으로 잘못 선별된 데이터를 선별 결과에서 제거하는 입력을 수신하도록 구성된다. 정형 데이터 처리부(140)는 이 제거 입력을 수신하여 선별 결과를 수정하고, 정형 데이터로 변환할 ESG 관리 항목 목록을 최종 선별할 수도 있다.
- [0119] 또한, 상기 방법은, ESG 관리 항목 DB(230)에 저장하고 있는 ESG 관리 항목에 대응한 항목이 문서 내에 포함되었는지 여부에 따라 ESG 문서 내 항목에 대해 선별 상태를 지정하는 단계(S250)를 포함한다. 일부 실시 예들에서, 상기 단계(S250)는 단계(S240)에서 스크리닝 결과에 이상이 없는 경우에 수행될 수도 있다.
- [0120] 상기 단계(S250)에서 상기 데이터 처리부(140)는, 파싱된 ESG 문서의 선별 상태를 지정하기 위해, 파싱된 ESG 문서 내에 해당 항목이 포함되어 있으면 상기 ESG 문서를 제1 선별 상태로 지정하고, 해당 항목이 포함되어 있지 않으면 상기 ESG 문서를 제2 선별 상태로 지정할 수도 있다.
- [0121] 일부 실시 예들에서, 상기 ESG 문서의 선별 상태를 지정하는 단계(S250)는, 선별 여부를 가리키는 체크 박스를 온/오프 처리하는 것으로 구현될 수도 있다. 상기 일부 실시 예들에서, 상기 제1 선별 상태 또는 제2 선별 상태의 지정은 선별 여부를 가리키는 체크 박스를 온/오프 처리하는 것으로 구현될 수도 있다. 예를 들어, 상기 정형 데이터 처리부(140)는 파싱된 ESG 문서 내에 해당 항목이 포함되어 있으면 체크 박스를 온 처리하여 상기 ESG 문서를 제1 선별 상태로 지정하고, 해당 항목이 포함되어 있지 않으면 체크 박스를 오프 처리하여 상기 ESG 문서를 제2 선별 상태로 지정할 수도 있다.
- [0122] 또한, 상기 방법은, 선별 상태의 지정 결과에 따라 미리 저장된 복수의 ESG 관리 항목에 대한 최종 입력 필드 값의 입력 칸을 활성화하거나 또는 비활성화하는 단계(S260)를 포함한다.
- [0124] *일부 실시 예들에서, 상기 정형 데이터 처리부(140)는 인터페이스 제공부(130)로부터 선별 처리 결과에 이상이 없음을 나타낸 신호를 수신할 수도 있다. 상기 정형 데이터 처리부(140)는 해당 항목의 '최종 입력 필드 값'의 입력 칸의 활성화/비활성화는 상기 신호의 수신에 반응하여 수행될 수도 있다(S260).
- [0125] 상기 정형 데이터 처리부(140)는 선별 처리가 이상 없음을 나타낸 신호를 수신하면, 체크박스를 온 처리하여 제1 선별 상태로 지정한 경우 해당 항목의 '최종 입력 필드 값'의 입력 칸을 활성화하거나, 또는 체크박스를 오프 처리하여 제2 선별 상태로 지정한 경우 해당 항목의 '최종 입력 필드 값'의 입력 칸을 비활성화할 수도 있다.
- [0126] 상기 단계(S260)에서 상기 정형 데이터 처리부(140)는 문서에 포함된 것으로 판단된 해당 항목의 '최종 입력 필드 값'의 입력 칸을 활성화 또는 비활성화할 수도 있다. 상기 입력 칸의 활성화 또는 비활성화 결과는 보조 툴 인터페이스 화면을 통해 제공될 수도 있다.
- [0127] 도 3은, 본 출원의 일 실시 예에 따른, 항목의 최종 입력 필드 값의 입력 칸을 포함한 보조 툴 인터페이스 화면을 도시한다.
- [0128] 도 3을 참조하면, 상기 보조 툴 인터페이스 화면은 문서에 포함된 것으로 판단된 해당 항목을 포함할 수도 있다..
- [0129] 다시 도 2를 참조하면, 상기 방법은, 문서에 포함된 것으로 판단된, ESG 관리 항목에 대응한 항목의 내용을 획득하는 단계(S270)를 포함한다.
- [0130] 도 2b는, 도 2a의 단계(S270)의 세부 흐름도이다.
- [0131] 정형 데이터 처리부(140)는 제1 선별 상태로 지정된 ESG 관리 항목의 입력 값을 획득하도록 구성된다.
- [0132] 정형 데이터 처리부(140)는 제1 선별 상태로 지정된 ESG 관리 항목에 대응한 항목에 대해서, 그 내용을 획득하기 위해 파싱된 ESG 문서 또는 등록된 다른 ESG 문서 내 데이터를 추출할 수도 있다.
- [0134] *정형 데이터 처리부(140)는, ESG 문서 DB(220) 내 등록된 항목 중 상기 문서에 포함된 항목에 대응한 ESG 관리

항목에 매칭하는 등록 항목의 등록 값을 상기 과싱된 ESG 문서에 포함된 항목 내용의 입력 값으로 복사할 수도 있다. 그러면, 정형 데이터 처리부(140)는 상기 특정 항목에 대해서는, 과싱된 ESG 문서에서 인식된 데이터(예컨대, 기록 값) 대신 데이터베이스 서버(200)에 저장된 다른 ESG 문서 내 상기 특정 항목의 등록 데이터를 그대로 활용할 수 있다.

- [0135] 또는, 정형 데이터 처리부(140)는 인터페이스 제공부(130)가 제공한 보조 툴 인터페이스 화면을 통해 상기 ESG 문서의 항목에 대한 입력 값을 입력받을 수도 있다.
- [0136] 일부 실시 예들에서, 상기 단계(S270)는, 제1 선별 상태로 지정된 ESG 관리 항목의 상기 ESG 문서 내 위치를 제공하는 단계(S271); 및 제공된 위치에 기초하여 해당 ESG 관리 항목의 내용을 획득하는 단계(S273)를 포함한다.
- [0137] 또한, 일부 실시 예들에서, 상기 단계(S273)는, 사용자의 내용 입력에 기초하거나 또는 제공된 위치에 대한 주변 데이터에 기초하여 '최종 필드 입력 값'의 내용을 획득하는 것일 수도 있다.
- [0138] 정형 데이터 처리부(140)는 미리 저장된 ESG 관리 항목에 대응한 코드 값 및/또는 데이터 세트 값에 매칭한 데이터의 위치를 제공하고, 제공된 위치의 ESG 관리 항목에 대한 사용자의 내용 입력이 수신되면 해당 입력 값을 '최종 필드 입력 값'의 내용으로 획득할 수도 있다. 정형 데이터 처리부(140)가 매칭한 데이터의 위치를 인터페이스 제공부(130)로 공급하면, 인터페이스 제공부(130)는 매칭한 데이터의 위치를 포함한 보조 툴 인터페이스 화면을 표시하여, 해당 항목의 내용에 대한 사용자 입력을 유도할 수도 있다(S270).
- [0139] 또한, 상기 정형 데이터 처리부(140)는 제공된 위치를 기준으로 일정 범위 이내에 배치된, 주변 데이터에 기초하여 매칭한 데이터가 실제로 ESG 관리 항목에 대응하는 지를 판단할 수도 있다(S270). 상기 범위는 해당 위치 주변의 텍스트 배치 구조(예컨대, 텍스트 블록, 텍스트 라인 등), 텍스트의 수에 기초하여 설정된다. 예를 들어, 상기 범위는 제공된 위치 하단의 텍스트 블록일 수도 있다. 그러면, 해당 항목의 내용은 상기 하단의 텍스트 블록 내 텍스트의 일부 또는 전부로 획득될 수도 있다.
- [0140] 일부 실시 예들에서, 상기 단계(S270)는, 기존 ESG 문서 내 데이터를 이용하겠다는 사용자 입력을 수신한 경우, 상기 기존 ESG 문서 내 데이터를 해당 항목에 대한 내용으로 획득하는 단계(S275);를 더 포함할 수도 있다. 상기 단계(S271, S273)는 기존 ESG 문서 내 데이터를 이용하지 않겠다는 사용자 입력을 수신한 이후에 수행될 수도 있다.
- [0141] 상기 기존 ESG 문서는 평가 대상 기업의 이전 ESG 문서를 포함한다.
- [0142] ESG 문서는 단일 평가 대상 기업이 시계열 순서에 따라 생성하는 것이 일반적이다. 평가 대상 기업의 변화는 상대적으로 느리므로, 근시일 단위(예컨대, 분기, 반기, 년도) 차이의 ESG 문서는 부분적으로 동일한 항목에 대해 동일한 내용을 가질 수도 있다. 이 경우, 상기 정형 데이터 처리부(140)는 기존 ESG 문서의 데이터를 이번에 새롭게 생성할 ESG 문서에 그대로 이용할 수도 있다.
- [0143] 또한, 상기 방법은, 단계(S250)에서 선별된 ESG 관리 항목의 목록 및 단계(S270)에서 획득된 각각의 ESG 관리 항목의 내용을 테이블 변환 처리하여 정형 데이터를 생성하는 단계(S280)를 포함한다.
- [0144] 상기 단계(S280)에서 생성된 정형 데이터는 비정형 ESG 데이터가 ESG 서비스 제공 시스템(1)에서 인식 가능한 데이터 구조를 갖도록 변환된 것이다.
- [0145] 단계(S270)에서 획득된 내용은 단계(S250)에서 선별된 각각의 ESG 관리 항목에 매핑된다. 그러면, 정형 데이터 처리부(140)에서 항목/내용 값으로 이루어진 테이블로 변환 처리됨으로써 정형화된 데이터로 변환될 수도 있다.
- [0146] 정형 데이터 처리부(140)는 테이블 변환 처리 결과를 정형화된 데이터로 데이터베이스 서버(200)로 전송하여 ESG 관리 항목 DB(230)에 저장할 수도 있다.
- [0147] 다시 도 1을 참조하면, 상기 ESG 서비스 제공 시스템(1)은 이미 저장된 ESG 관리 항목을 수정, 삭제 또는 생성하는, 일련의 관리 동작을 수행하도록 더 구성될 수도 있다.
- [0148] 도 4는, 본 출원의 일 실시 예에 따른, ESG 관리 항목을 관리하는 과정의 흐름도이다.
- [0149] 도 4를 참조하면, 상기 과정은, ESG 관리 항목에 대한 수정, 삭제 또는 생성 요청을 수신하기 위해 보조 툴 인터페이스 화면을 표시하는 단계(S291)를 포함한다.
- [0150] 일부 실시 예들에서, 상기 보조 툴 인터페이스 화면의 표시는 ESG 관리 항목에 대한 관리 기능을 개시하는 동작에 따라 수행될 수도 있다.

