



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2020년11월06일

(11) 등록번호 10-2176035

(24) 등록일자 2020년11월02일

(51) 국제특허분류(Int. Cl.)  
G06N 5/02 (2006.01) G06F 16/35 (2019.01)(52) CPC특허분류  
G06N 5/02 (2019.01)  
G06F 16/35 (2019.01)

(21) 출원번호 10-2019-0056190

(22) 출원일자 2019년05월14일

심사청구일자 2019년05월14일

(56) 선행기술조사문헌

US20150363509 A1\*

(뒷면에 계속)

(73) 특허권자

주식회사 엔씨소프트

서울특별시 강남구 테헤란로 509 (삼성동)

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(72) 발명자

이경호

서울특별시 서대문구 신촌로 50 연세대학교 제4공학관 812호

서승민

서울특별시 서대문구 신촌로 50 연세대학교 제4공학관 812호

(뒷면에 계속)

(74) 대리인

두호특허법인

전체 청구항 수 : 총 18 항

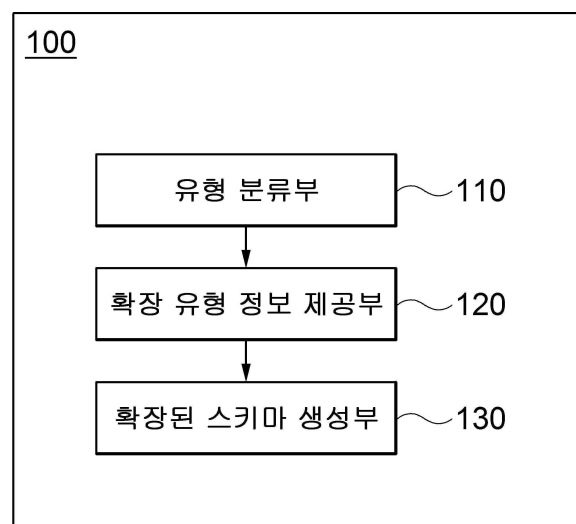
심사관 : 서광훈

(54) 발명의 명칭 지식 그래프 스키마 확장 방법 및 장치

## (57) 요약

지식 그래프 스키마 확장 방법 및 장치가 개시된다. 일 실시예에 따른 지식 그래프 스키마 확장 방법은, 사전 학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하는 동작, 상기 분류 결과에 기초하여 사용자에게 상기 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공하는 동작, 상기 사용자에게 의해 상기 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우, 상기 하나 이상의 신규 유형을 포함하도록 상기 시드 스키마를 확장하여 확장된 스키마를 생성하는 동작을 포함한다.

대표도 - 도1



(72) 발명자

**신동훈**

경기도 성남시 분당구 대왕판교로 644번길 12

**신동원**

경기도 성남시 분당구 대왕판교로 644번길 12

**유동현**

경기도 성남시 분당구 대왕판교로 644번길 12

**이우철**

경기도 성남시 분당구 대왕판교로 644번길 12

(56) 선행기술조사문헌

US09852231 B1

US20090012842 A1

W02018156184 A1

W02018226888 A1

US20160012336 A1

US20180075359 A1

\*는 심사관에 의하여 인용된 문헌

## 명세서

### 청구범위

#### 청구항 1

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨터 장치에서 수행되는 방법으로서,

사전 학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하는 동작;

상기 분류 결과에 기초하여 사용자에게 상기 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공하는 동작; 및

상기 사용자에 의해 상기 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우, 상기 하나 이상의 신규 유형을 포함하도록 상기 시드 스키마를 확장하여 확장된 스키마를 생성하는 동작을 포함하고,

상기 유형에 대한 정보를 제공하는 동작은, 상기 분류 결과에 기초하여 상기 하나 이상의 개체가 분류된 상기 복수의 유형 각각의 입도 점수(granularity score)를 산출하고, 상기 복수의 유형 중 상기 산출된 입도 점수가 기 설정된 값 이하인 유형을 상기 확장이 필요한 유형으로 결정하는 지식 그래프 스키마 확장 방법.

#### 청구항 2

청구항 1에 있어서,

상기 시드 스키마에 포함된 복수의 유형은, 복수의 개체 유형을 포함하는 지식 그래프 스키마 확장 방법.

#### 청구항 3

청구항 2에 있어서,

상기 시드 스키마에 포함된 복수의 유형은, 하나 이상의 관계 유형을 더 포함하는 지식 그래프 스키마 확장 방법.

#### 청구항 4

삭제

#### 청구항 5

청구항 1에 있어서,

상기 하나 이상의 신규 유형에 기초하여 상기 텍스트 데이터에 대한 상기 분류 결과를 갱신하는 동작; 및

상기 텍스트 데이터, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 학습 데이터로 이용하여 상기 유형 분류 모델을 재학습시키는 동작을 더 포함하는 지식 그래프 스키마 확장 방법.

#### 청구항 6

청구항 5에 있어서,

상기 갱신하는 동작은, 상기 하나 이상의 개체 중 상기 확장이 필요한 유형으로 분류된 개체 중 일부가 상기 신규 유형으로 분류되도록 상기 분류 결과를 갱신하는 지식 그래프 스키마 확장 방법.

#### 청구항 7

청구항 5에 있어서,

상기 재학습시키는 동작은, 상기 텍스트 데이터를 상기 유형 분류 모델의 입력 데이터로 이용하고, 상기 갱신된

분류 결과 및 상기 확장된 스키마를 상기 유형 분류 모델의 타겟 데이터로 이용하여 상기 유형 분류 모델을 재학습시키는 지식 그래프 스키마 확장 방법.

#### 청구항 8

청구항 5에 있어서,

상기 재학습된 유형 분류 모델을 이용하여 상기 텍스트 데이터에 포함된 하나 이상의 개체를 각각 상기 확장된 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하는 동작; 및

상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 사용자에게 상기 신규 유형에 대한 오류 정보를 제공하는 동작을 더 포함하는 지식 그래프 스키마 확장 방법.

#### 청구항 9

청구항 8에 있어서,

상기 오류 정보를 제공하는 동작은, 상기 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 시드 스키마의 엔트로피를 산출하고, 상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 확장된 스키마의 엔트로피를 산출하며, 상기 시드 스키마의 엔트로피 및 상기 확장된 스키마의 엔트로피에 기초하여 상기 오류 정보를 제공하는 지식 그래프 스키마 확장 방법.

#### 청구항 10

청구항 9에 있어서,

상기 오류 정보를 제공하는 동작은, 상기 확장된 스키마의 엔트로피가 상기 시드 스키마의 엔트로피보다 높은 경우, 상기 오류 정보를 제공하는 지식 그래프 스키마 확장 방법.

#### 청구항 11

하나 이상의 명령어를 저장하는 메모리; 및

상기 하나 이상의 명령어를 실행하는 하나 이상의 프로세서를 포함하고,

상기 하나 이상의 프로세서는,

사전 학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하고,

상기 분류 결과에 기초하여 사용자에게 상기 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공하고,

상기 사용자에게 의해 상기 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우, 상기 하나 이상의 신규 유형을 포함하도록 상기 시드 스키마를 확장하여 확장된 스키마를 생성하되,

상기 분류 결과에 기초하여 상기 하나 이상의 개체가 분류된 상기 복수의 유형 각각의 입도 점수(granularity score)를 산출하고, 상기 복수의 유형 중 상기 산출된 입도 점수가 기 설정된 값 이하인 유형을 상기 확장이 필요한 유형으로 결정하는 지식 그래프 스키마 확장 장치.

#### 청구항 12

청구항 11에 있어서,

상기 시드 스키마에 포함된 복수의 유형은, 복수의 개체 유형을 포함하는 지식 그래프 스키마 확장 장치.

#### 청구항 13

청구항 12에 있어서,

상기 시드 스키마에 포함된 복수의 유형은, 하나 이상의 관계 유형을 더 포함하는 지식 그래프 스키마 확장 장치.

