



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0081594  
(43) 공개일자 2023년06월07일

- |   |  |
|---|--|
| <p>(51) 국제특허분류(Int. Cl.)<br/>G06F 16/242 (2019.01) G06F 16/21 (2019.01)<br/>G06F 16/28 (2019.01) G06N 3/08 (2023.01)</p> <p>(52) CPC특허분류<br/>G06F 16/243 (2019.01)<br/>G06F 16/211 (2019.01)</p> <p>(21) 출원번호 10-2022-0117575<br/>(22) 출원일자 2022년09월19일<br/>심사청구일자 2022년09월19일</p> <p>(30) 우선권주장<br/>1020210169562 2021년11월30일 대한민국(KR)</p> | <p>(71) 출원인<br/>포항공과대학교 산학협력단<br/>경상북도 포항시 남구 청암로 77 (지곡동)</p> <p>(72) 발명자<br/>한옥신<br/>경상북도 포항시 남구 청암로 77<br/>강혁규<br/>경상북도 포항시 남구 청암로 77<br/>김현지<br/>경상북도 포항시 남구 청암로 77</p> <p>(74) 대리인<br/>특허법인이상</p> |
|---|--|

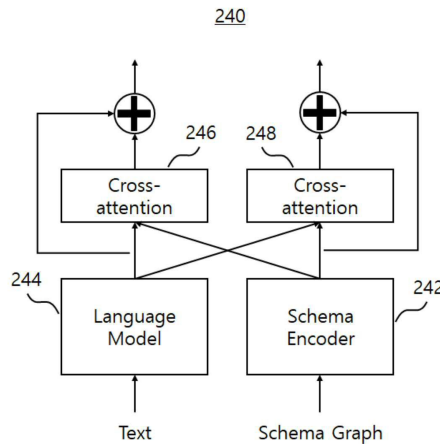
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **트랜스포머 신경망을 이용하여 관계형 데이터베이스에 대한 자연어 질의를 처리하는 장치 및 방법**

**(57) 요약**

본 발명의 목적을 달성하기 위한 일 실시예에 따른 자연어 질의 처리 장치는 사용자에게 의하여 입력된 자연어 질의를 수신하고, 자연어 질의에 기반하여 구조화된 질의를 생성하고, 프로세서는 자연어 질의에 기반하여 구조화된 질의를 생성함에 있어서, 자연어 질의에 대한 자연어 처리 결과 및 자연어 질의와 관련되는 데이터베이스 내의 하위 데이터베이스 간의 관계에 기반하여 추출되는 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 이용하여 구조화된 질의를 생성한다.

**대표도** - 도4



(52) CPC특허분류

*G06F 16/2423* (2019.01)

*G06F 16/284* (2019.01)

*G06N 3/08* (2023.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711126201
과제번호	2018-0-01398-004
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발(R&D, 정보화)
연구과제명	(SW 스타랩) 대화 가능하고 자동으로 튜닝하는 DBMS의 개발
기여율	1/1
과제수행기관명	포항공과대학교 산학협력단
연구기간	2021.01.01 ~ 2021.12.31

---

## 명세서

### 청구범위

#### 청구항 1

프로세서(processor);

를 포함하고,

상기 프로세서는

사용자에 의하여 입력된 자연어 질의를 수신하고,

상기 자연어 질의에 기반하여 구조화된 질의를 생성하고,

상기 프로세서는 상기 자연어 질의에 기반하여 상기 구조화된 질의를 생성함에 있어서,

상기 자연어 질의에 대한 자연어 처리 결과; 및

상기 자연어 질의와 관련된 데이터베이스 내의 하위 데이터베이스 간의 관계에 기반하여 추출되는 스키마 관계;

의 상호 간에 생성되는 교차 어텐션 결과를 이용하여 상기 구조화된 질의를 생성하는, 자연어 질의 처리 장치.

#### 청구항 2

제1항에 있어서,

상기 프로세서는 상기 자연어 처리 결과의 상기 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과, 및 상기 스키마 관계의 상기 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 이용하여 상기 구조화된 질의를 생성하고,

상기 교차 어텐션 결과는 상기 제1 교차 어텐션 결과 및 상기 제2 교차 어텐션 결과를 포함하는, 자연어 질의 처리 장치.

#### 청구항 3

제1항에 있어서,

상기 자연어 질의에 기반하여 상기 자연어 처리 결과를 생성하는 자연어 처리 모델; 및

상기 스키마 관계를 출력하는 스키마 인코더 모델;

을 더 포함하는, 자연어 질의 처리 장치.

#### 청구항 4

제3항에 있어서,

상기 스키마 인코더 모델은 입력 스키마 그래프에 기반하여 스키마 의미 표현을 상기 스키마 관계로서 출력하고,

상기 자연어 처리 모델은 상기 자연어 질의의 전처리 결과를 입력받고 자연어 요소 의미 표현을 상기 자연어 처리 결과로서 출력하는, 자연어 질의 처리 장치.

#### 청구항 5

제3항에 있어서,

상기 자연어 처리 결과의 상기 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과를 생성하는 제1 교차 어텐션 계층; 및

상기 스키마 관계의 상기 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 생성하는 제2 교차 어텐션 계층;

을 더 포함하고,

상기 교차 어텐션 결과는 상기 제1 교차 어텐션 결과 및 상기 제2 교차 어텐션 결과를 포함하는, 자연어 질의 처리 장치.

#### **청구항 6**

제3항에 있어서,

상기 자연어 처리 모델은 내부에 셀프 어텐션 기능을 가지는 제1 내부 어텐션 계층을 포함하고,

상기 스키마 인코더 모델은 내부에 셀프 어텐션 및 교차 어텐션 중 적어도 하나 이상의 기능을 가지는 제2 내부 어텐션 계층을 포함하는, 자연어 질의 처리 장치.

#### **청구항 7**

제6항에 있어서,

상기 프로세서는

상기 제1 내부 어텐션 계층의 출력에 기반한 상기 자연어 처리 모델 내부의 파라미터의 업데이트를 차단하는, 자연어 질의 처리 장치.

#### **청구항 8**

제3항에 있어서,

상기 프로세서는

상기 교차 어텐션 결과에 기반한 상기 자연어 처리 모델 내부의 파라미터의 업데이트를 차단하는, 자연어 질의 처리 장치.

#### **청구항 9**

제3항에 있어서,

상기 교차 어텐션 결과에 기반하여 상기 구조화된 질의를 생성하는 디코더 계층;

을 더 포함하는, 자연어 질의 처리 장치.

#### **청구항 10**

프로세서(processor)를 포함하는 컴퓨팅 시스템에 의하여 실행되는 자연어 질의 처리 방법에 있어서,

사용자에 의하여 입력되는 자연어 질의를 수신하는 단계; 및

상기 자연어 질의에 기반하여 구조화된 질의를 생성하는 단계;

를 포함하고,

상기 구조화된 질의를 생성하는 단계는,

상기 자연어 질의에 대한 자연어 처리 결과를 획득하는 단계;

상기 자연어 질의와 관련된 데이터베이스 내의 하위 데이터베이스 간의 관계에 기반하여 추출되는 스키마 관계를 획득하는 단계;

상기 자연어 처리 결과 및 상기 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 획득하는 단계; 및

상기 교차 어텐션 결과를 이용하여 상기 구조화된 질의를 생성하는 단계;

를 포함하는, 자연어 질의 처리 방법.