- [0151] 상기 보조 툴 인터페이스 화면은 수정 또는 삭제 대상 항목의 선택을 유도하거나 및/또는 기존의 항목과의 관계 구조에 기초하여 신규 대상 항목을 생성하는 것을 유도하기 위해 구성된다.
- [0152] 도 5는, 본 출원의 일 실시 예에 따른, 항목 레벨 구조를 표시한 보조 툴 인터페이스 화면을 도시한다.
- [0153] 도 5를 참조하면, 상기 보조 툴 인터페이스 화면은 대분류별 항목 세트를 각각의 동일한 항목 레벨별로 분류하여 표시할 수도 있다.
- [0154] 다시 도 4를 참조하면, 상기 과정은, 이미 저장된 ESG 관리 항목이 저장되었는지 확인하는 단계(S292); ESG 관리 항목이 저장되어 있는 경우, ESG 관리 항목에 대한 수정, 삭제 또는 생성 요청을 수신하는 단계(S293); 및 ESG 관리 항목이 저장되어 있지 않는 경우, ESG 관리 항목에 대한 생성 요청을 수신하는 단계(S294)를 포함한다.
- [0155] 상기 전자 장치(100)는 도 5의 보조 툴 인터페이스 화면을 통해 수정 또는 삭제 대상 항목을 선택받고, 대상 항목의 내용을 수정하거나 대상 항목을 ESG 관리 항목 DB(230)에서 삭제할 수도 있다. 또는, 항목 레벨 구조에 기초하여 신규 대상 항목이 추가될 수도 있다.
- [0156] 단계(S293)에서 전자 장치(100)는 도 5의 보조 툴 인터페이스 화면 상에 표시된 '항목 레벨 구조'에서 수정할 항목을 선택하여 항목 수정 요청을 수신할 수도 있다.
- [0157] 단계(S293)에서 생성 요청은 기존의 ESG 관리 항목 세트에 새로운 관리 항목을 추가하는 것이다. 상기 전자 장치(100)는 도 5의 보조 툴 인터페이스 화면 상에 표시된 '항목 레벨 구조'에 따라 항목 추가 요청을 수신할 수도 있다.
- [0158] 상기 항목 수정 요청, 항목 추가 요청 등은 보조 툴 인터페이스 화면을 통해 수신하며, 그러면, 상기 전자 장치(100)는 선택된 항목의 항목 레벨을 획득한다.
- [0159] 단계(S294)에서 생성 요청은 ESG 관리 항목을 신설하는 것이다. 이는 단계(S293)의 항목 추가와 유사하므로, 자세한 설명은 생략한다.
- [0160] 또한, 일부 실시 예들에서, 도 5의 상기 과정은, 단계(S293)의 요청 대상 항목이 최하위 레벨 항목인지 여부에 따라 입력칸을 활성화 또는 비활성화하는 단계를 더 포함할 수도 있다.
- [0161] 상기 활성화 또는 비활성화하는 단계는, 단계(S293, S294)의 요청 대상 항목이 최하위 레벨 항목이 아닐 경우 대상 항목의 내용에 대한 입력 칸을 비활성화하는 단계(S295); 및 단계(S293, S294)의 요청 대상 항목이 최하위 레벨 항목일 경우 대상 항목의 내용에 대한 입력 칸을 활성화하는 단계(S296)를 포함한다.
- [0162] 상기 단계(S296)에서 상기 입력 칸을 통해 항목에 대한 '답변유형', 항목 내용의 '최종 입력 필드 값', 항목을 나타낼 '코드 값', '데이터 세트 값' 중 하나 이상의 항목 정보를 입력받을 수도 있다.
- [0163] 상기 인터페이스 제공부(130)는 항목에 대한 '답변유형', 항목 내용의 '최종 입력 필드 값', 항목을 나타낼 '코드 값', '데이터 세트 값' 중 하나 이상의 항목 정보와 같이 대상 항목의 내용을 입력받기 위한 보조 툴 인터페이스 화면을 제공할 수도 있다.
- [0164] 일부 실시 예들에서, 상기 보조 툴 인터페이스 화면은 활성화된 입력 칸을 포함한 제1 서브 영역 및 항목 구조 레벨을 포함한 제2 서브 영역을 포함할 수도 있다. '답변유형', '최종 입력 필드 값', '코드 값', '데이터 세트 값'에 대한 입력은 도 5의 보조 툴 인터페이스 화면의 하단(즉, 제1 서브 영역)에 위치할 수도 있다. 제2 서브 영역에는 대분류별 항목 세트를 각각의 동일한 항목 레벨별로 분류하여 표시된다.
- [0165] 또한, 상기 과정은 단계(S296) 이후, 활성화된 입력 칸을 통해 수정 또는 생성할 대상 항목의 내용을 입력받는 단계(S297); 및 단계(S297)에서 입력된 항목의 내용을 ESG 관리 항목 DB(230)에 저장하는 단계(S298)를 더 포함할 수도 있다.
- [0166] 이와 같이 상기 ESG 서비스 제공 시스템(1)에 의해 비정형 ESG 데이터가 파싱 및 정형화되어 데이터베이스로 구축됨으로써, 빅데이터 관점에서 분석 및 가공 처리에 따라 적절하게 ESG 데이터를 활용할 수 있다.
- [0168] 추가적으로, 파싱된 ESG 문서의 텍스트는 ESG 문서를 자동완성하기 위한 기계학습 모델의 학습 데이터로 이용될 수도 있다. 이를 위해, 상기 ESG 서비스 제공 시스템(1)은 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템에 연결되어, 상기 파싱된 ESG 문서의 텍스트를 제공할 수도 있다.

- [0169] 도 6은, 본 출원의 일 측면에 따른, ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템의 개략도이고, 도 7은, 본 출원의 일 실시 예에 따른, ESG 자동완성 모델을 생성하는 과정의 개략도이다.
- [0170] 도 6 및 도 7을 참조하면, 상기 ESG 서비스 제공 시스템(1)은 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템과 연결됨으로써, 상기 ESG 서비스 제공 시스템(1)이 제공하는 ESG 서비스는 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하는 서비스를 더 포함한다. 상기 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템은 데이터 웨어하우스 시스템(300)과 연결된 서비스 서버(200-1)를 포함한다.
- [0171] 데이터 웨어하우스 시스템(300)은 클라우드 데이터 웨어하우스 시스템 또는 기타 데이터 웨어하우스 시스템을 간단히 지칭하는 것으로서, 데이터 분석 및 보고를 위해, 하나 이상의 다른 소스(disparate sources)로부터 통합된 데이터의 중앙 저장소를 포함한 네트워크-기반 저장 시스템이다. 상기 데이터 웨어하우스 시스템(300)은 기업용 분석 보고서를 생성하기 위해 사용될 수 있는 현재, 과거 데이터를 저장할 수도 있다. 이를 위해, 데이터 웨어하우스 시스템(300)은 비즈니스 인텔리전스 도구(business intelligence tools), 데이터를 추출하고, 변환하며, 이를 저장소로 로드하기 위한 도구, 및 메타데이터를 관리하고 검색하기 위한 도구를 제공한다.
- [0172] 상기 데이터 웨어하우스 시스템(300)은 함수(예를 들어, 구조화된 질의 언어(structured Query Language, SQL) 함수), 테이블 함수, 또는 프로시저(procedure)를 데이터 웨어하우스 시스템 외부에 있고, 웹 애플리케이션 프로그래밍 인터페이스(Web application programming interface)로서 노출된 원격 소프트웨어 코드로 바인딩(bind)하기 위한 기능을 제공하도록 구성될 수도 있다.
- [0173] 상기 데이터 웨어하우스 시스템(300)은 외부에서 관리되는 웹 API 관리 시스템 제공된 클라우드 컴퓨팅 서비스 플랫폼(예를 들어, Amazon Web Services®(AWS), Microsoft Azure® 또는 Google Cloud Services®)에서 외부에서 구현된 웹 엔드포인트(web endpoint)(예를 들어, 하이퍼텍스트 전송 프로토콜(HyperText Transfer Protocol, HTTP) 표현 상태 변경(Representational State Transfer, REST) 엔드포인트)에 의해 지원되는 (backed) 함수 및 저장된 프로시저를 작성하기(author) 위한 메커니즘을 사용자에게 제공한다. 사용자는 웹 엔드포인트를 제공하고(provisioning), 저장 플랫폼 내의 비즈니스 로직을 기초로 엔드포인트를 구성할 책임이 있다. 일부 경우에서, 웹 API 관리 시스템은 람다 함수(Lambda functions)에 대한 요청을 프록시하고, 다른 경우에서, 웹 API 관리 시스템은 요청을 변환하며, 이를 데이터 웨어하우스 시스템 외부에 있는 제3자 소프트웨어 구성요소로 전달한다. 데이터 웨어하우스 시스템(300)은 이의 외부 소프트웨어 구성요소에 의해 제공되는 외부 함수가 사용자 정의 함수, 사용자-정의 테이블 함수 및 저장된 프로시저와 같이 질의에서 사용되는 것을 가능하게 한다.
- [0174] 데이터 웨어하우스 시스템(300)은 원격 소프트웨어 구성요소에 의해 제공된 외부 기능의 호출을 가능하게 하기 위해 다양한 데이터 객체를 저장한다. 데이터 객체는 클라우드 컴퓨팅 플랫폼에 의해 제공되는 웹 API 관리 시스템을 통해 외부 기능을 호출하는 데 사용될 임시 보안 크리덴셜(credentials)을 획득하기 위해 네트워크-기반 데이터 웨어하우스에서 사용하는 정보를 저장한다. 질의의 실행 중에, 네트워크-기반 데이터 웨어 하우스 시스템은 임시 보안 크리덴셜을 사용하여 클라우드 컴퓨팅 서비스 플랫폼의 인증 시스템을 통해 타겟 엔드포인트로 인증하고, 질의에서 정의된 타겟 데이터의 배치(batches)로 엔드포인트에서 기능을 호출한다. 타겟 데이터는 예를 들어, 이진 데이터, 자바스크립트 객체 표기법(JavaScript Object Notation, JSON) 인코딩된 데이터, 또는 확장 가능한 마크업 언어(extensible Markup Language, XML)와 같은 다른 텍스트 포맷을 포함한다. 타겟 데이터는 HTTP 요청/응답과 함께 인라인으로(inline) 전달되거나, 또는 클라우드 컴퓨팅 서비스 플랫폼(예를 들어, Amazon® Simple Storage Service(S3®))에 의해 제공되는 일반적으로 액세스되는 저장소에 기록될 수 있다. 데이터 웨어하우스 시스템(300)에 저장된 사용자 데이터는 HTTP 요청 및 응답을 통해 전달하기에 적합한 포맷으로 인코딩된다.
- [0175] 도 7을 참조하면, 상기 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템에서 국문, 영문 또는 제3 언어로 기재된 ESG 문서로부터 ESG 텍스트를 추출하고 이를 상기 데이터 웨어하우스 시스템(300)에 저장한다.
- [0176] 상기 ESG 텍스트는 ESG 문서로부터 추출되었기 때문에, ESG 형식에 맞는 의미 및 표현을 가지는 텍스트이다. 상기 ESG 형식에 맞는 텍스트는 ESG 형식의 카테고리로 분류될 수도 있다. 상기 ESG 텍스트는 ESG 분야에서 활용되는 단어(예컨대, 키워드) 또는 문장을 포함할 수도 있다. 또한, 상기 ESG 텍스트는 ESG 분야와 관련된 타-분야에서 활용되는 단어, 문구(phrase) 또는 문장을 더 포함할 수도 있다. ESG 분야와 관련된 타-분야는 ESG 문서를 생성하는데 활용된 참조 문서의 분야일 수도 있다.