#### 청구항 14

삭제

#### 청구항 15

청구항 11에 있어서,

상기 하나 이상의 프로세서는,

상기 하나 이상의 신규 유형에 기초하여 상기 텍스트 데이터에 대한 상기 분류 결과를 갱신하고,

상기 텍스트 데이터, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 학습 데이터로 이용하여 상기 유형 분류 모델을 재학습시키는 지식 그래프 스키마 확장 장치.

#### 청구항 16

청구항 15에 있어서,

상기 하나 이상의 프로세서는, 상기 하나 이상의 개체 중 상기 확장이 필요한 유형으로 분류된 개체 중 일부가 상기 신규 유형으로 분류되도록 상기 분류 결과를 갱신하는 지식 그래프 스키마 확장 장치.

#### 청구항 17

청구항 15에 있어서,

상기 하나 이상의 프로세서는, 상기 텍스트 데이터를 상기 유형 분류 모델의 입력 데이터로 이용하고, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 상기 유형 분류 모델의 타겟 데이터로 이용하여 상기 유형 분류 모델을 재학습시키는 지식 그래프 스키마 확장 장치.

#### 청구항 18

청구항 15에 있어서,

상기 하나 이상의 프로세서는,

상기 재학습된 유형 분류 모델을 이용하여 상기 텍스트 데이터에 포함된 하나 이상의 개체를 각각 상기 확장된 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하고,

상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 사용자에게 상기 신규 유형에 대한 오류 정보를 제공하는 지식 그래프 스키마 확장 장치.

#### 청구항 19

청구항 18에 있어서,

상기 하나 이상의 프로세서는, 상기 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 시드 스키마의 엔트로피를 산출하고, 상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 확장된 스키마의 엔트로피를 산출하며, 상기 시드 스키마의 엔트로피 및 상기 확장된 스키마의 엔트로피에 기초하여 상기 오류 정보를 제공하는 지식 그래프 스키마 확장 장치.

#### 청구항 20

청구항 19에 있어서,

상기 하나 이상의 프로세서는, 상기 확장된 스키마의 엔트로피가 상기 시드 스키마의 엔트로피보다 높은 경우, 상기 오류 정보를 제공하는 지식 그래프 스키마 확장 장치.

### 발명의 설명

### 기술 분야

개시되는 실시예들은 지식 그래프 스키마를 확장하는 기술과 관련된다.

[0001]

## 배경 기술

- [0002] 지식 그래프(Knowledge graph)는 개체 인식, 개체 쌍 사이의 관계 추출 및 질문 응답과 같은 자연어 처리를 위한 필수적 자원으로 주목을 끌고 있다. 이러한, 지식 그래프 스키마(schema)를 생성하기 위해 개체 유형 할당(entity typing) 및 관계 추출(relation extraction) 방식을 통해 비정형 자연어 문서로부터 개체 또는 개체 쌍 사이의 관계의 의미 유형(semantic type)을 식별한다.
- [0003] 이와 관련하여, 종래의 기술은 특정 도메인에 대한 전문가가 수동으로 지식 그래프 스키마를 생성하기 때문에, 지식 그래프 스키마를 생성하는데 많은 시간과 높은 비용이 소요된다. 또한, 종래의 기술은 개체 또는 개체 쌍 사이의 관계를 미리 정의된 유형에만 할당해야 하므로, 개체 또는 개체 쌍 사이의 관계를 잘못된 유형에 할당하거나 지식 그래프 스키마를 확장할 수 없다는 문제점이 있다.

## 발명의 내용

### 해결하려는 과제

- [0004] 개시되는 실시예들은 지식 그래프 스키마 확장 방법 및 장치를 제공하기 위한 것이다.

### 과제의 해결 수단

- [0005] 일 실시예에 따른 지식 그래프 스키마 확장 방법은, 사전 학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하는 동작, 상기 분류 결과에 기초하여 사용자에게 상기 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공하는 동작 및 상기 사용자에게 의해 상기 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우, 상기 하나 이상의 신규 유형을 포함하도록 상기 시드 스키마를 확장하여 확장된 스키마를 생성하는 동작을 포함한다.
- [0006] 상기 시드 스키마에 포함된 복수의 유형은, 복수의 개체 유형을 포함할 수 있다.
- [0007] 상기 시드 스키마에 포함된 복수의 유형은, 하나 이상의 관계 유형을 더 포함할 수 있다.
- [0008] 상기 유형에 대한 정보를 제공하는 동작은, 상기 분류 결과에 기초하여 상기 하나 이상의 개체가 분류된 상기 복수의 유형 각각의 입도 점수(granularity score)를 산출하고, 상기 복수의 유형 중 상기 산출된 입도 점수가 기 설정된 값 이하인 유형을 상기 확장이 필요한 유형으로 결정할 수 있다.
- [0009] 상기 지식 그래프 스키마 확장 방법은, 상기 하나 이상의 신규 유형에 기초하여 상기 텍스트 데이터에 대한 상기 분류 결과를 갱신하는 동작 및 상기 텍스트 데이터, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 학습 데이터로 이용하여 상기 유형 분류 모델을 재학습시키는 동작을 더 포함할 수 있다.
- [0010] 상기 갱신하는 동작은, 상기 하나 이상의 개체 중 상기 확장이 필요한 유형으로 분류된 개체 중 일부가 상기 신규 유형으로 분류되도록 상기 분류 결과를 갱신할 수 있다.
- [0011] 상기 재학습시키는 동작은, 상기 텍스트 데이터를 상기 유형 분류 모델의 입력 데이터로 이용하고, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 상기 유형 분류 모델의 타겟 데이터로 이용하여 상기 유형 분류 모델을 재학습시킬 수 있다.
- [0012] 상기 지식 그래프 스키마 확장 방법은, 상기 재학습된 유형 분류 모델을 이용하여 상기 텍스트 데이터에 포함된 하나 이상의 개체를 각각 상기 확장된 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하는 동작 및 상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 사용자에게 상기 신규 유형에 대한 오류 정보를 제공하는 동작을 더 포함할 수 있다.
- [0013] 상기 오류 정보를 제공하는 동작은, 상기 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 시드 스키마의 엔트로피를 산출하고, 상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 확장된 스키마의 엔트로피를 산출하며, 상기 시드 스키마의 엔트로피 및 상기 확장된 스키마의 엔트로피에 기초하여 상기 오류 정보를 제공할 수 있다.
- [0014] 상기 오류 정보를 제공하는 동작은, 상기 확장된 스키마의 엔트로피가 상기 시드 스키마의 엔트로피보다 높은 경우, 상기 오류 정보를 제공할 수 있다.
- [0015] 일 실시예에 따른 지식 그래프 스키마 확장 장치는, 하나 이상의 명령어를 저장하는 메모리 및 상기 하나 이상

의 명령어를 실행하는 하나 이상의 프로세서를 포함하고, 상기 하나 이상의 프로세서는, 사전 학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하고, 상기 분류 결과에 기초하여 사용자에게 상기 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공하고, 상기 사용자에 의해 상기 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우, 상기 하나 이상의 신규 유형을 포함하도록 상기 시드 스키마를 확장하여 확장된 스키마를 생성한다.