**청구항 11**

제10항에 있어서,  
 상기 교차 어텐션 결과를 획득하는 단계는,  
 상기 자연어 처리 결과의 상기 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과를 획득하는 단계; 및  
 상기 스키마 관계의 상기 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 획득하는 단계;  
 를 포함하는, 자연어 질의 처리 방법.

**청구항 12**

제10항에 있어서,  
 상기 스키마 관계를 획득하는 단계는,  
 입력 스키마 그래프에 기반하여 스키마 의미 표현을 상기 스키마 관계로서 획득하고,  
 상기 자연어 처리 결과를 획득하는 단계는,  
 상기 자연어 질의 전처리 결과를 입력받고 자연어 요소 의미 표현을 상기 자연어 처리 결과로서 획득하는, 자연어 질의 처리 방법.

**청구항 13**

제10항에 있어서,  
 상기 자연어 처리 결과를 획득하는 단계는,  
 내부 어텐션 계층의 출력에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트가 차단된 상태에서 수행되는, 자연어 질의 처리 방법.

**청구항 14**

제10항에 있어서,  
 상기 자연어 처리 결과를 획득하는 단계는,  
 상기 교차 어텐션 결과에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트가 차단된 상태에서 수행되는, 자연어 질의 처리 방법.

**청구항 15**

프로세서(processor)를 포함하는 컴퓨팅 시스템에 의하여 실행되는 자연어 질의 처리 방법에 있어서,  
 자연어 질의와 관련되는 데이터베이스 내의 하위 데이터베이스 간의 스키마 관계를 일대일(one-to-one) 대응, 및 일대다(one-to-many) 대응 중 어느 하나로서 추출하는 단계; 및  
 상기 스키마 관계에 기반하여 상기 자연어 질의로부터 구조화된 질의를 생성하는 단계;  
 를 포함하는, 자연어 질의 처리 방법.

**청구항 16**

제15항에 있어서,  
 상기 하위 데이터베이스 간의 비교를 통해 상기 스키마 관계를 강제적(mandatory) 관계, 및 선택적(optional) 관계 중 어느 하나로서 예측하는 단계;  
 를 더 포함하는 자연어 질의 처리 방법.

**청구항 17**

제15항에 있어서,

상기 하위 데이터베이스 중 제1 하위 데이터베이스에 대응하는 제1 자연어 모델 및 제2 하위 데이터베이스에 대응하는 제2 자연어 모델을 이용하여 상기 스키마 관계를 인코딩하는 단계;

를 더 포함하는 자연어 질의 처리 방법.

**청구항 18**

제15항에 있어서,

상기 하위 데이터베이스 중 일부에 대한 마스크 컬럼 모델링을 이용하여 상기 스키마 관계의 적어도 일부를 미리 인코딩하는 단계;

를 더 포함하는 자연어 질의 처리 방법.

**청구항 19**

제15항에 있어서,

상기 자연어 질의에 기반하여 자연어 처리 결과를 획득하는 단계;

를 더 포함하고,

상기 구조화된 질의를 생성하는 단계는,

상기 자연어 처리 결과 및 상기 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 획득하는 단계; 및

상기 교차 어텐션 결과를 이용하여 상기 구조화된 질의를 생성하는 단계;

를 포함하는, 자연어 질의 처리 방법.

**청구항 20**

제15항에 있어서,

미리 학습된 언어 모델을 이용하여 상기 자연어 질의에 기반하여 자연어 처리 결과를 획득하는 단계;

를 더 포함하고,

상기 자연어 처리 결과를 획득하는 단계는, 상기 언어 모델의 내부 파라미터의 업데이트가 차단된 상태에서 수행되는, 자연어 질의 처리 방법.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 데이터베이스에 대한 질의 처리 방법 및 장치에 관한 것으로, 자연어 질의에 기반하여 규격화/구조화된 질의를 생성하는 기술에 관한 것이다. 특히 본 발명은 트랜스포머 신경망과 관계형 데이터베이스를 이용하여 자연어 질의에 기반하여 규격화/구조화된 질의를 생성하는 기술에 관한 것이다.

**배경 기술**

[0002] 최근 인터넷의 보급으로 사용자는 대단히 방대한 정보에 노출되고 있다. 이러한 방대한 정보 중 자신에게 필요한 정보만을 획득하기 위해서는 효과적인 검색이 필요하다.

[0003] 검색은 질의에 대한 응답으로 이루어진다. 일반적인 질의 응답 시스템은 질문에 대한 결과로 해답을 제시한다. 정보 요구자가 원하는 내용을 정확히 질의하고 탐색하는 것은 매우 어려운 일이다.

[0004] 원하는 결과를 얻기 위해서는 질의가 규격을 갖춘 구조화된 질의일 것이 요구된다. 그러나 구조화된 질의를 생성하기 위해서는 규격/규칙에 대한 전문적인 지식을 요구하는 경우가 있으므로 대부분의 사용자는 구조화된 질의를 스스로 생성해 내기 쉽지 않은 것이 현실이다.

[0005] 때문에 사용자는 자연어로 이루어진 질의를 생성하고, 자연어 질의를 규격화/구조화된 질의로 해석/변환하려는 시도가 있어 왔다.

[0006] 한국등록특허 KR 10-2150908호 "자연어 질의해석 방법 및 시스템"은 자연어 질의를 해석함에 있어 사용자의 의도를 더욱 명확히 해석하고 사용자의 의도에 효과적으로 부합하는 질의 처리 방법을 제안하였다.

[0007] 그러나 이러한 선행기술에 의해서도 자연어 질의를 구조화된 질의로 변환하는 과정이 완전하지 않으며, 자연어 질의를 구조화된 질의로 변환하는 효과적인 기법이 요구되고 있다.

**선행기술문헌**

**특허문헌**

- [0008] (특허문헌 0001) 한국등록특허 KR 10-2150908호 "자연어 질의해석 방법 및 시스템" (공개일 2020년 8월 27일)
- (특허문헌 0002) 한국등록특허 KR 10-2277787호 "신경망 기반 자연어로부터 SQL 질의 번역 시 사용되는 컬럼 및 테이블을 예측하는 방법" (공개일 2021년 7월 9일)
- (특허문헌 0003) 한국등록특허 KR 10-2345568호 "자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법" (공개일 2021년 12월 27일)

**비특허문헌**

- [0009] (비특허문헌 0001) "Attention Is All You Need", Ashish Vaswani 외, 31st Conference on Neural Information Processing Systems (NIPS 2017), (공개일 2017년 12월 6일)

**발명의 내용**

**해결하려는 과제**

- [0010] 상기와 같은 문제점을 해결하기 위한 본 발명의 목적은, 데이터베이스 관리 시스템뿐만 아니라 향후 검색 시스템에도 적용이 예상되는 검색 시스템으로서, 사용자가 원하는 정보에 키워드 또는 질의를 보냈을 때 관련된 자료나 정답을 찾아주는 기술로, 비정형 데이터(unstructured data) 뿐만 아니라 데이터베이스와 같은 정형 데이터를 사용할 때 활용할 수 있는 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 표현(meaning representation)으로 인코딩하는 장치 및 방법을 제공하는 것이다.
- [0011] 본 발명은 사용자가 구조화된 질의에 대한 지식 없이 자연어 질의를 입력하는 경우에도 자동으로 구조화된 질의를 생성하여 사용자가 원하는 질의 응답 결과를 제공하는 것을 목적으로 한다.
- [0012] 본 발명은 사용자가 입력한 자연어 질의만으로는 얻을 수 있는 스키마 정보를 미리 분석하고 인코딩함으로써 사용자의 의도에 부합하는 구조화된 질의를 생성하고 질의 응답 결과를 제공하는 것을 목적으로 한다.
- [0013] 본 발명은 사용자가 자연어 질의를 입력하는 경우에 구조화된 질의 및 질의 응답 결과를 생성하는 과정을 고도화하는 것을 목적으로 한다.
- [0014] 본 발명은 자연어 질의 입력에 응답하여 구조화된 질의를 생성하는 과정을 고도화하는 교차 어텐션 기반의 새로운 신경망 구조를 제안하는 것을 목적으로 한다.
- [0015] 본 발명은 자연어 질의 입력에 응답하여 구조화된 질의를 생성하는 과정을 고도화하는, 하위 데이터베이스 간의 관계에 기반한 스키마 관계 추출 기법을 제안하는 것을 목적으로 한다.