- [0177] 상기 ESG 문서는 정부 웹 페이지, 기업 웹 페이지, 기타 웹 사이트 등에서 이미 공개되어 있는 국/내외 평가 대상 기업의 ESG 지속가능 보고서, 또는 개별 기업의 ESG 담당자가 작성하여 배포한 기타 ESG 문서를 포함한다. 상기 평가 대상 기업은 사업체로서의 기업만을 지칭하지 않고 기관, 비영리 단체 등의 다양한 형태의 단체를 포괄하여 지칭할 수 있다.
- [0178] 일부 실시 예들에서, 상기 ESG 문서는, 도 1 및 도 6에 도시된, 정부 문서 DB(310), 고객 문서 DB(320), 및 웹 사이트 문서 DB(330) 중 하나 이상의 외부 DB로부터 획득되어 상기 데이터 웨어하우스 시스템(300)에 저장될 수도 있다.
- [0179] 상기 정부 문서 DB(310)는 정부 데이터 포털에 액세스하여 검색된 적어도 하나의 기업에 대한 ESG 문서를 수집할 수도 있다.
- [0180] 상기 고객 문서 DB(320)는 상기 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템과 관련된 고객, 예를 들어, 상기 시스템(1)에서 제공하는 서비스에 가입한 고객 기업의 데이터 포털에 액세스하여 해당 고객에 대한 ESG 문서를 수집할 수도 있다. 단일 고객에 대해 다수의 ESG 문서가 수집될 수도 있다.
- [0181] 상기 웹 사이트 문서 DB(330)는 주요 웹 사이트로 미리 지정된 하나 이상의 웹 사이트 각각에 액세스하고 액세스한 웹 사이트에서 공개된 ESG 문서를 검색하여 수집할 수도 있다.
- [0182] 서비스 서버(200-1)는 네트워크 서버로 구현되는 다수의 컴퓨터 시스템 또는 컴퓨터 소프트웨어이다. 여기서, 네트워크 서버란, 사설 인트라넷 또는 인터넷과 같은 컴퓨터 네트워크를 통해 다른 네트워크 서버와 통신할 수 있는 하위 장치와 연결되어 작업 수행 요청을 접수하고 그에 대한 작업을 수행하여 수행 결과를 제공하는 컴퓨터 시스템 및 컴퓨터 소프트웨어(네트워크 서버 프로그램)를 의미한다. 그러나 이러한 네트워크 서버 프로그램 이외에도, 네트워크 서버 상에서 동작하는 일련의 응용 프로그램과 경우에 따라서는 내부에 구축되어 있는 각종 데이터베이스를 포함하는 넓은 개념으로 이해되어야 할 것이다. 상기 서비스 서버(200-1)는 네트워크 서버, 웹 서버, 파일 서버, 슈퍼컴퓨터, 데스크탑 컴퓨터 등과 같은 컴퓨팅 장치들의 임의의 유형 또는 유형들의 조합으로 구현될 수도 있다. 이를 위해, 서비스 서버(200-1)는 데이터를 처리할 수 있는 적어도 하나의 프로세서, 데이터를 저장하는 메모리, 데이터를 송/수신하는 통신부를 포함한다.
- [0183] 서비스 서버(200-1)는 사용자의 ESG 문서를 보다 쉽게 작성하기 위해 ESG 형식에 맞는 텍스트(예컨대, 문장)을 추천하는, ESG 문서 자동완성 서비스를 제공할 수도 있다. 이를 위해, 서비스 서버(200-1)는 사용자의 전자 장치(100-1)로부터 수신한 입력 값을 미리 학습된 ESG 자동완성 모델에 적용하여 입력 값 다음 순서에 배치될 텍스트, 단어(예컨대, 키워드) 또는 문장을 예측한다. 또한, 상기 서비스 서버(200-1)는 예측 결과를 상기 사용자의 전자 장치(100-1)로 전달하도록 구성될 수도 있다.
- [0184] 일부 실시 예들에서, 상기 서비스 서버(200-1)는 학습부(210-1)를 더 포함할 수도 있다. 도 7에 도시된 것처럼, 상기 데이터 웨어하우스 시스템(300)은 저장한 ESG 텍스트 데이터를 ESG 자동완성 모델을 생성하기 위한 샘플 텍스트 데이터로 학습부(210-1)로 공급할 수도 있다.
- [0185] 학습부(210-1)는 샘플 텍스트 데이터를 자연어 처리하여 미리 설계된 ESG 자동완성 모델을 학습할 수도 있다. 특정 실시 예들에서, 상기 학습부(210-1)는 샘플 텍스트 데이터를 자연어 처리하는 전처리 동작, 샘플 텍스트의 토큰에 기초하여 트레이닝 데이터 세트를 형성하는 동작, 및 상기 트레이닝 데이터 세트를 이용하여 미리 설계된 신경망 구조를 갖는 ESG 자동완성 모델을 학습하는 동작을 수행하도록 구성될 수도 있다.
- [0186] 도 7에 도시된 것처럼, 학습부(210-1)는 수신한 샘플 텍스트의 말뭉치(예컨대, 문장)를 자연어 처리하는, 학습을 위한 전처리 동작을 수행한다.
- [0187] 상기 전처리 동작은, 토큰화 처리 동작 및/또는 특수 토큰 부가 동작을 포함할 수도 있다.
- [0188] 토큰화 처리 동작은 전처리 대상 텍스트의 말뭉치를 미리 설정된 최소 분할 단위인 토큰 단위로 분할하여 복수의 토큰을 생성하는 동작이다. 상기 토큰 단위는, 예를 들어 단어 단위 또는 형태소 단위로 미리 설정될 수도 있다. 문장이 학습부(210-1)로 제공되면, 상기 학습부(210-1)는 문장을 토큰 단위로 분할하여 해당 문장의 토큰 세트를 형성할 수도 있다. 각 토큰 세트는 단일 말뭉치로부터 생성된 복수의 토큰으로 이루어진다.
- [0189] 상기 특수 토큰 부가 동작은 미리 정의된 복수의 특수 토큰 중 하나 이상의 특수 토큰을 전처리 대상 텍스트의 말뭉치에 부가하는 동작이다. 상기 특수 토큰은 상기 ESG 자동완성 모델의 출력 텍스트가 보다 자연스러운 텍스트 스타일로 표현되는데 사용되는 특수 토큰이다. 텍스트를 가리키는 상기 토큰과 달리, 상기 특수 토큰은 텍스트를 가리키지 않을 수도 있다. 이하, 설명의 명료성을 위해, 말뭉치의 텍스트를 토큰 단위로 분할하여 생성된

토큰을 텍스트 토큰으로 지칭하여, 본 출원의 실시 예들을 보다 상세히 서술한다.