- [0016] 상기 시드 스키마에 포함된 복수의 유형은, 복수의 개체 유형을 포함할 수 있다.
- [0017] 상기 시드 스키마에 포함된 복수의 유형은, 하나 이상의 관계 유형을 더 포함할 수 있다.
- [0018] 상기 하나 이상의 프로세서는, 상기 분류 결과에 기초하여 상기 하나 이상의 개체가 분류된 상기 복수의 유형 각각의 입도 점수(granularity score)를 산출하고, 상기 복수의 유형 중 상기 산출된 입도 점수가 기 설정된 값 이하인 유형을 상기 확장이 필요한 유형으로 결정할 수 있다.
- [0019] 상기 하나 이상의 프로세서는, 상기 하나 이상의 신규 유형에 기초하여 상기 텍스트 데이터에 대한 상기 분류 결과를 갱신하고, 상기 텍스트 데이터, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 학습 데이터로 이용하여 상기 유형 분류 모델을 재학습시킬 수 있다.
- [0020] 상기 하나 이상의 프로세서는, 상기 하나 이상의 개체 중 상기 확장이 필요한 유형으로 분류된 개체 중 일부가 상기 신규 유형으로 분류되도록 상기 분류 결과를 갱신할 수 있다.
- [0021] 상기 하나 이상의 프로세서는, 상기 텍스트 데이터를 상기 유형 분류 모델의 입력 데이터로 이용하고, 상기 갱신된 분류 결과 및 상기 확장된 스키마를 상기 유형 분류 모델의 타겟 데이터로 이용하여 상기 유형 분류 모델을 재학습시킬 수 있다.
- [0022] 상기 하나 이상의 프로세서는, 상기 재학습된 유형 분류 모델을 이용하여 상기 텍스트 데이터에 포함된 하나 이상의 개체를 각각 상기 확장된 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류하고, 상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 사용자에게 상기 신규 유형에 대한 오류 정보를 제공할 수 있다.
- [0023] 상기 하나 이상의 프로세서는, 상기 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 시드 스키마의 엔트로피를 산출하고, 상기 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 상기 하나 이상의 개체에 대한 상기 확장된 스키마의 엔트로피를 산출하며, 상기 시드 스키마의 엔트로피 및 상기 확장된 스키마의 엔트로피에 기초하여 상기 오류 정보를 제공할 수 있다.
- [0024] 상기 하나 이상의 프로세서는, 상기 확장된 스키마의 엔트로피가 상기 시드 스키마의 엔트로피보다 높은 경우, 상기 오류 정보를 제공할 수 있다.

### 발명의 효과

- [0025] 개시되는 실시예들에 따르면, 유형 분류 모델에 의한 분류 결과에 기초하여 시드 스키마에 포함된 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공함으로써, 지식 그래프 스키마를 생성하는데 소요되는 시간과 비용을 줄이면서 도메인에 독립적인 시드 스키마로부터 특정 도메인의 지식이 반영된 확장된 스키마를 생성할 수 있다.
- [0026] 또한, 개시되는 실시예들에 따르면, 시드 스키마에 포함된 복수의 유형 각각에 대한 입도 점수에 기초하여 확장이 필요한 유형에 대한 정보를 제공함으로써, 사용자의 수고 없이 자동으로 확장이 필요한 유형을 결정할 수 있다.
- [0027] 또한, 개시되는 실시예들에 따르면, 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 사용자에 의해 입력된 신규 유형에 대한 오류 정보를 제공함으로써, 유형 분류 모델의 유형 분류 정확도를 높이고 확장된 스키마의 안정성을 높일 수 있다.

### 도면의 간단한 설명

- [0028] 도 1은 일 실시예에 따른 지식 그래프 스키마 확장 장치의 구성도이다.
- 도 2는 일 실시예에 따른 시드 스키마를 나타낸 도면이다.
- 도 3은 일 실시예에 따른 확장된 스키마를 나타낸 도면이다.

도 4는 추가적 실시예에 따른 지식 그래프 스키마 확장 장치의 구성도이다.

도 5는 일 실시예에 따른 오류 정보를 제공하는 예를 나타낸 도면이다.

도 6은 일 실시예에 따른 지식 그래프 스키마 확장 방법의 순서도이다.

도 7은 추가적 실시예에 따른 지식 그래프 스키마 확장 방법의 순서도이다.

도 8은 일 실시예에 따른 신규 유형에 대한 오류 정보를 제공하는 방법의 순서도이다.

도 9는 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도이다.

### 발명을 실시하기 위한 구체적인 내용

- [0029] 이하, 도면을 참조하여 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 이에 제한되지 않는다.
- [0030] 실시예들을 설명함에 있어서, 관련된 공지기술에 대한 구체적인 설명이 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 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. 상세한 설명에서 사용되는 용어는 단지 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 또한, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.
- [0031] 도 1은 일 실시예에 따른 지식 그래프 스키마(Knowledge graph schema) 확장 장치(100)의 구성도이다.
- [0032] 도 1을 참조하면, 일 실시예에 따른 지식 그래프 스키마 확장 장치(100)는 유형 분류부(110), 확장 유형 정보 제공부(120), 확장된 스키마 생성부(130)를 포함한다.
- [0033] 한편, 일 실시예에서, 유형 분류부(110), 확장 유형 정보 제공부(120), 확장된 스키마 생성부(130)는 물리적으로 구분된 하나 이상의 장치를 이용하여 구현되거나, 하나 이상의 프로세서 또는 하나 이상의 프로세서 및 소프트웨어의 결합에 의해 구현될 수 있으며, 도시된 예와 달리 구체적인 동작에 있어 명확히 구분되지 않을 수 있다.
- [0034] 도 1을 참조하면, 유형 분류부(110)는 사전 학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류한다.
- [0035] 이때, 텍스트 데이터는 예를 들어, 신문 기사, 뉴스, 소설 등과 같은 다양한 형태의 비정형 자연어 문서들을 포함할 수 있다. 텍스트 데이터는 자연어에 대한 하나 이상의 개체를 포함할 수 있다.
- [0036] 시드 스키마는 지식 그래프의 도메인의 종류에 관계 없이 다양한 도메인에 적용될 수 있는 스키마를 의미할 수 있다. 이때, 시드 스키마는 텍스트 데이터에 포함된 하나 이상의 개체를 분류하기 위한 복수의 의미 유형(semantic type)을 포함할 수 있다.
- [0037] 일 실시예에 따르면, 시드 스키마에 포함된 복수의 유형은 복수의 개체 유형을 포함할 수 있다. 이때, 개체 유형은 개체의 특성이나 상태를 나타내는 정보일 수 있다. 예를 들어, 텍스트 데이터에 'Donald Trump'라는 개체가 포함된 것으로 가정한다. 이때, 'Donald Trump' 개체는 인간 또는 정치인과 관련된 개체이므로, 시드 스키마에 포함된 복수의 유형 중 'Person' 또는 'Politician'라는 유형으로 분류될 수 있다. 이때, 'Person' 유형 또는 'Politician' 유형이 개체 유형에 해당될 수 있다.
- [0038] 또한, 시드 스키마에 포함된 복수의 유형은 하나 이상의 관계 유형을 더 포함할 수 있다. 이때, 관계 유형은 개체 쌍 사이의 관계 즉, 개체 쌍 사이의 연관성을 나타내는 정보일 수 있다. 예를 들어, 'Donald Trump is president of USA'라는 텍스트 데이터가 있다고 가정한다. 이때, 'is president of'는 'Donald Trump' 개체와 'USA' 개체 사이의 관계를 나타내는 것으로, 'Donald Trump' 개체 및 'USA' 개체 사이의 관계 유형에 해당될 수 있다.