**과제의 해결 수단**

- [0016] 본 발명의 목적을 달성하기 위한 일 실시예에 따른 자연어 질의 처리 장치는, 프로세서(processor)를 포함한다. 프로세서는 사용자에게 의하여 입력된 자연어 질의를 수신하고, 자연어 질의에 기반하여 구조화된 질의를 생성하고, 프로세서는 자연어 질의에 기반하여 구조화된 질의를 생성함에 있어서, 자연어 질의에 대한 자연어 처리 결과 및 자연어 질의와 관련되는 데이터베이스 내의 하위 데이터베이스 간의 관계에 기반하여 추출되는 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 이용하여 구조화된 질의를 생성한다.
- [0017] 프로세서는 자연어 처리 결과의 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과, 및 스키마 관계의

자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 이용하여 구조화된 질의를 생성할 수 있고, 교차 어텐션 결과는 제1 교차 어텐션 결과 및 제2 교차 어텐션 결과를 포함할 수 있다.

- [0018] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 자연어 질의에 기반하여 자연어 처리 결과를 생성하는 자연어 처리 모델; 및 스키마 관계를 출력하는 스키마 인코더 모델을 더 포함할 수 있다.
- [0019] 스키마 인코더 모델은 입력 스키마 그래프에 기반하여 스키마 의미 표현을 스키마 관계로서 출력할 수 있다.
- [0020] 자연어 처리 모델은 자연어 질의의 전처리 결과를 입력받고 자연어 요소 의미 표현을 자연어 처리 결과로서 출력할 수 있다.
- [0021] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 자연어 처리 결과의 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과를 생성하는 제1 교차 어텐션 계층; 및 스키마 관계의 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 생성하는 제2 교차 어텐션 계층을 더 포함할 수 있다.
- [0022] 자연어 처리 모델은 내부에 셀프 어텐션 기능을 가지는 제1 내부 어텐션 계층을 포함할 수 있고, 스키마 인코더 모델은 내부에 셀프 어텐션 및 교차 어텐션 중 적어도 하나 이상의 기능을 가지는 제2 내부 어텐션 계층을 포함할 수 있다.
- [0023] 프로세서는 제1 내부 어텐션 계층의 출력에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트를 차단할 수 있다.
- [0024] 프로세서는 교차 어텐션 결과에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트를 차단할 수 있다.
- [0025] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 교차 어텐션 결과에 기반하여 구조화된 질의를 생성하는 디코더 계층을 더 포함할 수 있다.
- [0026] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은 프로세서(processor)를 포함하는 컴퓨팅 시스템에 의하여 실행될 수 있다. 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은 사용자에게 의하여 입력되는 자연어 질의를 수신하는 단계; 및 자연어 질의에 기반하여 구조화된 질의를 생성하는 단계를 포함하고, 구조화된 질의를 생성하는 단계는, 자연어 질의에 대한 자연어 처리 결과를 획득하는 단계; 자연어 질의와 관련된 데이터베이스 내의 하위 데이터베이스 간의 관계에 기반하여 추출되는 스키마 관계를 획득하는 단계; 자연어 처리 결과 및 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 획득하는 단계; 및 교차 어텐션 결과를 이용하여 구조화된 질의를 생성하는 단계를 포함한다.
- [0027] 교차 어텐션 결과를 획득하는 단계는, 자연어 처리 결과의 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과를 획득하는 단계; 및 스키마 관계의 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 획득하는 단계를 포함할 수 있다.
- [0028] 스키마 관계를 획득하는 단계에서는, 입력 스키마 그래프에 기반하여 스키마 의미 표현이 스키마 관계로서 획득될 수 있다. 자연어 처리 결과를 획득하는 단계에서는, 자연어 질의의 전처리 결과가 입력되고 자연어 요소 의미 표현이 자연어 처리 결과로서 획득될 수 있다.
- [0029] 자연어 처리 결과를 획득하는 단계는, 내부 어텐션 계층의 출력에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트가 차단된 상태에서 수행될 수 있다.
- [0030] 자연어 처리 결과를 획득하는 단계는, 교차 어텐션 결과에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트가 차단된 상태에서 수행될 수 있다.
- [0031] 본 발명의 일 실시예에 따른 프로세서(processor)를 포함하는 컴퓨팅 시스템에 의하여 실행되는 자연어 질의 처리 방법은 자연어 질의와 관련된 데이터베이스 내의 하위 데이터베이스 간의 스키마 관계를 일대일(one-to-one) 대응, 및 일대다(one-to-many) 대응 중 어느 하나로서 추출하는 단계; 및 스키마 관계에 기반하여 자연어 질의로부터 구조화된 질의를 생성하는 단계를 포함한다.
- [0032] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 하위 데이터베이스 간의 비교를 통해 스키마 관계를 강제적(mandatory) 관계, 및 선택적(optional) 관계 중 어느 하나로서 예측하는 단계를 더 포함할 수 있다.
- [0033] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 하위 데이터베이스 중 제1 하위 데이터베이스에 대응하는 제1 자연어 모델 및 제2 하위 데이터베이스에 대응하는 제2 자연어 모델을 이용하여 스키마 관계를 인코딩하는 단계를 더 포함할 수 있다.



- [0034] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 하위 데이터베이스 중 일부에 대한 마스크 컬럼 모델링을 이용하여 스키마 관계의 적어도 일부를 미리 인코딩하는 단계를 더 포함할 수 있다.
- [0035] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 자연어 질의에 기반하여 자연어 처리 결과를 획득하는 단계를 더 포함할 수 있다. 이때 구조화된 질의를 생성하는 단계는, 자연어 처리 결과 및 스키마 관계의 상호간에 생성되는 교차 어텐션 결과를 획득하는 단계; 및 교차 어텐션 결과를 이용하여 구조화된 질의를 생성하는 단계를 포함할 수 있다.
- [0036] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 미리 학습된 언어 모델을 이용하여 자연어 질의에 기반하여 자연어 처리 결과를 획득하는 단계를 더 포함할 수 있다. 이때 자연어 처리 결과를 획득하는 단계는, 언어 모델의 내부 파라미터의 업데이트가 차단된 상태에서 수행될 수 있다.