- [0190] 상기 특수 토큰은, 예를 들어, 모르는 단어에 대한 토큰, 배치 데이터의 길이를 맞추기 위한 토큰, 문장의 시작을 가리키는 토큰, 문장의 종결을 가리키는 토큰, 및/또는 문장 텍스트 이외 기타 토큰을 포함할 수도 있다. 특수 토큰이 샘플 텍스트에 추가되고 특수 토큰이 추가된 샘플 텍스트의 말뭉치가 토큰화될 경우, 특수 토큰이 추가된 샘플 텍스트의 말뭉치의 토큰 세트가 입력되는 상기 ESG 자동완성 모델은 입력 토큰의 텍스트를 보다 용이하게 디코딩할 수도 있다.
- [0191] 일부 실시 예들에서, 상기 특수 토큰 부가 동작은 상기 전처리 대상 텍스트의 말뭉치를 토큰으로 분할할 필요가 없거나 또는 ESG 자동완성 모델의 출력 텍스트가 표현되는 언어에 대해 미리 연관된 언어 규칙이 있을 경우에 수행될 수도 있다.
- [0192] 이러한 전처리 동작을 통해, 상기 학습부(210-1)는 샘플 텍스트의 말뭉치가 토큰 단위로 분할된 텍스트 토큰으로 이루어진 토큰 세트, 분할 없이 특수 토큰만이 추가된 토큰(들)로 이루어진 토큰 세트, 또는 특수 토큰이 샘플 텍스트의 말뭉치에 추가된 뒤 분할된 텍스트 토큰으로 이루어진 토큰 세트를 이용하여 트레이닝 데이터 세트를 형성한다.
- [0193] 일부 실시 예들에서, 상기 학습부(210-1)는 복수의 트레이닝 샘플로 이루어진 트레이닝 데이터 세트를 형성할 수도 있다. 각각의 트레이닝 샘플은 트레이닝 데이터 및 레이블 데이터를 포함한다.
- [0194] 상기 복수의 트레이닝 샘플 각각은 복수의 샘플 텍스트의 말뭉치로부터 획득된 토큰 세트로부터 각각 형성될 수도 있다. 각 트레이닝 샘플의 트레이닝 데이터, 레이블 데이터는 동일한 샘플 텍스트의 말뭉치에 대응한 토큰 세트로부터 획득된다.
- [0195] 상기 트레이닝 데이터는 샘플 텍스트의 말뭉치로부터 획득된 토큰 세트에 포함된 토큰(들)을 해당 샘플 텍스트의 말뭉치 상의 순서대로 배열한 토큰 배열을 포함할 수도 있다. 예를 들어, 트레이닝 데이터는 해당 샘플 문장에서 맨 처음의 토큰으로부터 맨 끝의 토큰까지의 토큰 배열을 포함할 수도 있다. 트레이닝 데이터의 토큰 배열은 입력 텍스트(즉, 샘플 텍스트)의 말뭉치의 전부에 대응한다.
- [0196] 상기 레이블 데이터는 트레이닝 데이터와 동일한 샘플 텍스트의 말뭉치로부터 획득된 토큰 세트에 포함된 토큰(들) 중 일부 토큰을 제거하고 나머지 토큰을 샘플 텍스트의 말뭉치 상의 순서대로 배열한 토큰 배열을 포함할 수도 있다. 즉, 레이블 데이터의 토큰 배열은 트레이닝 데이터의 토큰 배열에서 일부 토큰이 제거된 것이므로, 잔여 토큰 배열로 지칭될 수도 있다.
- [0197] 상기 잔여 토큰 배열을 이루는 토큰은 샘플 텍스트의 말뭉치 상에서 미리 설정된 순번의 텍스트 토큰부터 맨 끝의 텍스트 토큰을 포함할 수도 있다. 일부 실시 예들에서, 상기 미리 설정된 순번은 맨 처음의 다음 순번일 수도 있다. 그러면, 레이블 데이터는 해당 샘플 텍스트의 말뭉치(예컨대, 샘플 문장)에서 맨 처음의 텍스트 토큰의 다음 텍스트 토큰으로부터 맨 끝의 텍스트 토큰을 포함한 잔여 토큰 배열을 포함할 수도 있다. 즉, 레이블 데이터의 잔여 토큰 배열은 입력 텍스트(즉, 샘플 텍스트)의 말뭉치의 전부에서 상기 미리 설정된 순번의 직전 순번의 텍스트 토큰까지가 제거된, 일부 텍스트의 말뭉치에 대응한다.
- [0198] 학습부(210-1)에 의해 생성된 상기 트레이닝 데이터, 레이블 데이터를 포함한 트레이닝 데이터 세트는 ESG 자동완성 모델을 학습하는데 이용된다.
- [0199] 상기 ESG 자동완성 모델은 입력 텍스트를 자연어 처리하여 적어도 부분적으로 복원하도록 설계된 뉴럴 네트워크를 포함한다. 예를 들어, 상기 ESG 자동완성 모델은 PLM 또는 기타 자연어 처리 기반 네트워크 모델일 수도 있다. 상기 ESG 자동완성 모델은 입력 텍스트의 말뭉치의 다음 순번에 배치되기에 적합한 텍스트의 말뭉치를 복원하도록 구성될 수도 있다. 일 예에서, 상기 ESG 자동완성 모델은 인코더 및 디코더를 포함할 수도 있다.
- [0200] 학습부(210-1)에 의해, 상기 ESG 자동완성 모델은 입력 텍스트의 말뭉치와 잔여 텍스트의 말뭉치 간의 연관 관계를 추론하여 입력 텍스트의 말뭉치의 그 다음 순번으로 배치되기 적합한 텍스트의 말뭉치를 예측하도록 학습된다. 여기서 텍스트의 말뭉치는 단어(예컨대, 키워드), 문구 또는 문장일 수도 있다. 잔여 텍스트의 말뭉치는 전술한 미리 설정된 순번의 텍스트 토큰으로부터 맨 끝의 텍스트 토큰까지의 잔여 토큰 배열이 나타난 텍스트의 말뭉치이다.
- [0201] 상기 학습부(210)에 의해, ESG 자동완성 모델은 학습을 위한 샘플 텍스트의 말뭉치에서 특정 텍스트 말뭉치의 다음에 위치한 텍스트 말뭉치를 적절하게 복원하는 동작을 반복함으로써, 모델 내 파라미터가 특정 텍스트 말뭉

치의 다음에 위치하기 적합한 텍스트 말뭉치를 예측하도록 학습된다.

- [0202] 이하, 인코더 및 디코더를 포함한 상기 일 예시의 ESG 자동완성 모델을 이용하여 학습 과정에 대해 보다 상세히 서술한다.
- [0203] 학습부(210-1)는 트레이닝 세트 내 각 트레이닝 샘플의 트레이닝 데이터를 상기 ESG 자동완성 모델에 입력한다. 상기 인코더는 입력 토큰을 자연어 처리하여 입력 토큰의 특징을 추출하도록 구성된다. 상기 특징은 벡터 형태 또는 맵과 같은 행렬 형태로 추출될 수도 있다. 상기 일 예에서, 상기 인코더의 입력 토큰은 트레이닝 데이터에 포함된 토큰 배열이다.
- [0204] 디코더는 입력 토큰의 특징에 기초하여 트레이닝 데이터의 토큰 배열 중에서 미리 설정된 순번의 텍스트 토큰으로부터 맨 끝의 텍스트 토큰을 포함한 잔여 배열 토큰을 복원하도록 구성된다. 전술한 트레이닝 데이터에 포함된 토큰 배열이 인코더에 입력되면, 상기 디코더는 상기 미리 설정된 순번의 텍스트 토큰으로부터 맨 끝의 텍스트 토큰까지 복원할 수도 있다. 디코더의 복원 값은 입력 텍스트의 말뭉치에서 맨 처음의 텍스트 토큰의 다음 텍스트 토큰으로부터 맨 끝까지의 텍스트 토큰을 복원한 결과로서, 각 트레이닝 샘플별 예측 값으로 활용된다.
- [0205] 도 8은, 본 출원의 일 실시 예에 따른, ESG 자동완성 모델에서 출력 텍스트를 예측하는 동작의 개략도이다.
- [0206] 도 8을 참조하면, 상기 디코더는 그리드 탐색 방식으로 잔여 배열 토큰을 복원할 수도 있다. 상기 디코더는 특정 토큰의 다음 순서에 배치 가능한 복수의 후보 토큰 중 가능성이 가장 높은 토큰을 선택하여, 잔여 배열 토큰을 복원한다. 도 8에 도시된 것처럼 복수의 후보 토큰은 ESG 분야의 텍스트로 이루어진 토큰이다. 디코더는 잔여 배열 토큰의 앞/뒤 맥락에 기초하여 잔여 배열 토큰의 위치에 배치될 가능성이 가장 높은 토큰을 선택한다.
- [0207] 상기 학습부(210-1)는 추천 텍스트를 추천하는 기능을 학습하기 위해, ESG 자동완성 모델의 파라미터를 조정할 수도 있다. 상기 ESG 자동완성 모델의 파라미터는 미리 설계된 인공 신경망 구조의 파라미터를 포함할 수도 있다.
- [0208] 상기 학습부(210-1)는 산출된 예측 값(예컨대, 디코더의 출력 값)과 실제 값 간의 오차가 감소하거나 최소화되도록 ESG 자동완성 모델의 파라미터를 조정할 수도 있다. 여기서, 실제 값은 출력 값을 산출하는데 이용된 트레이닝 샘플에 함께 포함된 레이블 데이터의 값이다.
- [0209] 상기 출력 값과 실제 값 간의 오차가 최소화될 경우, 학습된 ESG 자동완성 모델은 입력 텍스트의 말뭉치의 다음 순번에 배치될 텍스트의 말뭉치를 정확하게 예측할 수도 있다. ESG 자동완성 모델이 입력 텍스트의 말뭉치와 잔여 텍스트의 말뭉치 간의 연관 관계를 보다 정확하게 추론했다면, 상기 입력 텍스트의 말뭉치에서 미리 설정된 순번 이전까지의 텍스트 토큰을 제외한 나머지 텍스트의 말뭉치가 보다 정확하게 예측될 것이다.
- [0210] 이러한 과정을 통해 학습 완료된 ESG 자동완성 모델은 ESG 문서 상에서 입력 텍스트의 말뭉치의 다음 위치에 배치되기 적합한 텍스트를 추천 텍스트로 예측하도록 구성된다. 상기 출력 텍스트는 ESG 문서의 형식을 만족하는 텍스트로서, 입력 텍스트의 맥락에 연결되는 맥락을 가지면서 ESG 분야에 특화된 표현을 갖는 텍스트일 수도 있다.
- [0211] 일부 실시 예들에서, 상기 ESG 자동완성 모델은 단일 입력 텍스트의 말뭉치가 입력될 경우, 상기 단일 입력 텍스트의 말뭉치의 다음 순번에 배치될 텍스트의 말뭉치를 하나 이상의 텍스트 말뭉치로 산출할 수도 있다. 예를 들어, 상기 ESG 자동완성 모델은 단일 문장이 입력될 경우 상기 단일 문장의 다음 순번에 배치되기에 적합한 후보 문장을 하나 또는 복수 개 산출할 수도 있다. 또한, 일부 실시 예들에서, 상기 ESG 자동완성 모델은 복수의 추천 텍스트의 말뭉치를 산출하고, 각 추천 텍스트의 말뭉치별 확률 값에 기초하여 일부 또는 전부의 추천 텍스트의 말뭉치를 최종 추천 텍스트의 말뭉치로 출력할 수도 있다.
- [0212] 예를 들어, 상기 ESG 자동완성 모델은 입력 텍스트의 다음에 위치할 추천 텍스트를 산출하는 동작을 수행할 때 마다 모델의 시드 값이 임의의 값을 갖도록 설정될 수도 있다. 동일한 입력 텍스트에 대해서 해당 추천 텍스트를 산출하는 동작을 복수 회 반복할 경우, 상기 복수 회 중 2이상의 서로 다른 추천 텍스트의 말뭉치를 산출할 수도 있다.
- [0213] 이 경우, 상기 추천 텍스트의 개수는 동일한 입력 텍스트에 대해 해당 추천 텍스트를 산출하는 동작의 반복 회수를 가리키는, 미리 지정된 하이퍼 파라미터에 의존할 수도 있다.
- [0214] 일부 실시 예들에서, 상기 ESG 자동완성 모델은 각 추천 텍스트의 말뭉치별 확률 값이 미리 설정된 임계 확률 값 이상인 일부 추천 텍스트의 말뭉치를 최종 추천 텍스트의 말뭉치로 출력할 수도 있다. 그러면, 상기 시스템