- [0039] 한편, 개시되는 실시예에서 개체 유형 및 관계 유형은 종류만 다를 뿐, 지식 그래프 스키마 확장 장치(100)의 각 구성이 동작함에 있어서 동일하게 적용될 수 있다. 따라서, 이하의 실시예들은 개체 유형을 위주로 설명하였으나, 해당 실시예들은 관계 유형에도 동일하게 적용될 수 있다.
- [0040] 유형 분류 모델은 예를 들어, 합성곱 신경망(Convolution Neural Network, CNN) 기반의 모델일 수 있다. 일 실시예에서, 유형 분류 모델은 텍스트 데이터, 텍스트 데이터에 포함된 하나 이상의 개체 각각에 기 할당된 유형에 대한 데이터 및 시드 스키마를 포함하는 학습 데이터에 기초하여 사전 학습될 수 있다. 구체적으로, 유형 분류 모델은 텍스트 데이터를 입력 데이터로 이용하고, 기 할당된 유형에 대한 데이터 및 시드 스키마를 타겟 데이터로 이용하여 사전 학습될 수 있다.
- [0041] 또한, 유형 분류 모델은 예를 들어, 주의 집중 메커니즘(attention mechanism)이 적용된 주의 집중 기반 합성곱 신경망(Attention based Convolution Neural Network) 모델일 수 있다. 구체적으로, 유형 분류 모델은 컨벌루션 층(convolution layer)의 출력에서의 주의 집중 가중치(attention weight)를 결정할 수 있다. 구체적으로, 유형 분류 모델은 시드 스키마의 계층적 정보(hierarchical information)를 임베딩(embedding)하여 주의 집중 가중치를 산출할 수 있다. 또한, 후술할 바와 같이 유형 분류 모델을 재학습시키는 경우에는 확장된 스키마의 계층적 정보를 임베딩하여 주의 집중 가중치를 산출할 수 있다. 이에 따라, 유형 분류 모델은 훈련되는 파라미터의 수를 줄일 수 있고, 과적합(overfitting)에 덜 취약할 수 있다.
- [0042] 확장 유형 정보 제공부(120)는 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 사용자에게 시드 스키마에 포함된 복수의 유형 중 확장이 필요한 유형에 대한 정보를 제공한다.
- [0043] 구체적으로, 확장 유형 정보 제공부(120)는 시드 스키마에 포함된 복수의 유형 각각으로 분류된 개체에 기초하여 복수의 유형 각각에 대한 확장이 필요한지 여부를 판단할 수 있다.
- [0044] 예를 들어, 확장 유형 정보 제공부(120)는 복수의 유형 중 특정 유형으로 분류된 개체의 수가 사용자가 설정한 기준보다 많은 경우, 해당 유형에 대해 확장이 필요한 것으로 판단할 수 있다. 이 경우, 확장 유형 정보 제공부(120)는 사용자에게 해당 유형을 확장할 필요성이 있으며, 이에 따라 해당 유형을 확장하기 위해 시드 스키마에 신규 유형을 추가할 것인지에 대한 정보를 제공할 수 있다.
- [0045] 일 실시예에 따르면, 확장 유형 정보 제공부(120)는 분류 결과에 기초하여 하나 이상의 개체가 분류된 복수의 유형 각각의 입도 점수(granularity score)를 산출하고, 복수의 유형 중 산출된 입도 점수가 기 설정된 값 이하인 유형을 확장이 필요한 유형으로 결정할 수 있다.
- [0046] 구체적으로, 확장 유형 정보 제공부(120)는 사전 학습된 유형 분류 모델을 이용한 텍스트 데이터에 대한 분류 결과에 기초하여 시드 스키마에 포함된 복수의 유형 각각으로 분류된 개체들에 대한 정보를 획득할 수 있다.
- [0047] 예를 들어, 확장 유형 정보 제공부(120)는 텍스트 데이터에 대한 분류 결과로서 텍스트 데이터에 포함된 하나 이상의 개체 각각이 시드 스키마에 포함된 복수의 유형 각각으로 분류될 확률을 획득할 수 있다. 이때, 확장 유형 정보 제공부(120)는 하나 이상의 개체 각각이 복수의 유형 각각으로 분류될 확률에 기초하여 시드 스키마에 포함된 복수의 유형 각각으로 분류된 개체들에 대한 정보를 획득할 수 있다.
- [0048] 또한, 확장 유형 정보 제공부(120)는 사전 학습된 유형 예측 모델에 의해 하나 이상의 개체가 분류된 복수의 유형 중 확장이 필요한 유형을 판단하기 위한 입도 점수를 산출할 수 있다. 이때, 입도 점수는 시드 스키마에 포함된 복수의 유형에 대해 확장이 필요한지 여부를 판단하기 위한 수치일 수 있다. 예를 들어, 확장 유형 정보 제공부(120)는 산출된 복수의 유형 각각에 대한 입도 점수에 기초하여 복수의 유형 각각으로 분류된 개체의 수가 사용자에게 의해 설정된 기준보다 많은지 판단할 수 있다.
- [0049] 이때, 입도 점수는 아래 수학적 식 1을 통해 산출할 수 있다.

### 수학적 식 1

$$G(t_i) = - \frac{\sum_{j=1}^k \log(P(t_i|e_j))}{k}$$

[0050]

[0051] 수학적 식 1에서  $G$  는 입도 점수,  $t_i$  는 시드 스키마에 포함된 복수의 유형 중 하나,  $e_j$  는 텍스트 데이터에 포함

된 하나 이상의 개체 중 하나,  $k$  는 텍스트 데이터에 포함된 개체의 수,  $P(t_i|e_j)$  는 개체  $e_j$  가 유형  $t_i$  으로 분류될 확률을 의미한다.

- [0052] 한편, 확장 유형 정보 제공부(120)는 복수의 유형 중 산출된 입도 점수가 사용자에게 의해 기 설정된 값 이하인 유형이 존재하는 경우, 해당 유형으로 분류된 개체의 수가 많은 것으로 판단할 수 있다. 따라서, 확장 유형 정보 제공부(120)는 해당 유형을 확장이 필요한 유형으로 결정하여 동일 계층의 유형을 추가하거나 하위 계층의 유형을 추가할 필요가 있는 것으로 판단할 수 있다. 이때, 확장 유형 정보 제공부(120)는 결정된 유형에 대한 정보와 함께 결정된 유형을 확장할 것인지에 대한 질문 메시지를 사용자에게 제공할 수 있다.
- [0053] 확장된 스키마 생성부(130)는 사용자에게 의해 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우, 하나 이상의 신규 유형을 포함하도록 시드 스키마를 확장하여 확장된 스키마를 생성한다.
- [0054] 구체적으로, 확장된 스키마 생성부(130)는 확장이 필요한 유형에 대한 정보를 제공받은 사용자로부터 시드 스키마에 추가할 하나 이상의 신규 유형을 입력받을 수 있다. 또한, 확장된 스키마 생성부(130)는 시드 스키마에 포함된 복수의 유형에 신규 유형이 추가된 확장된 스키마를 생성할 수 있다.
- [0055] 도 2는 일 실시예에 따른 시드 스키마를 나타낸 도면이다. 또한, 도 3은 일 실시예에 따른 확장된 스키마를 나타낸 도면이다.
- [0056] 도 2 및 3을 참조하면, 도시된 시드 스키마에서 좌측에 있는 유형일수록 상위 유형에 해당하고, 우측에 있는 유형일수록 하위 유형인 것으로 가정한다. 예를 들어, 시드 스키마에 포함된 'Object' 유형(201)은 'Agent' 유형(202)의 상위 유형이다. 또한, 유형 분류 모델은 특정 도메인에 대한 텍스트 데이터 및 시드 스키마를 포함하는 학습 데이터에 기초하여 사전 학습된 것으로 가정한다.
- [0057] 지식 그래프 스키마 확장 장치(100)는 시드 스키마에 포함된 복수의 유형 각각에 대한 입도 점수를 산출할 수 있다. 이때, 복수의 유형 중 산출된 입도 점수가 기 설정된 값 이하인 유형이 존재하는 경우, 지식 그래프 스키마 확장 장치(100)는 해당 유형을 확장이 필요한 유형으로 결정할 수 있다. 예를 들어, 'Object' 유형(201) 및 'Organization' 유형(203) 각각에 대한 입도 점수가 기 설정된 값 이하인 것으로 판단되면, 지식 그래프 스키마 확장 장치(100)는 사용자에게 'Object' 유형(201) 및 'Organization' 유형(203)에 대해 확장할 필요성이 있다는 정보를 제공할 수 있다. 동시에, 지식 그래프 스키마 확장 장치(100)는 사용자에게 'Object' 유형(201) 및 'Organization' 유형(203)을 확장하기 위한 신규 유형의 추가 여부를 묻기 위한 정보를 제공할 수 있다.
- [0058] 이때, 사용자는 제공받은 정보에 응답하여 'Object' 유형(201) 및 'Organization' 유형(203)을 각각 확장하기 위하여 신규 유형인 'Award' 유형(301) 및 'Sports League' 유형(302)을 추가할 수 있다. 지식 그래프 스키마 확장 장치(100)는 사용자에게 의해 추가된 신규 유형에 기초하여 시드 스키마를 확장할 수 있다. 예를 들어, 지식 그래프 스키마 확장 장치(100)는 'Award' 유형(301) 및 'Sports League' 유형(302)을 각각 'Object' 유형(201) 및 'Organization' 유형(203)의 하위 계층으로 추가할 수 있다. 이 경우, 확장된 스키마는 시드 스키마에 포함된 복수의 유형에 'Award' 유형(301) 및 'Sports League' 유형(302)이 더 추가된 것일 수 있다.
- [0059] 한편, 상술한 예에서는 확장된 스키마를 생성하는 것과 관련하여, 신규 유형을 확장이 필요한 유형의 하위 계층으로 추가한 것으로 설명하였으나, 반드시 이에 한정되는 것은 아니고, 실시예에 따라 신규 유형을 확장이 필요한 유형과 동일 계층으로 추가할 수도 있다.
- [0060] 도 4는 추가적 실시예에 따른 지식 그래프 스키마 확장 장치(400)의 구성도이다.
- [0061] 도 4를 참조하면, 지식 그래프 스키마 확장 장치(400)는 분류 결과 갱신부(410), 모델 재학습부(420) 및 오류 정보 제공부(430)를 더 포함할 수 있다. 도 4에 도시된 예에서 유형 분류부(110), 확장 유형 정보 제공부(120), 확장된 스키마 생성부(130)는 도 1에 도시된 구성과 동일한 구성이므로 이에 대한 중복적인 설명은 생략한다.
- [0062] 한편, 일 실시예에서, 유형 분류부(110), 확장 유형 정보 제공부(120), 확장된 스키마 생성부(130), 분류 결과 갱신부(410), 모델 재학습부(420) 및 오류 정보 제공부(430)는 물리적으로 구분된 하나 이상의 장치를 이용하여 구현되거나, 하나 이상의 프로세서 또는 하나 이상의 프로세서 및 소프트웨어의 결합에 의해 구현될 수 있으며, 도시된 예와 달리 구체적인 동작에 있어 명확히 구분되지 않을 수 있다.
- [0063] 분류 결과 갱신부(410)는 하나 이상의 신규 유형에 기초하여 텍스트 데이터에 대한 분류 결과를 갱신한다.
- [0064] 일 실시예에 따르면, 분류 결과 갱신부(410)는 하나 이상의 개체 중 유형 예측 모델에 의해 확장이 필요한 유형