**발명의 효과**

- [0037] 본 발명의 실시예에 따르면, 자연어 질의를 SQL 질의와 같은 규격화/구조화된 질의로 번역하는데 있어서 필요한 자연어 질의와 데이터베이스 스키마에 대한 의미 표현을 더욱 정확하게 계산하여 번역 시스템의 성능을 높일 수 있다.
- [0038] 본 발명의 실시예에 따르면 사용자가 구조화된 질의에 대한 지식 없이 자연어 질의를 입력하는 경우에도 자동으로 구조화된 질의를 생성하여 사용자가 원하는 질의 응답 결과를 제공할 수 있다.
- [0039] 본 발명의 실시예에 따르면 사용자가 입력한 자연어 질의만으로는 얻을 수 있는 스키마 정보를 미리 분석하고 인코딩함으로써 사용자의 의도에 부합하는 구조화된 질의를 생성하고 질의 응답 결과를 제공할 수 있다.
- [0040] 본 발명의 실시예에 따르면 사용자가 자연어 질의를 입력하는 경우에 구조화된 질의 및 질의 응답 결과를 생성하는 과정을 고도화할 수 있다.
- [0041] 본 발명의 실시예에 따르면 자연어 질의 입력에 응답하여 구조화된 질의를 생성하는 과정을 고도화하는 교차 어텐션 기반의 새로운 신경망 구조를 구현할 수 있다.
- [0042] 본 발명의 실시예에 따르면 자연어 질의 입력에 응답하여 구조화된 질의를 생성하는 과정을 고도화하는, 하위 데이터베이스 간의 관계에 기반한 스키마 관계를 추출할 수 있다.

**도면의 간단한 설명**

- [0043] 도 1은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치를 포함하는 시스템을 도시하는 개념도이다.
- 도 2는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 의하여 실행되는 자연어 처리 시나리오의 일 예를 도시하는 개념도이다.
- 도 3은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 의하여 실행되는 자연어 질의 처리 과정을 설명하기 위한 가상의 사용자에 의한 증가적인 질의 생성 과정을 도시하는 개념도이다.
- 도 4는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치의 내부 구조로서 자연어 처리를 위한 언어 모델, 스키마 인코더, 및 교차 어텐션(Cross-attention) 계층을 도시하는 개념도이다.
- 도 5는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치의 내부 구조로서 언어 모델의 내부 파라미터의 업데이트가 차단되는 경우를 도시하는 개념도이다.
- 도 6은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치의 내부 구조로서 교차 어텐션 결과에 기반하여 구조화된 질의를 생성하는 디코더 계층을 더 포함하는 실시예를 도시하는 개념도이다.
- 도 7은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 의하여 실행되는 자연어 질의 처리 방법을 도시하는 동작 흐름도이다.
- 도 8은 도 7의 단계 S350 및 그 전후의 과정을 더욱 상세하게 도시하는 동작 흐름도이다.
- 도 9는 본 발명의 다른 일 실시예에 따른 자연어 질의 처리 방법을 도시하는 동작 흐름도이다.
- 도 10은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 자동차 관련된 자연어 질의와 스키마가 주어졌을 때의 스키마 링크 과정의 일 실시예를 설명하기 위한 개념도이다.

도 11은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 하나의 자연어 질의가 데이터베이스 내의 하위 데이터베이스(테이블) 간의 관계에 따라 서로 다른 구조화된 질의로 번역될 수 있음을 설명하기 위한 개념도이다.

도 12는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 데이터베이스 레코드 값을 활용하여 하위 데이터베이스(테이블) 간의 관계를 추출하는 실시예를 도시하는 개념도이다.

도 13은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 스키마 인코딩 예시를 도시하는 개념도이다.

도 14는 본 발명의 일 실시예에 따른 일반화된 자연어 처리 장치 또는 컴퓨팅 시스템을 도시하는 블록도이다.

**발명을 실시하기 위한 구체적인 내용**

- [0044] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 실시예를 가질 수 있는 바, 특정 실시예들을 도면에 예시하여 상세하게 설명하고자 한다. 그러나, 이는 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였다.
- [0045] 제1, 제2, A, B 등의 용어는 다양한 구성요소들을 설명하는 데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다. "및/또는"이라는 용어는 복수의 관련된 기재된 항목들의 조합 또는 복수의 관련된 기재된 항목들 중의 어느 항목을 포함한다.
- [0046] 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.
- [0047] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0048] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0049] 상기 제시한 선행기술인 한국등록특허 KR 10-2150908호 "자연어 질의해석 방법 및 시스템", 한국등록특허 KR 10-2277787호 "신경망 기반 자연어로부터 SQL 질의 번역 시 사용되는 컬럼 및 테이블을 예측하는 방법", 한국등록특허 KR 10-2345568호 "자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법" 등에 의하여 개시된 자연어 질의를 구조화된 질의로 번역하는 방법, 이때 필요한 메타 데이터인 스키마를 해석하기 위하여 필요한 데이터베이스의 하위 데이터베이스 간의 관계를 분석하는 방법과, "Attention Is All You Need", Ashish Vaswani 외, 31st Conference on Neural Information Processing Systems (NIPS 2017) (2017월 12월 6일)에서 개시된 트랜스포머 신경망 구조 등을 포함하여 선행문헌들에 개시된 사항은 본 발명의 목적에 부합하는 범위 내에서 본 발명의 구성의 일부 또는 전부로서 포함될 수 있다. 당업자라면 상기 선행기술 문헌들의 내용으로부터 본 발명의 목적 및 구성과의 연관성을 자명하게 유추할 수 있을 것이므로 본 발명의 취지를 흐릴 수 있는 지나치게 자세한 설명은 생략하며, 상기 선행기술 문헌을 소개하는 것으로 설명에 갈음한다.
- [0050] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 본 발명을 설명함에 있어 전체적인 이해를 용이하게 하기 위하여 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.