(1)은 사용자의 선택 명령에 대응한 추천 텍스트의 말뭉치를 상기 입력 텍스트의 다음 위치에 배치할 수도 있다. 사용자의 선택에 따라 배치된 추천 텍스트의 말뭉치는 다다음 순서의 추천 텍스트의 말뭉치를 산출하기 위한 새로운 입력 텍스트로 이용된다. 이러한 과정의 반복으로 인해, 사용자는 몇 번의 선택 명령을 통해 ESG 문서의 내용의 일부 또는 전부를 완성할 수도 있다.

- [0215] 특정 실시 예들에서, 상기 디코더는 핏팅 레이어를 포함할 수도 있다. 상기 핏팅 레이어는 ESG 자동완성 모델이 입력 텍스트의 다음 순번의 텍스트로서 상기 ESG 형식에 맞는 의미 및/또는 표현의 텍스트를 예측하기 위해 상기 디코더의 말단에 배치된다. 일부 실시 예들에서, 상기 디코더가 (예컨대, 소프트맥스와 같은) 확률 함수를 포함할 경우, 상기 핏팅 레이어는 상기 확률 함수 이전의 데이터 경로 상에 배치될 수도 있다.
- [0216] 상기 학습부(210-1)는, 상기 핏팅 레이어를 이용하여 상기 ESG 자동완성 모델이 입력 텍스트에 함축된 의미를 적어도 부분적으로 학습하는, 강화 학습 동작을 수행할 수도 있다.
- [0217] 상기 핏팅 레이어는 추천 문장이 출력되기 이전에, 입력 텍스트가 입력될 경우 예측되어 출력될 다음 순번의 텍스트(예컨대, 단어, 문구 또는 문장)에 대해서 ESG 형식에 보다 가까운 데이터에 가중치를 부여하도록 구성된다. 일부 실시 예들에서, ESG 형식에 적합한 정도에 따라 보다 높은 가중치 값이 부여될 수도 있다. 상기 핏팅 레이어에 의해 상기 ESG 자동완성 모델은 ESG 형식에 맞는 ESG 분야의 텍스트에 함축된 의미론적 부분을 적어도 부분적으로 학습할 수 있다.
- [0218] 일부 실시 예들에서, 상기 ESG 형식에 적합한 정도는 ESG 문서에 나타난 빈도에 기초할 수도 있다.
- [0219] 만약 상기 ESG 자동완성 모델이 GPT(Generative Pre-trained Transformer) 모델에 기반하여 설계될 경우를 가정해보자. 상기 ESG 자동완성 모델이 핏팅 레이어를 포함하지 않고 입력 텍스트의 다음 단어, 문구 또는 문장을 예측하도록 미리 학습될 경우, 여러 카테고리의 말뭉치 데이터가 상기 다음 단어, 문구 또는 문장으로 예측될 수도 있다. 상기 특정 실시 예들에서, 상기 ESG 자동완성 모델은 핏팅 레이어를 통해 ESG 형식 카테고리의 말뭉치 데이터에서 입력 텍스트의 다음 단어, 문구 또는 문장을 구현할 수도 있다.
- [0220] 그 결과, 상기 핏팅 레이어를 포함한 ESG 자동완성 모델은 ESG 문서의 형식을 만족하면서 입력 텍스트의 말뭉치의 다음 위치에 배치되기에 적합한(즉, 가능성이 가장 높은) 텍스트의 말뭉치를 예측하도록 학습된다.
- [0221] 상기 서비스 서버(200-1)는 도 7과 같이 학습된 ESG 자동완성 모델을 활용하여 사용자가 ESG 문서를 작성하는 작업을 지원한다. 일부 실시 예들에서, 상기 서비스 서버(200-1)는 사용자의 입력 텍스트의 다음 위치에 배치될 추천 텍스트를 예측하여 사용자의 ESG 문서 완성을 지원하는 ESG 문서 생성부(250-1)를 더 포함할 수도 있다. 상기 ESG 문서 생성부(250-1)는 상기 ESG 자동완성 모델을 이용하도록 구성된다. 상기 ESG 문서 생성부(250-1)의 동작에 대해서는 아래의 도 10을 참조해 보다 상세히 서술한다.
- [0222] 상기 전자 장치(100-1)는 서비스 서버(200-1)의 예측 동작을 개시하기 위한 입력 텍스트를 입력하고, 서비스 서버(200-1)에서 예측한 추천 텍스트 또는 입력 텍스트 및 예측한 추천 텍스트를 포함한, 적어도 부분적으로 자동 완성된 ESG 문서의 내용을 출력하도록 구성된다. 또한, 일부 실시 예들에서, 상기 전자 장치(100-1)는 예측한 추천 텍스트 중 문서 내용으로 이용할 하나의 텍스트를 선택하도록 더 구성될 수도 있다.
- [0223] 상기 전자 장치(100-1)는 서비스 서버(200-1)와 통신하는 클라이언트 단말 장치로서, 데이터를 처리할 수 있는 적어도 하나의 프로세서, 데이터를 저장하는 메모리, 데이터를 송/수신하는 통신부를 포함한다. 상기 전자 장치는, 예를 들어, 랩탑 컴퓨터, 기타 컴퓨팅 장치, 태블릿, 셀룰러 폰, 스마트 폰, 스마트 워치, 스마트 글래스, 헤드 마운트 디스플레이(HMD), 기타 모바일 장치, 기타 웨어러블 장치일 수도 있다.
- [0224] 상기 전자 장치(100-1), 서비스 서버(200-1) 또는 시스템(1)이 다른 구성요소를 포함할 수도 있다는 것이 통상의 기술자에게 명백할 것이다. 예를 들어, 상기 서비스 서버(200-1)는 데이터 엔트리를 위한 입력 장치, 및 인쇄 또는 다른 데이터 표시를 위한 출력 장치를 포함하는, 본 명세서에 서술된 동작에 필요한 다른 하드웨어 요소를 포함할 수도 있다. 또한, 상기 시스템(1)은 서비스 서버(200-1)와 외부 장치(예를 들어, 사용자 단말, 또는 외부 데이터베이스 등) 사이를 연결하는 네트워크, 네트워크 인터페이스 및 프로토콜 등을 더 포함할 수 있다.
- [0225] 대안적인 실시 예들에서, 상기 학습부(210-1)는 서비스 서버(200-1)의 내부에 포함되는 것으로 제한되지 않는다. 상기 학습부(210-1)는 서비스 서버(200-1)의 외부 구성요소로 구현될 수도 있다. 이 경우, 상기 학습부(210-1)는 데이터 웨어하우스 시스템(300), 서비스 서버(200-1)와 전기 통신하도록 구성된다. 상기 서비스 서버(200-1)는 외부의 학습부(210-1)로부터 미리 학습된 ESG 자동완성 모델을 수신하고, 수신된 모델을 활용하여 사

용자의 ESG 문서를 보다 쉽게 작성하기 위해 ESG 형식에 맞는 문장을 추천하는, ESG 텍스트 자동완성 서비스를 제공할 수도 있다.