으로 분류된 개체 중 일부가 신규 유형으로 분류되도록 분류 결과를 갱신할 수 있다.

- [0065] 예를 들어, 사전 학습된 유형 분류 모델에 의해 텍스트 데이터에 포함된 A 개체 및 B 개체가 시드 스키마에 포함된 복수의 유형 중 X 유형으로 분류된 것으로 가정한다. 이때, X 유형이 확장이 필요한 유형으로 결정되어 사용자에게 의해 Y 유형이 추가된 경우, 분류 결과 갱신부(410)는 A 개체 및 B 개체 중 A 개체가 Y 유형으로 분류되도록 텍스트 데이터에 대한 분류 결과를 갱신할 수 있다.
- [0066] 한편, 분류 결과 갱신부(410)가 사용자에게 의해 추가된 신규 유형으로 분류되도록 분류 결과를 갱신하기 위해 유형 예측 모델에 의해 확장이 필요한 유형으로 분류된 개체 중 일부를 결정하는 방식은 복수의 개체 중 일부를 랜덤하게 결정하거나 사용자의 설정에 의해 결정하는 것일 수 있다.
- [0067] 모델 재학습부(420)는 텍스트 데이터, 갱신된 분류 결과 및 확장된 스키마를 학습 데이터로 이용하여 유형 예측 모델을 재학습시킨다.
- [0068] 이때, 갱신된 분류 결과에 기초하여 모델 재학습부(420) 또는 사용자에게 의해 텍스트 데이터에 포함된 하나 이상의 개체에 각각 유형이 할당될 수 있다. 갱신된 분류 결과에 기초하여 유형이 할당된 하나 이상의 개체를 포함하는 텍스트 데이터는 유형 분류 모델을 재학습시키기 위한 학습 데이터로 이용될 수 있다.
- [0069] 일 실시예에 따르면, 모델 재학습부(420)는 텍스트 데이터를 유형 예측 모델의 입력 데이터로 이용하고, 갱신된 분류 결과 및 확장된 스키마를 유형 예측 모델의 타겟 데이터로 이용하여 유형 예측 모델을 재학습시킬 수 있다.
- [0070] 구체적으로, 모델 재학습부(420)는 텍스트 데이터를 유형 예측 모델의 입력 데이터로 이용하고, 갱신된 분류 결과에 기초하여 텍스트 데이터에 포함된 하나 이상의 개체 각각에 기 할당된 유형에 대한 데이터 및 확장된 스키마를 유형 예측 모델의 타겟 데이터로 이용하여 유형 예측 모델을 재학습시킬 수 있다.
- [0071] 또한, 일 실시예에 따르면, 유형 분류부(110)는 재학습된 유형 분류 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 확장된 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류할 수 있다. 이때, 확장 유형 정보 제공부(130)는 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 확장된 스키마를 다시 확장할 필요가 있는지 여부를 판단하고, 판단 결과에 따라 확장된 스키마를 다시 확장하는 과정을 반복하여 수행할 수 있다.
- [0072] 오류 정보 제공부(430)는 재학습된 유형 예측 모델에 의한 분류 결과에 기초하여 사용자에게 신규 유형에 대한 오류 정보를 제공할 수 있다.
- [0073] 구체적으로, 오류 정보 제공부(430)는 사전 학습된 유형 분류 모델을 이용한 텍스트 데이터에 대한 분류 결과 및 재학습된 유형 분류 모델을 이용한 텍스트 데이터에 대한 분류 결과에 기초하여 사전 학습된 유형 분류 모델 및 재학습된 유형 예측 모델 각각의 유형 분류 안정성을 판단할 수 있다. 또한, 오류 정보 제공부(430)는 재학습된 유형 예측 모델의 유형 분류 안정성이 사전 학습된 유형 분류 모델의 유형 분류 안정성보다 낮으면, 사용자에게 신규 유형에 대한 오류 정보를 제공할 수 있다.
- [0074] 이때, 유형 분류 안정성은 스키마에 포함된 복수의 유형 중 개체가 분류될 유형이 분명한지를 나타내는 정보일 수 있다. 즉, 개체가 스키마에 포함된 복수의 유형 중 특정 유형으로 분류될 확률이 복수의 유형 중 해당 유형을 제외한 나머지 유형으로 분류될 확률보다 높을수록 유형 분류 안정성이 높은 것으로 판단할 수 있다.
- [0075] 일 실시예에 따르면, 오류 정보 제공부(430)는 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 하나 이상의 개체에 대한 시드 스키마의 엔트로피를 산출하고, 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 하나 이상의 개체에 대한 확장된 스키마의 엔트로피를 산출하며, 시드 스키마의 엔트로피 및 확장된 스키마의 엔트로피에 기초하여 오류 정보를 제공할 수 있다.
- [0076] 구체적으로, 사전 학습된 유형 분류 모델 및 재학습된 유형 예측 모델 각각의 유형 분류 안정성은 하나 이상의 개체가 분류된 복수의 유형을 포함하는 시드 스키마의 엔트로피(entropy) 및 하나 이상의 개체가 분류된 복수의 유형을 포함하는 확장된 스키마의 엔트로피에 기초하여 판단할 수 있다. 이때, 엔트로피는 텍스트 데이터에 포함된 하나 이상의 개체가 유형 예측 모델에 의해 스키마에 포함된 복수의 유형 각각으로 분류될 확률과 관련된 정보일 수 있다.
- [0077] 이때, 엔트로피는 아래의 수학적 식 2에 의해 산출될 수 있다.

## 수학식 2

$$E = - \sum_i \sum_j P(t_j | e_i) * \log_2 P(t_j | e_i)$$

[0078]

[0079]

수학식 2에서  $E$  는 엔트로피를 의미한다.

[0080]

시드 스키마의 엔트로피는 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 텍스트 데이터에 포함된 하나 이상의 개체가 각각 시드 스키마에 포함된 복수의 유형 각각으로 분류될 확률과 관련된 정보일 수 있다. 또한, 확장된 스키마의 엔트로피는 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 텍스트 데이터에 포함된 하나 이상의 개체가 각각 확장된 스키마에 포함된 복수의 유형 각각으로 분류될 확률과 관련된 정보일 수 있다.