- [0051] 도 1은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치를 포함하는 시스템을 도시하는 개념도이다.
- [0052] 도 1을 참조하면, 시스템은 데이터베이스에 대한 자연어 질의를 번역하는 시스템으로 이해될 수 있다. 또한 시스템은 데이터베이스에 사용자(100)가 자연어 질의를 입력하는 경우, 자연어 질의에 대응하는 응답 결과를 자동으로 생성하는 시스템으로 이해될 수 있다.
- [0053] 시스템 내의 질의 번역기(240)는 트랜스포머 신경망을 내부에 포함할 수 있으며, 트랜스포머 신경망에 기반하여 자연어 질의와 관계형 데이터베이스의 스키마 추출 결과를 의미 표현(meaning representation)으로 인코딩할 수 있다. 질의 번역기(240)는 인코딩된 결과를 디코딩하여 구조화된 질의로 변환하고, 답변 엔진(260)에 전달할 수 있다.
- [0054] 일반적인 데이터베이스 관리 시스템(DBMS)을 사용하기 위해서는 일반적인 사용자는 해당 데이터베이스의 구조 및 구조화된 질의를 생성하기 위한 문법을 이해하고 있어야 한다. 이때 일반적인 사용자는 구조화된 질의를 적절하게 작성하고 입력하기 위하여 해당 데이터베이스의 구조에 따르는 스키마(schema)에 대한 지식을 필요로 한다. 이때 구조화된 질의는 일반적으로 널리 알려진 SQL(Structured Query Language)일 수 있고, 알려진 다른 형태의 구조화된 질의일 수 있다. 일반적인 사용자가 필요로 하는 스키마에 대한 지식 및 구조화된 질의에 대한 지식의 예시는 후술할 도 3에서 설명하기로 한다.
- [0055] 도 1의 시스템에서는 사용자(100)는 구조화된 질의 및/또는 스키마에 대한 직접적인 지식을 필요로 하지 않을 수 있다. 사용자(100)는 대화 관리자(220)에 자연어 질의를 입력할 수 있다. 이때 자연어 질의의 입력 방식은 키보드/키패드를 경유하는 텍스트 입력일 수도 있고, 음성 입력일 수도 있다. 음성 입력의 경우에는 대화 관리자(220)에 의하여 음성 인식 결과를 자연어 질의로 변환하는 전처리 과정을 거칠 수 있다.
- [0056] 대화 관리자(220)는 자연어로 형성된 질의를 질의 번역기(240)로 전달할 수 있다. 질의 번역기(240)는 자연어 질의의 입력을 구조화된 질의로 번역할 수 있다. 질의 번역기(240)는 구조화된 질의를 응답 엔진(260)으로 전달할 수 있다.
- [0057] 응답 엔진(260)은 공지된 인공지능 기반 대화형 비서와 같은 기능을 포함할 수 있다. 응답 엔진(260)은 구조화된 질의에 대응하는 데이터베이스 내의 검색 결과를 제공할 수 있다. 응답 엔진(260)은 검색 결과를 자연어 응답 및/또는 시각화된 결과로서 사용자(100)에게 제공할 수 있다.
- [0058] 설명의 편의상 도 1에 도시되지는 않았지만 도 1의 대화 관리자(220), 질의 번역기(240), 및 응답 엔진(260)은 적어도 하나 이상의 프로세서와 전자적으로 연결되고, 프로세서에 의하여 관리 및 제어될 수 있다. 프로세서는 적어도 하나 이상의 메모리와 전자적으로 연결되고, 메모리에 저장된 적어도 하나 이상의 명령어가 프로세서에 의하여 실행됨으로써 도 1의 대화 관리자(220), 질의 번역기(240), 및 응답 엔진(260)의 기능이 제공될 수 있다. 본 발명의 또 다른 실시예에서는 도 1의 대화 관리자(220), 질의 번역기(240), 및 응답 엔진(260) 중 적어도 일부가 전용 하드웨어의 형태로 구현될 수도 있다.
- [0059] 본 발명의 목적을 달성하기 위한 일 실시예에 따른 시스템/장치는 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 있는 표현으로 인코딩하는 기능을 포함할 수 있다.
- [0060] 시스템/장치는 다국어 음성인식이 필요한 상황에서 단말기의 하드웨어적 제약으로 인해 단일어만 음성인식이 가능한 경우 적어도 하나의 인공지능 모델로 다국어 음성인식을 수행하고, 자동으로 사용자의 발화를 인식하여 처리할 수 있는 음성인식 서비스를 제공할 수 있다.
- [0061] 시스템/장치는 영어(EN)와 한국어(KO)의 오디오 입력을 7개 층의 CNN(Convolutional Neural Network) 구조의 Feature Extractor를 거친 후, 24개 층의 Transformer Encoder를 거쳐 음성 특징 정보를 추출하고, 한국어(KO), 영어(EN) 또는 한국어와 영어를 모두(ALL) 출력가능한 CTC Projection 출력 계층을 통해 최종 음성인식 결과를 출력할 수 있다.
- [0062] 시스템/장치는, 각 CTC Projection 출력 계층의 출력 단위는 1byte(8bit)에 해당하는 유니코드(Unicode) 단위이고, 이를 재조합해 최종 결과를 생성할 수 있다. 이때 byte 단위의 정보는 8진수 표현 방식으로 0x00 ~ 0xff로 표현될 수 있다. byte 단위의 유니코드 표현은 전세계의 모든 언어를 256(2의8승)개의 8bit(1byte)의 조합으로 표현할 수 있으므로, 지원하는 언어의 수가 증가하더라도 전체 모델의 크기는 256 단계의 출력을 지원하는 CTC Projection 출력 계층만큼만 증가할 수 있다.
- [0063] 시스템/장치는, 인식하는 모든 언어가 공통적으로 공유되고, 각 언어별로 별도의 CTC Projection 출력 계층을 가지도록 구현될 수 있다. 음성 언어 분류가 확실하지 않은 경우 모든 언어가 동시에 학습되어 모든 언어가 출

력이 가능한 범용(ALL) 출력 계층이 이용될 수 있다. 범용 출력 계층은 전용(KO or EN) 출력 계층보다 성능이 떨어지지만, 음성 언어 인식이 제대로 이뤄지지 않았을 때 잘못된 언어의 출력을 생성하는 오작동을 방지할 수 있다.

- [0064] 시스템/장치는, 다국어 음성인식의 성능을 증진시키기 위해, 음성 언어 분류 모델을 탑재할 수 있다.
- [0065] 시스템/장치의 음성 언어 분류 모델은, Wav2Vec2.0 모델을 기반으로 하며, 음성 정보를 25ms(밀리초) 단위로 나누어 Wav2Vec2.0 모델을 이용해 음성 정보를 추출하고, Projector 계층을 통해 25ms 단위의 추출된 음성 정보들의 평균값(Mean Pooling 연산 수행)을 음성 언어 분류기(Classifier)를 통해 분류할 수 있다.
- [0066] 시스템/장치는, 입력으로 받은 오디오 데이터를 Wav2Byte 음성인식기와 Wav2Vec2.0기반 음성 언어 분류기에 각각 입력하고, Wav2Byte 구조의 음성인식기의 언어별 CTC Projection 출력 계층을 선택할 때 Wav2Vec2.0 기반 음성 언어 분류기의 출력(언어 분류 정보)를 이용해 출력 계층을 선택하여 최종 음성 인식 결과를 생성할 수 있다. 시스템/장치는 음성 언어 분류의 확실성 (Confidence)가 70% 미만일 경우에는 오작동을 방지하기 위하여 전용 언어 출력 계층이 아닌 범용(ALL) 언어 출력 계층을 이용해 음성 인식 결과를 생성할 수 있다.
- [0067] 도 2는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 의하여 실행되는 자연어 처리 시나리오의 일 예를 도시하는 개념도이다.
- [0068] 도 2를 참조하면, 도 1의 대화 관리자(220), 질의 번역기(240), 및 응답 엔진(260)을 경유하여 사용자(100)가 데이터베이스 관리 시스템(DBMS)(200)과 대화형 인터페이스를 이용하여 질의 및 검색 결과를 제공받는 시나리오가 도시된다.
- [0069] 사용자(100)는 300,000을 초과하는 ATM 거래량을 가지는 도시에 대하여 자연어 질의를 입력할 수 있다. DBMS(200)는 300,000을 초과하는 거래량을 가지는 도시는 검색되지 않으나, 300,000과 가장 가까운 299,580 건을 가지는 도시로서 서울(Seoul)이 검색되었음을 리포트할 수 있다. 도 1에서 사용자(100)는 은행 또는 금융기관 종사자를 가정하였으나, 본 발명의 사상은 이러한 가정에 국한되어 해석되지 아니 한다.
- [0070] 이때 도 1과 도 2를 함께 참고하면, 도 1의 질의 번역기(240) 및 응답 엔진(260) 중 적어도 하나 이상은 상호 협력하여, 또는 단독으로 구조화된 질의에 따른 검색 결과에 기반하여 보조적인 구조화된 질의를 추가로 생성할 수 있고 이에 따른 지능형 응답을 사용자(100)에게 제공할 수 있다. 예를 들어 질의 검색 결과 적절한 수의 결과가 검색되면 원본 검색 결과가 그대로 사용자(100)에게 제공될 수 있고, 적절한 수의 결과가 검색되지 않으면 보조 검색 결과가 함께 사용자(100)에게 제공될 수도 있다.
- [0071] 사용자(100)는 첫번째 질의 응답 결과를 수신한 후 두번째 질의를 입력할 수 있다. 이때 사용자(100)는 무엇에 대한 검색인 지를 일부 생략한 채로 자연어 질의를 입력할 수 있고, DBMS(200)는 스키마 관계에 기반하여 사용자(100)의 의도에 부합하는 구조화된 질의를 생성하고, 데이터베이스에서 질의를 검색하고, 질의 응답 결과를 사용자(100)에게 제공할 수 있다.
- [0072] 도 3은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 의하여 실행되는 자연어 질의 처리 과정을 설명하기 위한 가상의 사용자에 의한 증가적인 질의 생성 과정을 도시하는 개념도이다.
- [0073] 도 3을 참조하면, 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 사용자(100)는 자연어 질의를 입력할 수 있다. 이때 자연어 질의 처리 장치 내부의 동작을 설명하기 위하여 가상의 사용자(120)를 가정한다. 가상의 사용자(120)는 사용자(100)의 자연어 질의에 응답하여 구조화된 질의를 생성할 수 있다.
- [0074] 가상의 사용자(120)가 생성하는 구조화된 질의는 도 3의 시스템 내에서 사용자(100)의 자연어 질의와 증가적인 질의일 수 있다.
- [0075] 도 3의 실시예를 달리 설명하면, 종래 기술에서는 가상의 사용자(120)는 도 3에서 도시된 것처럼 구조화된 질의를 직접 작성하여 입력할 필요가 있었다. 이때 가상의 사용자(120)는 300,000건 이상의 거래가 발생한 도시를 확인하기 위하여, 해당 사용자는 세 개의 테이블 스키마를 숙지해야 하고 복잡한 중첩 SQL 질의를 작성할 수 있어야 함을 알 수 있다.
- [0076] 그러나 일반 대중들은 SQL 언어에 대해 알지 못하며, SQL 질의에 대한 지식을 가진 전문가일 지라도 데이터베이스의 특수성까지를 고려하여 자신의 의도를 정확히 반영한 질의를 작성하는 것은 쉽지 않다.
- [0077] 또한, 종래 기술의 DBMS의 사용자는 SQL 질의를 제대로 작성하였는지 스스로 확인하기 어렵다.