- [0226] 상기 서비스 서버(200-1)의 학습부(210-1)는, 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계 학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 단계;를 수행할 수 있다. 상기 기계학습 모델은 복수의 트레이닝 샘플로 이루어진 트레이닝 데이터를 이용하여 학습되는 것으로서, 입력 토큰의 특징을 추출하도록 구성된 인코더 - 상기 입력 토큰은 트레이닝 데이터에 포함된 토큰 배열임 -, 및 입력 토큰의 특징에 기초하여 트레이닝 데이터의 토큰 배열 중에서 미리 설정된 순번의 텍스트 토큰으로부터 맨 끝의 텍스트 토큰을 포함한 잔여 배열 토큰을 복원하도록 구성된 디코더를 포함한다. 상기 파싱된 ESG 문서의 텍스트를 ESG 문서를 자동 완성하기 위한 기계학습 모델의 학습 데이터로 이용하여 상기 기계학습 모델을 학습하는 단계는, 상기 트레이닝 데이터 세트 내 트레이닝 데이터를 상기 학습부(210-1)에 구현된 기계학습 모델에 입력하여 해당 트레이닝 샘플에 대한 예측 값을 산출하는 단계 - 상기 기계학습 모델 내 인코더는 상기 트레이닝 데이터로 입력된 입력 토큰의 특징을 추출하고, 상기 기계학습 모델 내 디코더는 상기 입력 토큰의 특징에 기초하여 복원된 잔여 배열 토큰의 복원 결과를 예측 값으로 출력함; 및 해당 트레이닝 샘플에 대해 산출된 예측 값과 상기 해당 트레이닝 샘플에 포함된 레이블 데이터의 값인 실제 값 간의 오차를 감소하도록 상기 기계학습 모델의 파라미터를 조정하는 단계;를 포함할 수 있다.
- [0227] 도 9는, 본 출원의 다른 일 측면에 따른, ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하는 방법의 흐름도이다.
- [0228] 도 9의 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하는 방법(이하, ESG 텍스트 추천 방법)은 도 1의 ESG 서비스 제공 시스템(1)에 연결된 도 6의 ESG 문서를 자동 완성하는데 적합한 텍스트를 추천하도록 구성된 시스템과 같은 하나 이상의 컴퓨팅 장치에 의해 수행될 수도 있다. 여기서, ESG 텍스트는 ESG 문서의 형식을 만족하는 텍스트를 지칭한다.
- [0229] 도 9를 참조하면, 상기 ESG 텍스트 자동완성 방법은, 사용자의 전자 장치(100-1)로부터 상기 ESG 문서의 내용 중 일부 내용으로 기재될 ESG 텍스트를 완성하기 위한 기초 텍스트를 수신하는 단계(S410)를 포함한다. 상기 기초 텍스트는 다음 순번에 배치될 텍스트를 예측하는 단서로 활용된다.
- [0230] 상기 기초 텍스트를 수신하는 단계(S410)에서 기초 텍스트는 ESG 문서에 기재될 단어(예컨대, 키워드), 문구 또는 문장이다. 상기 기초 텍스트는, 예를 들어 해당 ESG 문서를 완성하는데 요구되는 필수 텍스트 중 일부 또는 전부의 텍스트, 또는 ESG 문서의 내용 전체 중 일부분의 내용을 나타낸 텍스트일 수도 있다.
- [0231] 완성할 ESG 문서의 내용 전체가 해당 ESG 문서에 기재되기 위해, 내용 전체 중 일부분에 속하는 기초 텍스트가 입력되면, 나머지 내용 중 일부 또는 전부가 추천되어 ESG 문서가 자동 완성된다.
- [0232] 일부 실시 예들에서, 상기 단계(S410)의 기초 텍스트는 적합한 텍스트를 추천하도록 구성된 시스템이 운영하는 특정 웹 페이지에 액세스할 경우 제공되는 인터페이스 화면을 통해 입력될 수도 있다.
- [0233] 상기 인터페이스 화면은 완성하길 원하는 ESG 형식을 표현하도록 구성될 수도 있다. 상기 인터페이스 화면을 통해 사용자는 자신이 완성할 ESG 문서의 형식에 내용을 직접 입력하거나 또는 자동 완성 결과를 실시간으로 제공받을 수도 있다.
- [0235] 또한, 상기 ESG 텍스트 추천 방법은, 상기 단계(S410)에서 수신한 기초 텍스트를 미리 학습된 ESG 자동완성 모델에 입력하여 기초 텍스트의 다음 순번에 배치될 추천 텍스트를 예측하는 단계(S420)를 포함한다.
- [0236] 일부 실시 예들에서, 상기 추천 텍스트를 산출하는 단계(S420)는, 상기 기초 텍스트의 말뭉치를 전처리하여 입력 데이터를 형성하는 단계, 및 형성된 입력 데이터를 상기 미리 학습된 ESG 자동완성 모델에 입력하여 적어도 하나의 추천 텍스트를 산출하는 단계를 포함할 수도 있다.
- [0237] 상기 단계(S420)에서 수행되는 전처리 동작은 토큰화 처리 동작 및/또는 특수 토큰 부가 동작을 포함한다. 상기 전처리 동작은 ESG 자동완성 모델의 트레이닝 데이터를 생성하는데 이용된 전처리 동작에 대응할 수도 있다. 예를 들어, 상기 트레이닝 데이터가 토큰화 처리 동작만 전처리해서 생성될 경우, 서비스 서버(200-1)는 추천 텍스트를 산출하기 위해 기초 텍스트의 말뭉치를 토큰화 처리 동작하여 입력 데이터를 생성할 수도 있다.
- [0238] 상기 단계(S420)의 ESG 자동완성 모델은 단계(S410)의 기초 텍스트의 다음 순번에 배치되기에 적합한, ESG 텍스트의 말뭉치를 예측하도록 미리 학습된다. 상기 ESG 자동완성 모델은 기초 텍스트의 다음 순번에 배치될 가능성이 있는 하나 이상의 텍스트의 말뭉치를 산출하고, 산출된 하나 이상의 텍스트의 말뭉치 중 적어도 하나의 텍

트의 말뭉치를 추천 텍스트의 말뭉치로 출력하도록 구성될 수도 있다.

- [0239] 상기 ESG 텍스트의 말뭉치는 단어(예컨대, 키워드), 문구 또는 문장일 수도 있다.
- [0240] 일부 실시 예들에서, 상기 ESG 자동완성 모델은 하나 이상의 ESG 텍스트의 말뭉치를 추천 텍스트의 말뭉치로 출력할 수도 있다.
- [0241] 서비스 서버(200-1)는 출력된 하나 이상의 추천 텍스트의 말뭉치를 전자 장치(100-1)로 전송할 수도 있다. 상기 전자 장치(100-1)는 수신한 추천 텍스트를 사용자에게 제공할 수도 있다. 일부 실시 예들에서, 상기 전자 장치(100-1)는 기초 텍스트 및 각각의 추천 텍스트를 순서대로 배열한 결과를 제공할 수도 있다.
- [0242] 도 10은, 본 출원의 일 실시 예에 따른, 추천 텍스트의 말뭉치의 출력 결과를 도시한다.
- [0243] 도 10을 참조하면, 상기 단계(S410)에서 "대상(은) 온실가스 감축을 위해 적극적인 설비 투자와 활동을 꾸준히 수행하고 있습니다"의 문장이 기초 텍스트로 입력될 경우, 해당 문장의 다음 순번에 배치되기 적합한 하나 이상의 추천 텍스트가 상기 ESG 자동완성 모델로부터 출력될 수도 있다. 상기 ESG 자동완성 모델은 복수의 추천 문장을 산출하고, 상기 복수의 추천 문장별 확률 값에 기초하여 도 10에 도시된 것처럼 제1 문장 내지 제5 문장을 추천 텍스트로 출력할 수도 있다. 상기 제1 문장 내지 제5 문장 각각은 기초 문장의 다음 순번에 배치되어 제공될 수도 있다.
- [0244] 상기 단계(S420)에서 이용되는 ESG 자동완성 모델은 도 7, 도 8을 참조해서 위에서 서술하였는 바, 자세한 설명은 생략한다.
- [0245] 다시 도 9를 참조하면, 상기 ESG 텍스트 추천 방법은, ESG 문서의 형식에 추천 텍스트를 기재하는 단계(S430)를 더 포함할 수도 있다.
- [0246] 상기 단계(S420)에서 단일 추천 텍스트가 출력될 경우, 출력된 단일 추천 텍스트는 기초 텍스트의 다음 순번에 곧바로 기재될 수도 있다.
- [0247] 상기 단계(S420)에서 복수의 추천 텍스트가 출력될 경우, 상기 단계(S430)에서 기재되는 추천 텍스트는 복수의 추천 텍스트 중 어느 하나의 텍스트이다. 기재될 추천 텍스트는 사용자의 선택에 의해 결정될 수도 있다. 예를 들어, 사용자의 전자 장치(100-1)는 도 10의 화면을 표시하여 제1 문장 내지 제5 문장 중 어느 하나의 추천 문장을 기초 문장의 다음 순번에 기재할 추천 문장으로 선택하는 사용자 명령을 유도한다. 어느 하나의 문장이 선택되면, 전자 장치(100-1)는 선택된 추천 문장의 정보를 서비스 서버(200-1)로 전송하여 상기 기초 텍스트 및 추천 텍스트의 내용으로 상기 ESG 문서의 현재 내용을 업데이트한다.
- [0248] 또한, 상기 ESG 텍스트 추천 방법은, 단계(S420)에서 예측된 추천 텍스트를 상기 ESG 자동완성 모델에 입력하여 상기 단계(S420)의 추천 텍스트의 다음 순번에 배치될 새로운 추천 텍스트를 예측하는 단계(S440)를 더 포함할 수도 있다. 그러면, ESG 자동완성 모델은 추천 텍스트의 맥락에 연결되는 맥락을 가지면서 ESG 분야에 특화된 표현을 갖는 새로운 추천 텍스트를 출력한다.
- [0249] 또한, 상기 ESG 텍스트 추천 방법은, 단계(S440)에서 출력된 새로운 추천 텍스트를 단계(S430)의 ESG 문서의 형식에 추가 기재하는 단계(S450)를 더 포함할 수도 있다. 상기 단계(S430)의 ESG 문서의 형식은 기초 텍스트 및 추천 텍스트의 내용을 포함한 상태이다.
- [0250] 상기 단계(S450)에서 새로운 추천 텍스트가 기재되면, 작성 중인 ESG 문서는 단계(S410)의 기초 텍스트, 단계(S420)의 추천 텍스트 및 단계(S440)의 새로운 추천 텍스트의 내용을 포함하도록 업데이트된다. 단계(S450)의 동작은, 단계(S430)의 동작과 유사하므로 자세한 설명은 생략한다.
- [0251] 일부 실시 예들에서, 상기 ESG 텍스트 추천 방법은, ESG 문서 전체 내용이 완성될 때까지 상기 단계(S440, S450)를 반복하는 단계(S460)를 더 포함할 수도 있다.
- [0252] 그러면, 사용자는 단계(S410)에서 전자 장치(100-1)를 통해 최소한의 텍스트를 기초 텍스트로 입력하여, ESG 문서를 자동 완성할 수도 있다.
- [0254] 하드웨어를 이용하여 본 발명의 실시 예를 구현하는 경우에는, 본 출원의 실시 예들을 수행하도록 구성된 ASICs(application specific integrated circuits) 또는 DSPs(digital signal processors), DSPDs(digital signal processing devices), PLDs(programmable logic devices), FPGAs(field programmable gate arrays) 등이 본 출원의 구성요소에 구비될 수 있다.