[0081]

또한, 오류 정보 제공부(430)는 확장된 스키마의 엔트로피가 시드 스키마의 엔트로피보다 높은 경우, 오류 정보를 제공할 수 있다.

[0082]

예를 들어, 확장된 스키마의 엔트로피가 시드 스키마의 엔트로피보다 높은 경우, 신규 유형이 추가된 확장된 스키마에 의해 재학습된 유형 예측 모델의 유형 분류 안정성이 낮아진 것으로 판단할 수 있다. 이때, 오류 정보 제공부(430)는 사용자에게 신규 유형에 오류가 있다는 정보를 제공할 수 있다. 또한, 오류 정보 제공부(430)는 사용자로부터 신규 유형을 확장된 스키마에서 삭제하거나 다른 신규 유형을 입력받을 수 있다.

[0083]

다른 예로, 오류 정보 제공부(430)는 확장된 스키마의 엔트로피가 시드 스키마의 엔트로피보다 낮거나 동일한 경우, 신규 유형이 추가된 확장된 스키마에 의해 재학습된 유형 예측 모델의 유형 분류 안정성이 높아진 것으로 판단할 수 있다. 이 경우, 사용자에게 오류 정보를 제공하지 않고, 상술한 스키마를 확장하는 과정을 반복하여 수행함으로써 특정 도메인에 대한 지식 그래프 스키마를 확장할 수 있다.

[0084]

도 5는 일 실시예에 따른 오류 정보를 제공하는 예를 나타낸 도면이다.

[0085]

도 5를 참조하면, 확장된 스키마는 시드 스키마에서 'Person' 유형(501)의 하위 계층으로 'Sportsperson' 유형(502)을 추가한 것으로 가정한다. 또한, 사용자가 유형 예측 모델을 재학습시키기 위해 'Sportsperson' 유형(502)을 추가하면서 텍스트 데이터에 포함된 축구선수 이름과 관련된 'LionelMessi' 개체(504)가 'Sportsperson' 유형(502)으로 분류되도록 유형 예측 모델의 분류 결과를 갱신한 것으로 가정한다.

[0086]

지식 그래프 스키마 확장 장치(400)는 'LionelMessi' 개체(504)를 포함하는 하나 이상의 개체에 대한 시드 스키마의 엔트로피 및 'LionelMessi' 개체(504)를 포함하는 하나 이상의 개체에 대한 확장된 스키마의 엔트로피를 산출할 수 있다. 이때, 산출 결과에 의하면, 시드 스키마의 엔트로피는 확장된 스키마의 엔트로피보다 높을 수 있다. 왜냐하면, 'LionelMessi' 개체(504)는 선수와 관련된 유형인 'Athlete' 유형(503)이 존재함에도 불구하고, 사용자에 의해 새로 추가된 'Sportsperson' 유형(502)이 할당되었다. 따라서, 확장된 스키마에 포함된 복수의 유형에서 선수와 관련된 유형이 중복되므로, 'LionelMessi' 개체(504)는 재학습된 유형 분류 모델에 의해 확장된 스키마에 포함된 복수의 유형 중 하나로 분류될 확률이 현저하게 떨어지게 된다.

[0087]

이 경우, 지식 그래프 스키마 확장 장치(400)는 'Sportsperson' 유형(502)에 대해 오류가 있다는 메시지를 사용자에게 제공할 수 있다. 이에 따라, 사용자는 지식 그래프 스키마 확장 장치(400)는 확장된 스키마에서 'Sportsperson' 유형(502)을 삭제하거나 다른 신규 유형을 추가할 수 있다. 또한, 사용자는 'Sportsperson' 유형(502)을 유지하고, 'Athlete' 유형(503)을 삭제할 수도 있다.

[0088]

도 6은 일 실시예에 따른 지식 그래프 스키마 확장 방법의 순서도이다.

[0089]

도 6에 도시된 방법은 예를 들어, 도 1에 도시된 지식 그래프 스키마 확장 장치(100)에 의해 수행될 수 있다.

[0090]

도 6을 참조하면, 지식 그래프 스키마 확장 장치(100)는 사전 학습된 유형 예측 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 시드 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류한다(610).

[0091]

이후, 지식 그래프 스키마 확장 장치(100)는 분류 결과에 기초하여 사용자에게 복수의 유형 중 확장이 필요한



유형에 대한 정보를 제공한다(620).

- [0092] 이때, 지식 그래프 스키마 확장 장치(100)는 분류 결과에 기초하여 하나 이상의 개체가 분류된 복수의 유형 각각에 대한 입도 점수를 산출하고, 복수의 유형 중 산출된 입도 점수가 기 설정된 값 이하인 유형을 확장이 필요한 유형으로 결정할 수 있다.
- [0093] 이후, 지식 그래프 스키마 확장 장치(100)는 사용자에게 의해 시드 스키마에 추가할 하나 이상의 신규 유형이 입력된 경우(630), 하나 이상의 신규 유형을 포함하도록 시드 스키마를 확장하여 확장된 스키마를 생성한다(640).
- [0094] 도 7은 추가적 실시예에 따른 지식 그래프 스키마 확장 방법의 순서도이다.
- [0095] 도 7에 도시된 방법은 예를 들어, 도 4에 도시된 지식 그래프 스키마 확장 장치(400)에 의해 수행될 수 있다.
- [0096] 도 7에 도시된 예에서, 710 동작 내지 730 동작은 도 6에 도시된 610 동작 내지 630 동작과 동일하므로 이에 대한 중복적인 설명은 생략한다.
- [0097] 도 7을 참조하면, 지식 그래프 스키마 확장 장치(400)는 하나 이상의 신규 개체 유형에 기초하여 텍스트 데이터에 대한 분류 결과를 갱신한다(750).
- [0098] 이때, 지식 그래프 스키마 확장 장치(400)는 하나 이상의 개체 중 확장이 필요한 유형으로 분류된 개체 중 일부가 신규 유형으로 분류되도록 분류 결과를 갱신할 수 있다.
- [0099] 이후, 지식 그래프 스키마 확장 장치(400)는 텍스트 데이터, 갱신된 분류 결과 및 확장된 스키마를 학습 데이터로 이용하여 유형 예측 모델을 재학습시킨다(760).
- [0100] 이때, 지식 그래프 스키마 확장 장치(400)는 텍스트 데이터를 유형 예측 모델의 입력 데이터로 이용하고, 갱신된 분류 결과 및 확장된 스키마를 유형 예측 모델의 타겟 데이터로 이용하여 유형 예측 모델을 재학습시킬 수 있다.
- [0101] 도 8은 일 실시예에 따른 신규 유형에 대한 오류 정보를 제공하는 방법의 순서도이다.
- [0102] 도 8에 도시된 방법은 예를 들어, 도 4에 도시된 지식 그래프 스키마 확장 장치(400)에 의해 수행될 수 있다.
- [0103] 도 8을 참조하면, 지식 그래프 스키마 확장 장치(400)는 재학습된 유형 예측 모델을 이용하여 텍스트 데이터에 포함된 하나 이상의 개체를 각각 확장된 스키마에 포함된 복수의 유형 중 하나 이상의 유형으로 분류할 수 있다(810).
- [0104] 이후, 지식 그래프 스키마 확장 장치(400)는 재학습된 유형 예측 모델에 의한 분류 결과에 기초하여 사용자에게 신규 유형에 대한 오류 정보를 제공할 수 있다.
- [0105] 구체적으로, 지식 그래프 스키마 확장 장치(400)는 사전 학습된 유형 분류 모델에 의한 분류 결과에 기초하여 하나 이상의 개체에 대한 시드 스키마의 엔트로피를 산출하고, 재학습된 유형 분류 모델에 의한 분류 결과에 기초하여 하나 이상의 개체에 대한 상기 확장된 스키마의 엔트로피를 산출할 수 있다(820).
- [0106] 이후, 지식 그래프 스키마 확장 장치(400)는 확장된 스키마의 엔트로피가 시드 스키마의 엔트로피보다 높은 경우(830), 오류 정보를 제공할 수 있다(840).
- [0107] 도 9는 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경(10)을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술되지 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.
- [0108] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 도 1 및 4에 도시된 지식 그래프 스키마 확장 장치(100, 400)에 포함되는 하나 이상의 컴포넌트일 수 있다.
- [0109] 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0110] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다

른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.