- [0078] 예를 들어 가상의 사용자(120)가 SQL 기반 질의를 정확히 작성하더라도, DBMS가 반환하는 질의 응답 결과가 검색 결과 없음으로 제공될 경우에 사용자는 본인이 SQL 질의를 정확히 작성하였는지 확인할 수 없을 것이다.
- [0079] 하지만 도 3의 실시예에서는 사용자(100)는 DBMS(200)에 자연어로 질의를 할 수 있고, 사용자(200)는 자신이 원하는 정보를 말로 정확하게 표현할 수 있으며 복잡한 하위 데이터베이스(테이블)의 스키마를 이해할 필요가 없다.
- [0080] 본 발명의 일 실시예에서는 신경망 기반의 자연어 처리 기술을 이용할 수 있다. 이러한 자연어 처리 기술은 공지된 트랜스포머 신경망 등을 이용할 수 있으며, 필요에 따라 본 발명의 구성의 일부로서 포함될 수 있다.
- [0081] 최근 몇 년간 신경망 기반의 여러 자연어 질의 번역 기술들이 등장하였다. 초기에는 순환 신경망/장단기 기억 네트워크(recurrent neural networks/Long Short Term Memory networks) 기반의 연구들이 소개되었으나, 입력 길이가 길어졌을 때 기울기 폭주/손실(gradient exploding/vanishing) 문제가 발생하여 최근 연구에서는 트랜스포머 구조의 신경망이 많이 사용되고 있다.
- [0082] 트랜스포머 구조를 활용하는 대표적인 기술로 RAT-SQL이 있다. RAT-SQL은 인코더-디코더(encoder-decoder) 모델로 자연어 질의와 스키마를 사전 학습된 언어 모델 BERT에 동시에 넣어 의미 표현(meaning representation)을 계산하고 관계 인식 트랜스포머 신경망을 이용하여 입력 토큰들간의 명시적인 관계정보를 활용하여 토큰들의 최종 의미 표현을 계산한다.
- [0083] 자연어 질의와 스키마를 종래 기술의 트랜스포머 신경망에 입력하기 위해서는 스키마 정보를 연속적인 단어로 만들어 토큰화 과정을 거쳐야 한다. RAT-SQL은 schema에 포함되는 테이블 이름과 컬럼 이름을 일렬로 나열하여 [SEP](분리자)라는 특별한 토큰을 사이에 두고 자연어 질의와 연결한다.
- [0084] 또한, 관계인식 트랜스포머를 사용하기 위해서는 토큰들끼리의 명시적인 관계를 추출해야 하는데, 스키마 토큰들 사이의 관계는 스키마 그래프 상의 연결 관계를 사용하고, 자연어 질의와 스키마 토큰 사이의 관계는 스키마 링킹(schema-linking)을 통해 알아낸다. 이렇게 계산된 최종 의미 표현을 가지고 디코더에서 SQL phrase를 순차적으로 생성한다.
- [0085] RAT-SQL 이후 다양한 자연어 질의 번역 기술들이 추가로 등장하고 있으나, 대부분 RAT-SQL의 인코더 방식을 따른다. 인코딩 방식에서 모델 파라미터(model parameter)를 효과적으로 초기화하는 방식이 제안되었고, 이후 디코더가 새롭게 제안되었다.
- [0086] RAT-SQL 모델은 사전 학습된 언어 모델 BERT를 활용해 자연어 질의 번역 성능을 많이 끌어 올릴 수 있다.
- [0087] 하지만 종래 기술인 RAT-SQL이 BERT를 사용하는 방식은 몇 가지의 문제점이 있다.
- [0088] 첫 번째로 BERT는 자연어 문장을 받도록 학습되었는데 RAT-SQL은 자연어 질의 문장과 함께 테이블과 컬럼 이름을 나열하여 입력받는다. 이러한 경우 BERT는 새로운 입력 형식에 대해서 효과적인 의미 표현을 추출하기 위하여 추가학습을 해야 하지만 자연어 질의와 스키마 쌍에 대한 학습 데이터가 많지 않다.
- [0089] 두 번째로 BERT는 입력 길이는 토큰 개수 512개로 제한이 있어 현재와 같은 구조는 컬럼과 테이블의 개수가 많은 스키마에 대해서는 적용이 불가능하다는 단점이 있다.
- [0090] 세 번째로는 입력 토큰들의 관계정보를 효과적으로 사용하지 못한다는 단점이 있다.
- [0091] 종래 기술인 RAT-SQL은 컬럼과 테이블을 비롯한 토큰들 사이의 연결 관계를 활용하기 위해서 관계 인식 트랜스포머에 토큰끼리의 연결 관계를 명시적으로 입력할 수 있다.
- [0092] 하지만, 관계 인식 트랜스포머는 인코더의 뒤편에서 사용되는 모듈로, 입력 토큰이 처음 주어진 앞부분에서 토큰들 간의 관계정보가 전혀 활용되지 않는다는 문제가 있다. 이러한 모델 구조는 잘못된 중간 의미 표현을 만들어 모델의 뒷부분에 영향을 줄 가능성이 높다.
- [0093] 또 다른 종래 기술들은 자연어 질의를 해석하는 과정을 고도화하기 위하여 메타 데이터 또는 스키마를 이용한다. 예를 들어 앞에서 설명한 한국등록특허 KR 10-2150908호 "자연어 질의해석 방법 및 시스템"에서는 형태소 분석에 기반하여 결정되는 각 자연어 덩들의 제1 유형들을 스키마를 이용하여 더욱 상세한 하위 유형들인 제2 유형/성분 패턴으로 구체화하는 구성이 개시된다.
- [0094] 그러나 이러한 종래 기술들은 자연어 질의를 구성하는 각 덩들의 의미를 더욱 상세하게 추출할 수는 있지만 여전히 구조화된 질의(예를 들어 SQL 질의)를 생성하기 위하여 필요한 정보를 모두 추출하지는 못한다.