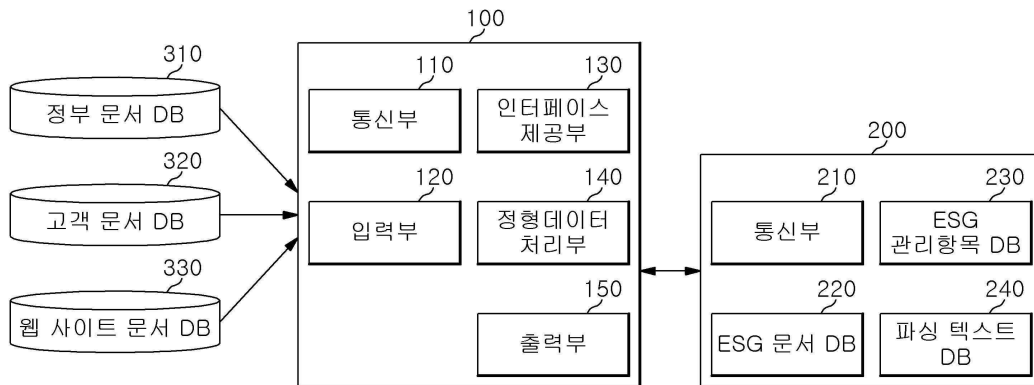
[0255] 이상에서 설명한 본 출원의 실시 예들에 따른 ESG 보조 툴을 이용하여 정형화된 ESG 데이터로 이루어진 ESG 데이터베이스를 구축하는 방법 및 이를 수행하는 ESG 서비스 제공 시스템(1)에 의한 동작은 적어도 부분적으로 컴퓨터 프로그램으로 구현되어, 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 예를 들어, 프로그램 코드를 포함하는 컴퓨터-판독가능 매체로 구성되는 프로그램 제품과 함께 구현되고, 이는 기술된 임의의 또는 모든 단계, 동작, 또는 과정을 수행하기 위한 프로세서에 의해 실행될 수 있다.

[0256] 상기 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터에 의하여 읽힐 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등을 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 본 실시예를 구현하기 위한 기능적인 프로그램, 코드 및 코드 세그먼트(segment)들은 본 실시예가 속하는 기술 분야의 통상의 기술자에 의해 용이하게 이해될 수 있을 것이다.

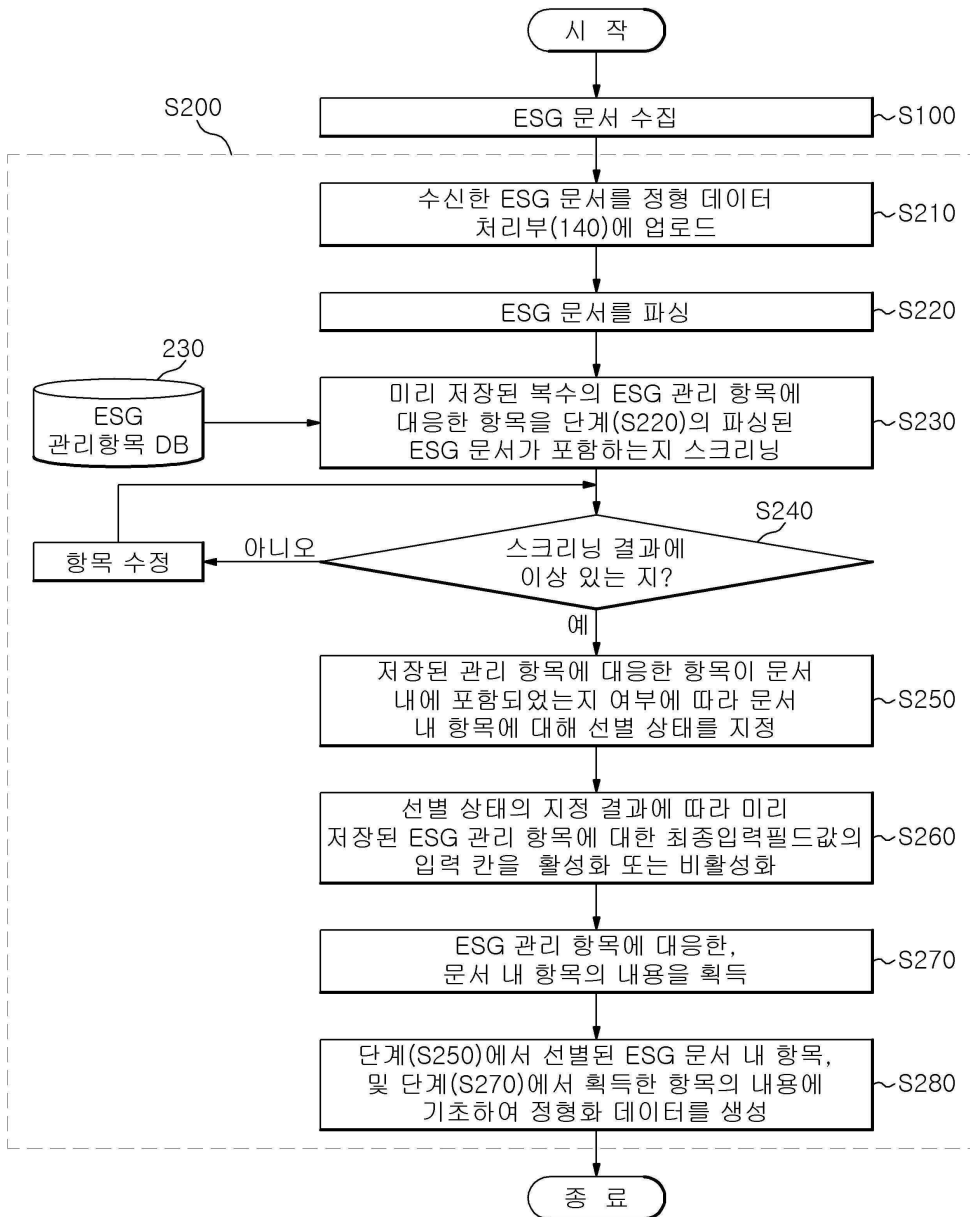
[0257] 이상에서 살펴본 본 발명은 도면에 도시된 실시예들을 참고로 하여 설명하였으나 이는 예시적인 것에 불과하며 당해 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 실시예의 변형이 가능하다는 점을 이해할 것이다. 그러나, 이와 같은 변형은 본 발명의 기술적 보호범위 내에 있다고 보아야 한다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위의 기술적 사상에 의해서 정해져야 할 것이다.