[0111] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.

[0112] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와 연결될 수도 있다.

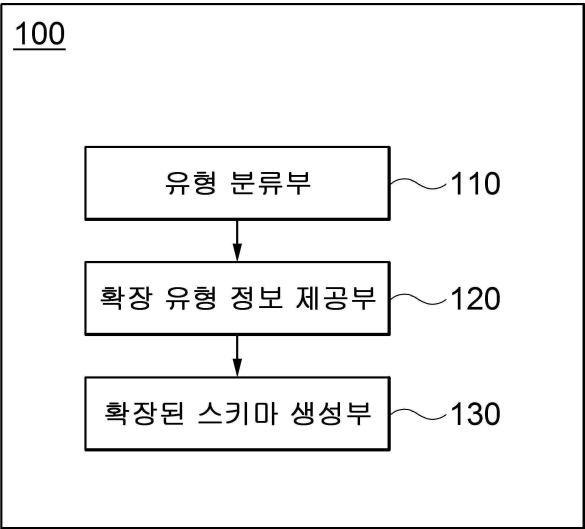
[0113] 이상에서는 실시예들을 중심으로 기술적 특징들을 설명하였다. 하지만, 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 하고, 권리 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 권리범위에 포함된 것으로 해석되어야 할 것이다.

### 부호의 설명

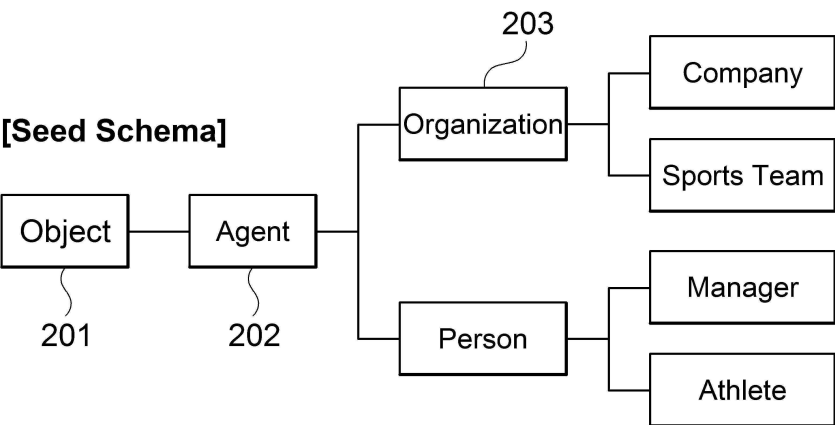
- [0114]
- 10: 컴퓨팅 환경
  - 12: 컴퓨팅 장치
  - 14: 프로세서
  - 16: 컴퓨터 판독 가능 저장 매체
  - 18: 통신 버스
  - 20: 프로그램
  - 22: 입출력 인터페이스
  - 24: 입출력 장치
  - 26: 네트워크 통신 인터페이스
  - 100, 400: 지식 그래프 스키마 확장 장치
  - 110: 유형 분류부
  - 120: 확장 유형 정보 제공부
  - 130: 확장된 스키마 생성부
  - 410: 분류 결과 갱신부
  - 420: 모델 재학습부
  - 430: 오류 정보 제공부