- [0095] 이러한 종래 기술의 문제점을 해결하기 위하여 본 발명은 자연어 질의 처리 장치를 고도화할 수 있는 새로운 내부 구조를 제안한다.
- [0096] 도 4는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치의 내부 구조로서 자연어 처리를 위한 언어 모델, 스키마 인코더, 및 교차 어텐션(Cross-attention) 계층을 도시하는 개념도이다.
- [0097] 도 4는 본 발명의 일 실시예의 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 있는 표현으로 인코딩하는 장치의 자연어 질의와 스키마를 각자 인코딩한 후 두 관계를 고려하여 의미 표현을 다시 계산하는 과정을 설명하기 위한 개념도이다. 도 4에서는 도시되지 않았지만 도 4의 각 구성요소들은 프로세서와 전자적으로 연결되고, 프로세서에 의하여 각 구성요소들의 동작 및 기능이 제어 및 관리될 수 있다.
- [0098] 도 4를 참조하면, 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 사용자(100)에 의하여 입력된 자연어 질의를 수신하고, 자연어 질의에 기반하여 구조화된 질의를 생성하는 질의 번역기(240)일 수 있다.
- [0099] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 자연어 질의에 기반하여 구조화된 질의를 생성함에 있어서, 자연어 질의에 대한 1) 자연어 처리 결과 및 자연어 질의와 관련된 데이터베이스 내의 하위 데이터베이스 간의 관계에 기반하여 추출되는 스키마 관계의 상호 간에 생성되는 2) 교차 어텐션 결과를 이용하여 구조화된 질의를 생성할 수 있다.
- [0100] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 자연어 처리 결과의 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과, 및 스키마 관계의 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 이용하여 구조화된 질의를 생성할 수 있고, 교차 어텐션 결과는 제1 교차 어텐션 결과 및 제2 교차 어텐션 결과를 포함할 수 있다.
- [0101] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 자연어 질의에 기반하여 자연어 처리 결과를 생성하는 자연어 처리 모델로서 언어 모델(244); 및 스키마 관계를 출력하는 스키마 인코더(242) 모델을 더 포함할 수 있다.
- [0102] 스키마 인코더(242) 모델은 입력 스키마 그래프에 기반하여 스키마 의미 표현을 스키마 관계로서 출력할 수 있다.
- [0103] 자연어 처리 모델인 언어 모델(244)은 자연어 질의의 전처리 결과를 입력받고 자연어 요소(예를 들어 파싱된 단어, 형태소, 토큰) 의미 표현을 자연어 처리 결과로서 출력할 수 있다.
- [0104] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 자연어 처리 결과의 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과를 생성하는 제1 교차 어텐션 계층(246); 및 스키마 관계의 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 생성하는 제2 교차 어텐션 계층(248)을 더 포함할 수 있다.
- [0105] 자연어 처리 모델인 언어 모델(244)은 내부에 셀프 어텐션의 기능을 가지는 제1 내부 어텐션 계층을 포함할 수 있고, 스키마 인코더(242) 모델은 내부에 셀프 어텐션 및 교차 어텐션 중 적어도 하나 이상의 기능을 가지는 제2 내부 어텐션 계층을 포함할 수 있다.
- [0106] 도 5는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치의 내부 구조로서 언어 모델의 내부 파라미터의 업데이트가 차단되는 경우를 도시하는 개념도이다.
- [0107] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 제1 내부 어텐션 계층의 출력에 기반한 자연어 처리 모델/언어 모델(244) 내부의 파라미터의 업데이트를 차단할 수 있다.
- [0108] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 교차 어텐션 결과에 기반한 자연어 처리 모델/언어 모델(244) 내부의 파라미터의 업데이트를 차단할 수 있다.
- [0109] 도 5를 참조하면, 자연어 처리 과정과 스키마 인코딩 과정이 분리됨으로써 스키마 인코더(242) 모델만 자연어 처리 과정과 독립적으로 사전 학습할 수 있다. 이는 본 발명과 종래 기술을 구분하는 큰 차이점으로서 본 발명의 실시예에 따른 자연어 질의 처리 기법을 더욱 고도화할 수 있는 요소이기도 하다.
- [0110] 스키마 인코더(242) 모델의 사전 학습 과정을 자연어 처리 과정과 분리함으로써, 웹에서 쉽게 구할 수 있는 하위 데이터베이스(테이블)의 정보로부터 스키마를 생성하여 스키마 인코더(242) 모델을 비지도 학습할 수 있다. 스키마 인코더(242) 모델을 사전 학습함으로써 종래 기술의 문제점이었던 스키마 입력 형태가 사전 학습된 모델의 입력 형태와 맞지 않는 문제가 해결될 수 있다.