도면

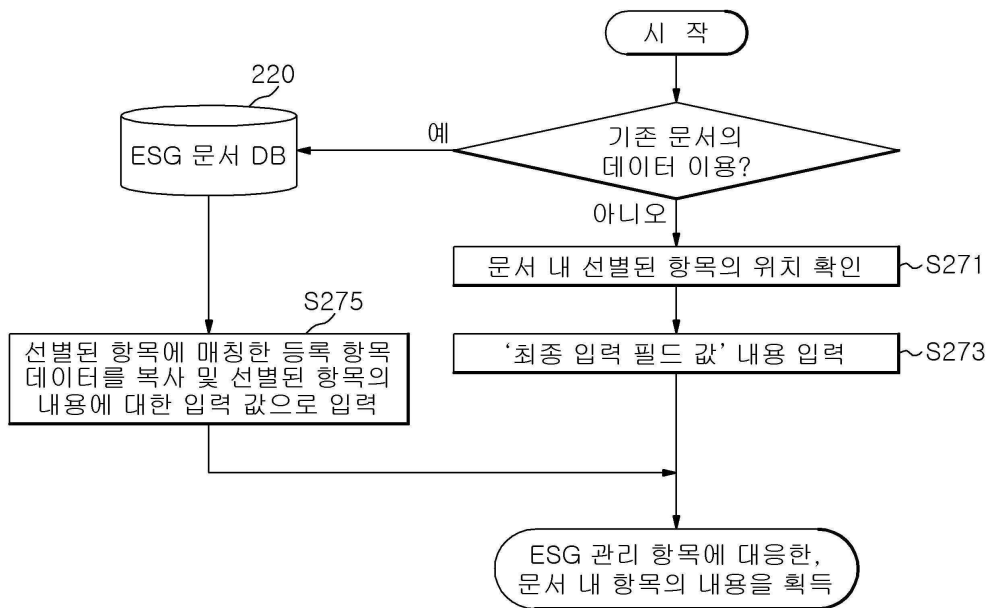
도면1



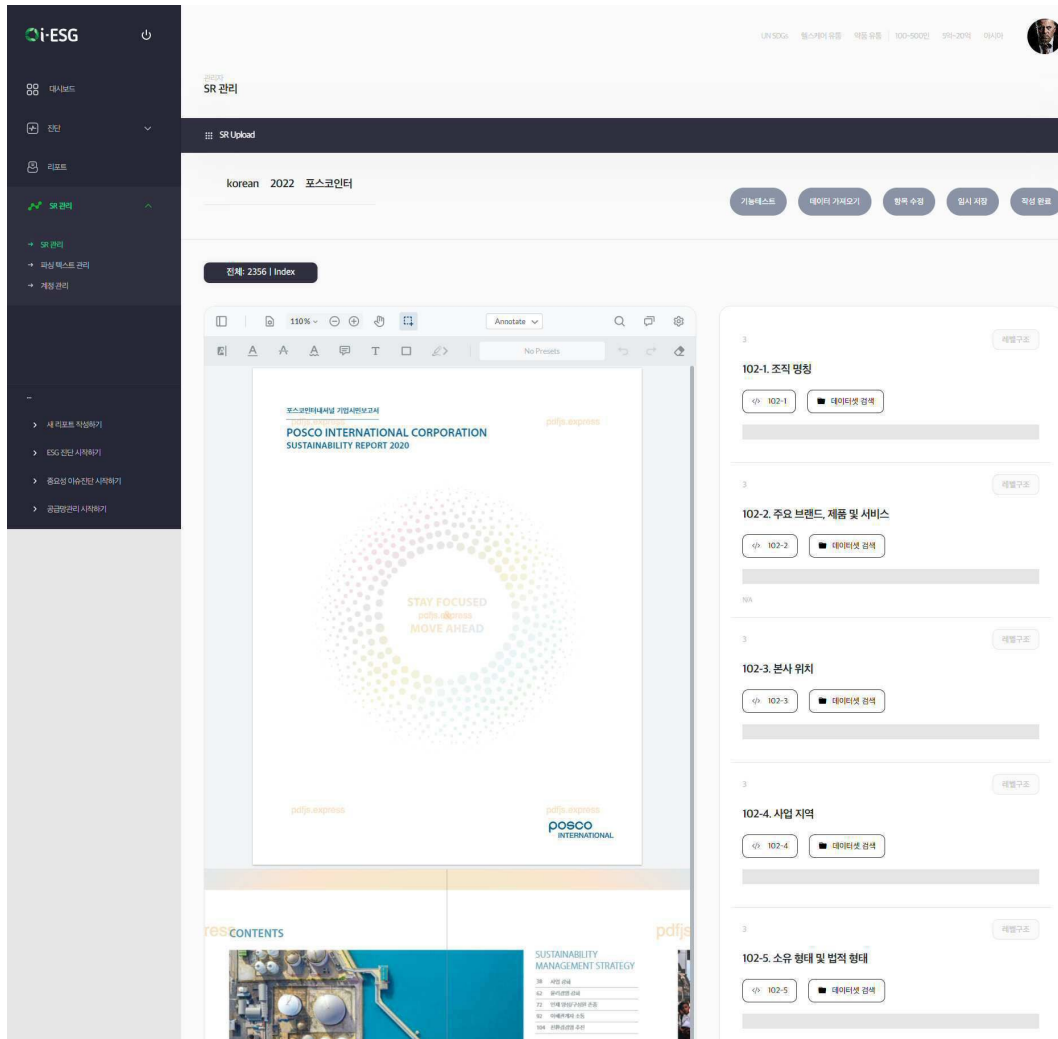
도면2a



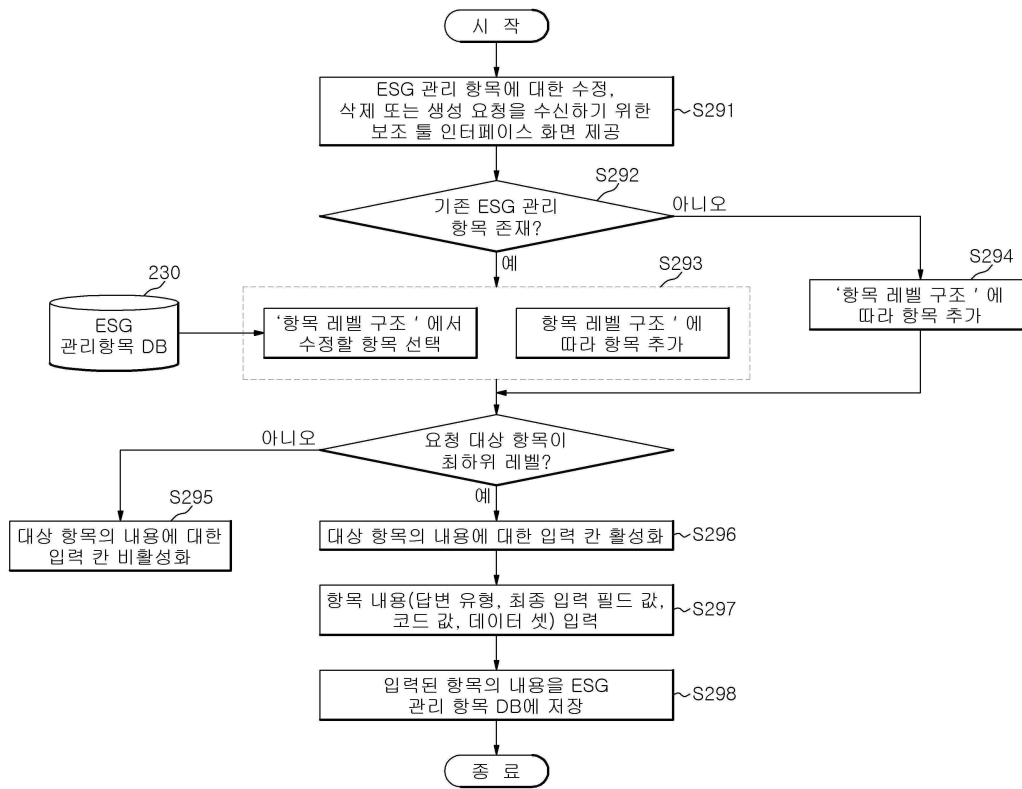
도면2b



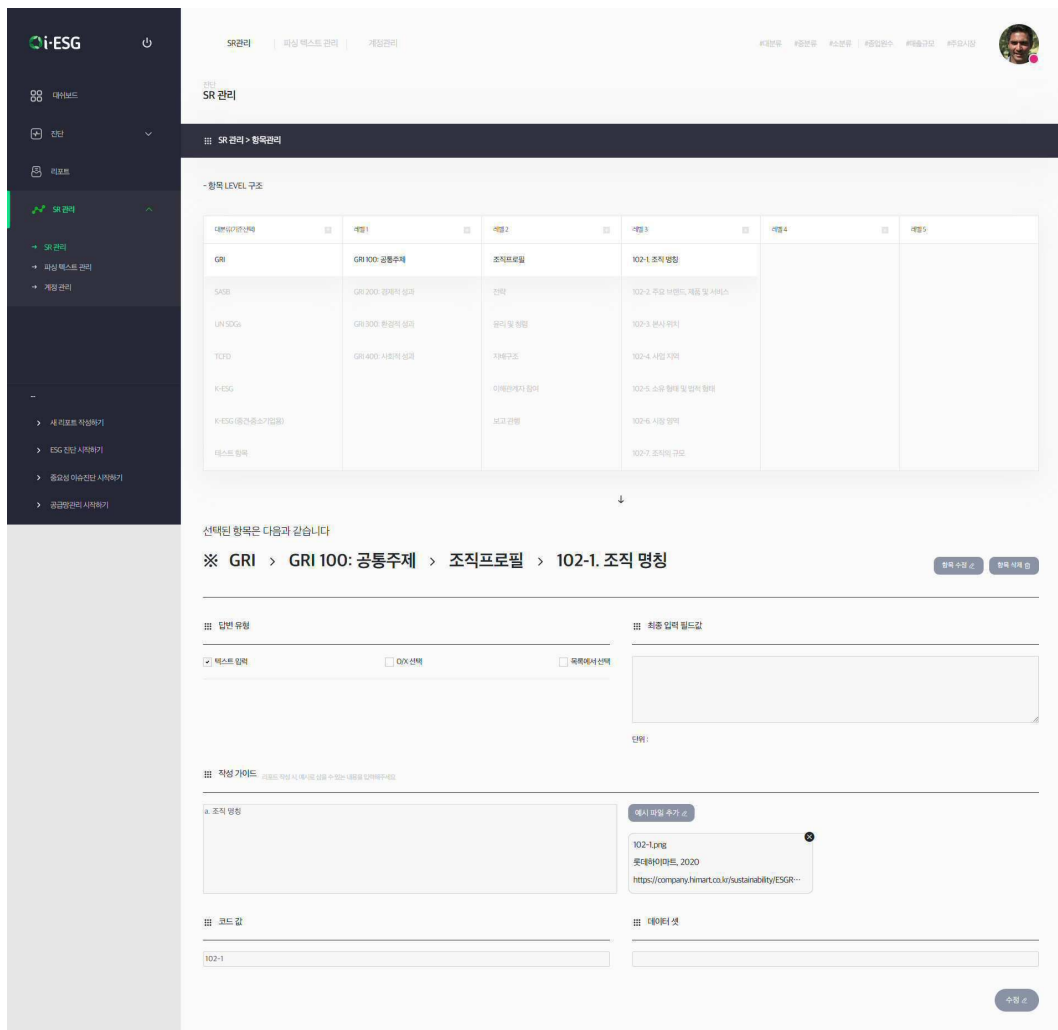
도면3



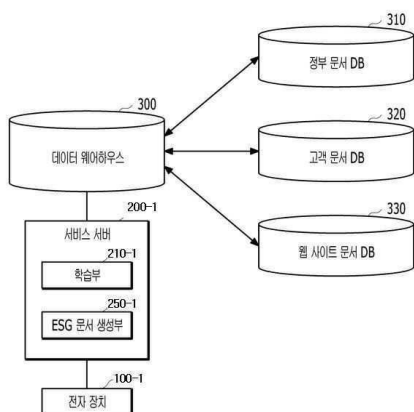
도면4



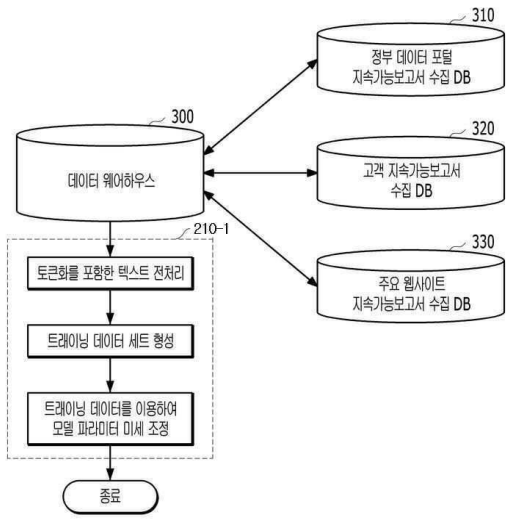
도면5



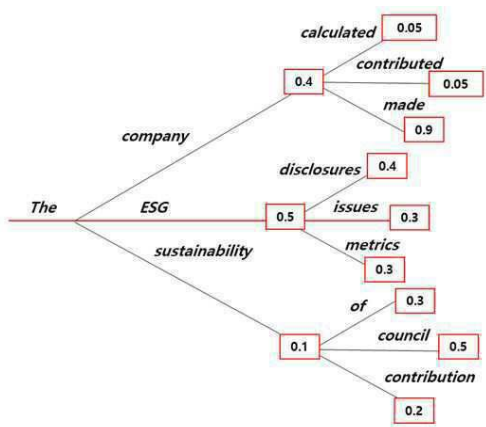
도면6



도면7



도면8



도면9