도면

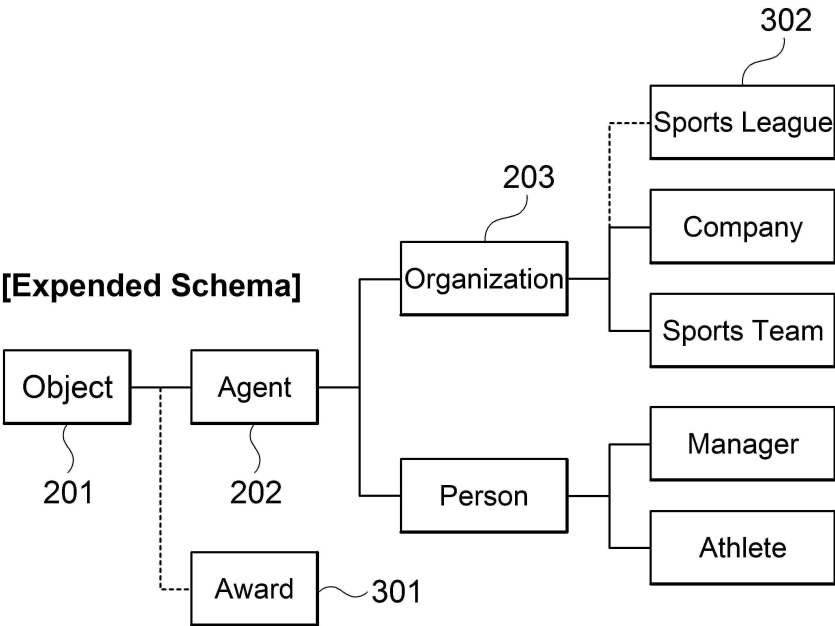
도면1



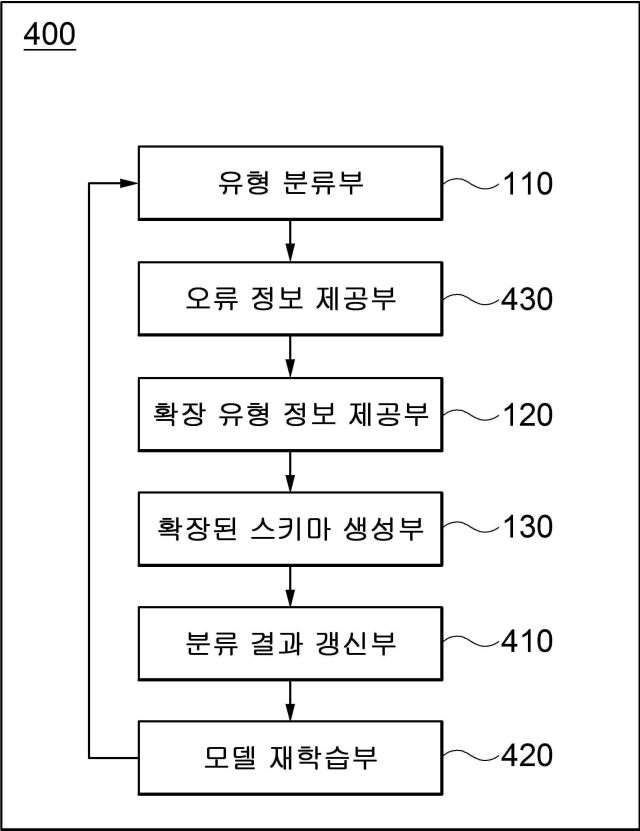
도면2



도면3

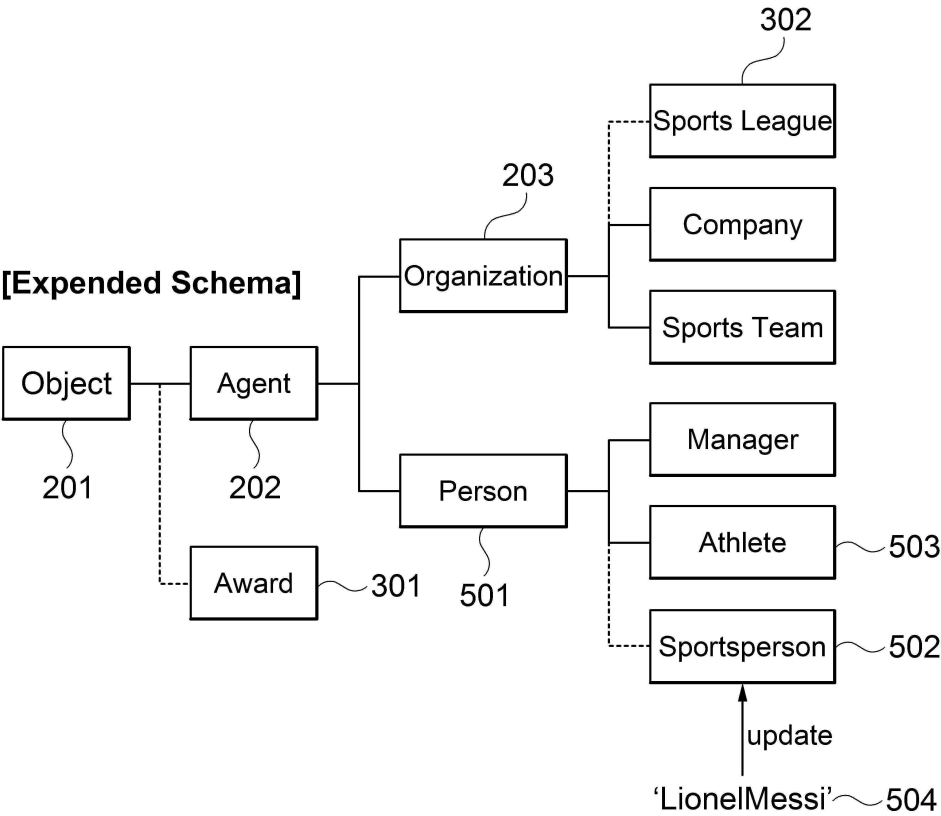


도면4

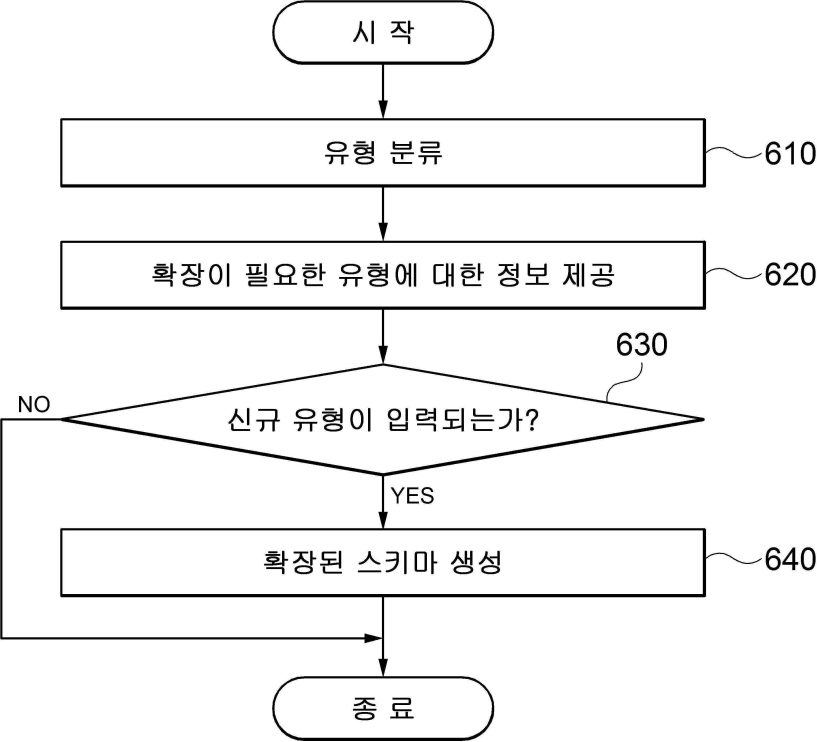




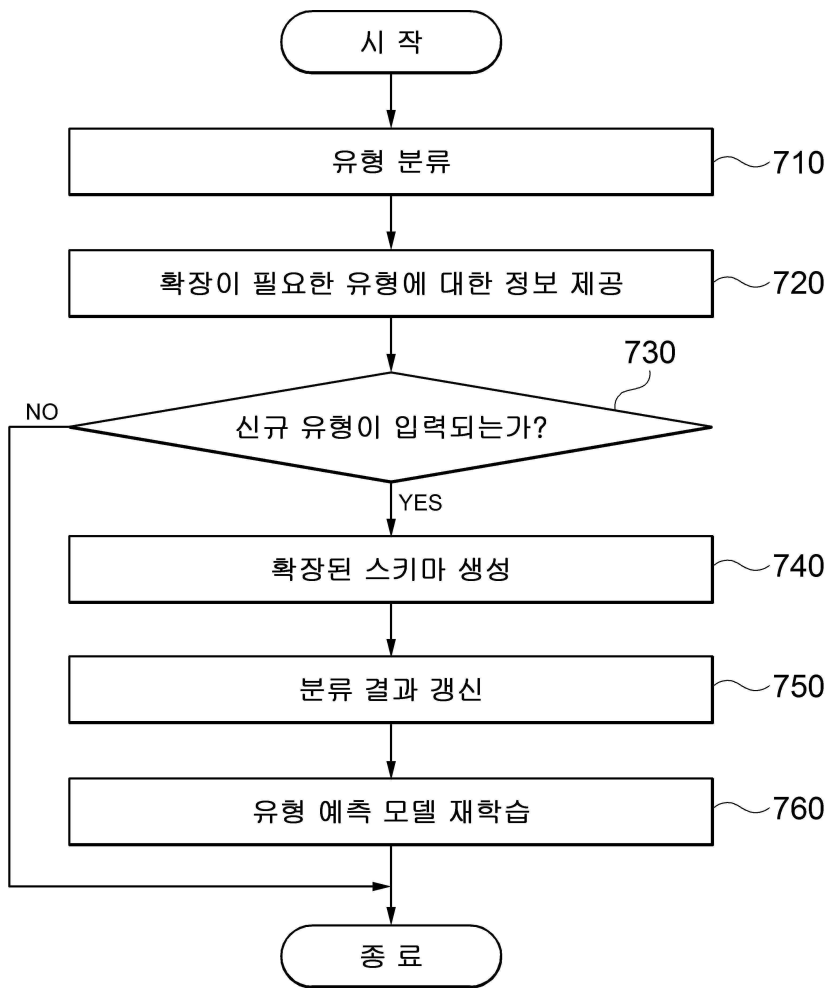
도면5



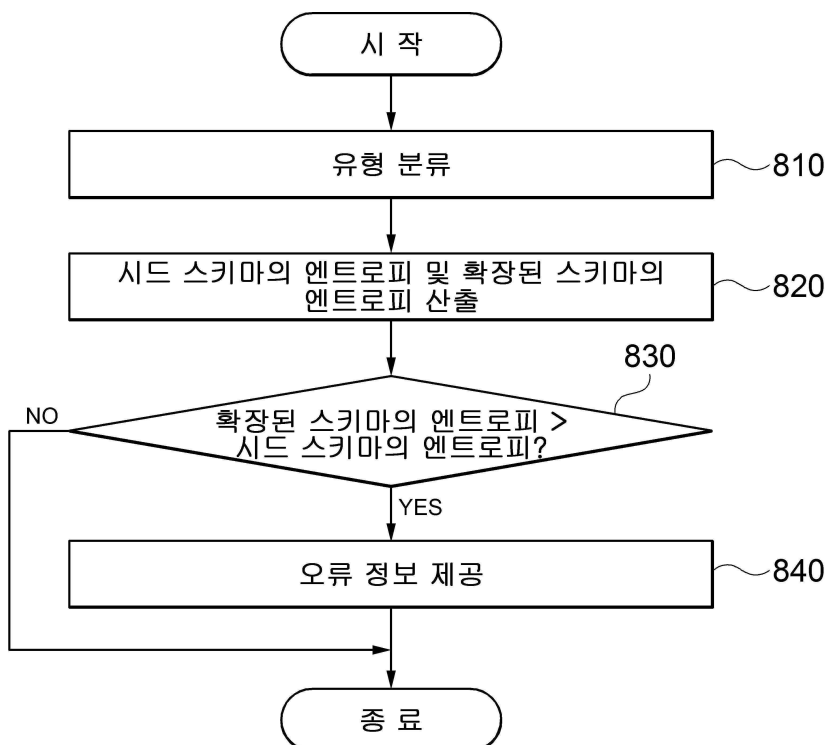
도면6



도면7



도면8



도면9

10