- [0111] 사전 학습을 위한 2가지 학습 과정은, 마스크 컬럼 모델링(masked column modeling) 및/또는 테이블 관계 예측(table relationship prediction)을 이용할 수 있다.
- [0112] 테이블 내의 연관성을 학습할 수 있는 마스크 컬럼 모델링은 마스크 언어 모델링과 유사하게 15%의 확률로 테이블 컬럼의 이름을 마스크한 후 해당 컬럼의 이름을 예측하는 방식으로 수행될 수 있다. 이때 컬럼의 이름을 예측하기 위하여 사용되는 의미 표현으로는 BERT를 통과한 이후에 2층의 완전 연결 계층(fully-connected layer)를 통해서 나온 의미 표현이 이용될 수 있고, 모델이 마스크된 컬럼의 이름을 예측하는 분류문제를 풀도록 제어될 수 있다.
- [0113] 테이블 간의 연결성을 잘 이해하기 위한 학습으로서 수행될 수 있는 테이블 관계 예측 기법은 관계 인식 트랜스포머에 들어가는 관계 정보를 15%의 확률로 마스크하고, 해당 위치에 어떤 관계가 배치되어야 하는 지 맞추는 분류문제를 모델이 풀도록 제어됨으로써 수행될 수 있다.
- [0114] 마스크 컬럼 모델링(masked column modeling) 및/또는 테이블 관계 예측(table relationship prediction)의 더욱 구체적인 과정은 후술할 도 10 내지 도 13에 의하여 설명하기로 한다.
- [0115] 다시 도 4 및 도 5를 참조하면, 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 인코딩 단계에서 자연어 질의의 자연어 관계와 스키마 관계를 개별적으로 고려하여 인코딩 결과를 생성할 수 있다. 언어 모델(244) 및 스키마 인코더(242)는 서로 독립적으로 각자의 입력에 대한 의미 표현을 계산할 수 있다.
- [0116] 자연어 질의에 기반하여 구조화된 질의를 효과적으로 생성하기 위해서는, 스키마와 자연어 간의 관련성을 정확히 인식하는 것이 중요하다. 이를 위하여 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 교차 어텐션(cross-attention)을 이용할 수 있다.
- [0117] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 교차 어텐션을 이용하여 각각 인코딩된 두 입력(자연어 처리 결과 및 스키마 관계) 간의 정보가 서로 참조되어 활용될 수 있도록 구성될 수 있다. 교차 어텐션에서 어텐션 점수(attention score)를 계산하는 수학적식은 관계 정보를 뜻하는  $r$ 을 포함할 수 있다. 이때 관계 정보  $r$ 로서 스키마 링크를 통해서 계산된 관계가 이용될 수 있다. 교차 어텐션을 활용해 자연어 처리 관계와 스키마 관계의 최종적인 의미 표현이 계산될 수 있다.
- [0118] 이와 같은 구성을 통하여 본 발명의 실시예는 자연어 질의를 SQL 질의와 같은 규격화/구조화된 질의로 번역하는 데 있어서 필요한 자연어 처리 결과와 데이터베이스 스키마에 대한 의미 표현을 더욱 정확하게 계산하여 번역 시스템의 성능을 높일 수 있다.
- [0119] 도 6은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치의 내부 구조로서 교차 어텐션 결과에 기반하여 구조화된 질의를 생성하는 디코더 계층을 더 포함하는 실시예를 도시하는 개념도이다.
- [0120] 본 발명의 일 실시예에 따른 자연어 질의 처리 장치는 교차 어텐션 결과에 기반하여 구조화된 질의를 생성하는 디코더 계층(250)을 더 포함할 수 있다.
- [0121] 도 4 내지 도 6의 실시예를 참조하면, 구조화된 질의를 생성하기 위하여 자연어 질의 내의 자연어 처리 결과 및 자연어 질의에 직접적으로 포함되지 않을 수 있는 스키마 관계를 상호 참조하여 추가적으로 의미 표현을 인코딩할 수 있다. 스키마 관계는 자연어 질의 내에 포함되는 정보 및 자연어 질의 내에 포함되지 않는 정보를 모두 포함할 수 있다.
- [0122] 스키마 관계에 대한 자연어 처리 결과의 제2 교차 어텐션(248)을 이용하여 스키마 관계 중 자연어 질의 번역을 위하여 집중해야 할 하위 데이터베이스(테이블/컬럼) 정보가 얻어질 수 있다. 결과적으로 전체 스키마 중 일부분의 테이블/컬럼이 유의미하게 선택되어 이용될 수 있다.
- [0123] 자연어 처리 결과에 대한 스키마 관계의 제1 교차 어텐션(246)을 이용하여 자연어 질의를 구성하는 자연어 요소(파싱 결과, 단어, 형태소, 토큰)들 중 어떤 요소가 어떤 테이블/컬럼과 관련되는 지에 대한 정보를 얻을 수 있다. 이때 얻어지는 자연어 질의 내의 단어와 데이터베이스 내의 테이블/컬럼 간의 의미 대응 관계를 이용하여 구조화된 질의가 생성될 수 있다.
- [0124] 도 7은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에 의하여 실행되는 자연어 질의 처리 방법을 도시하는 동작 흐름도이다.
- [0125] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은 프로세서(processor)를 포함하는 컴퓨팅 시스템에 의하여 실행될 수 있다. 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은 사용자(100)에 의하여 입력되는 자연어

질의를 수신하는 단계(S310); 및 자연어 질의에 기반하여 구조화된 질의를 생성하는 단계(S320)를 포함하고, 구조화된 질의를 생성하는 단계(S320)는, 자연어 질의에 대한 자연어 처리 결과를 획득하는 단계(S340); 자연어 질의와 관련되는 데이터베이스 내의 하위 데이터베이스(테이블/컬럼) 간의 관계에 기반하여 추출되는 스키마 관계를 획득하는 단계(S330)); 자연어 처리 결과 및 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 획득하는 단계(S350); 및 교차 어텐션 결과를 이용하여 구조화된 질의를 생성하는 단계(S360)를 포함한다.

- [0126] 도 8은 도 7의 단계 S350 및 그 전후의 과정을 더욱 상세하게 도시하는 동작 흐름도이다.
- [0127] 교차 어텐션 결과를 획득하는 단계(S350)는, 자연어 처리 결과의 스키마 관계와의 관련도를 나타내는 제1 교차 어텐션 결과를 획득하는 단계(S352); 및 스키마 관계의 자연어 처리 결과와의 관련도를 나타내는 제2 교차 어텐션 결과를 획득하는 단계(S354)를 포함할 수 있다.
- [0128] 스키마 관계를 획득하는 단계(S330)에서는, 입력 스키마 그래프에 기반하여 스키마 의미 표현이 스키마 관계로서 획득될 수 있다. 자연어 처리 결과를 획득하는 단계(S340)에서는, 자연어 질의의 전처리 결과가 입력되고 자연어 요소(파싱 결과, 단어, 형태소, 토큰) 의미 표현이 자연어 처리 결과로서 획득될 수 있다.
- [0129] 자연어 처리 결과를 획득하는 단계(S340)는, 내부 어텐션 계층의 출력에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트가 차단된 상태에서 수행될 수 있다.
- [0130] 자연어 처리 결과를 획득하는 단계(S340)는, 교차 어텐션 결과에 기반한 자연어 처리 모델 내부의 파라미터의 업데이트가 차단된 상태에서 수행될 수 있다.
- [0131] 도 9는 본 발명의 다른 일 실시예에 따른 자연어 질의 처리 방법을 도시하는 동작 흐름도이다.
- [0132] 본 발명의 일 실시예에 따른 프로세서(processor)를 포함하는 컴퓨팅 시스템에 의하여 실행되는 자연어 질의 처리 방법은 자연어 질의와 관련되는 데이터베이스 내의 하위 데이터베이스 간의 스키마 관계를 일대일(one-to-one) 대응, 및 일대다(one-to-many) 대응 중 어느 하나로서 추출하는 단계(S430); 및 스키마 관계에 기반하여 자연어 질의로부터 구조화된 질의를 생성하는 단계(S460)를 포함한다. 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 자연어 질의를 수신하는 단계(S410)를 더 포함할 수 있다.
- [0133] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 하위 데이터베이스 간의 비교를 통해 스키마 관계를 강제적(mandatory) 관계, 및 선택적(optional) 관계 중 어느 하나로서 예측하는 단계를 더 포함할 수 있다.
- [0134] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 하위 데이터베이스 중 제1 하위 데이터베이스에 대응하는 제1 자연어 모델 및 제2 하위 데이터베이스에 대응하는 제2 자연어 모델을 이용하여 스키마 관계를 인코딩하는 단계를 더 포함할 수 있다.
- [0135] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 하위 데이터베이스 중 일부에 대한 마스크 컬럼 모델링을 이용하여 스키마 관계의 적어도 일부를 미리 인코딩하는 단계를 더 포함할 수 있다.
- [0136] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 자연어 질의에 기반하여 자연어 처리 결과를 획득하는 단계를 더 포함할 수 있다. 이때 구조화된 질의를 생성하는 단계는, 자연어 처리 결과 및 스키마 관계의 상호 간에 생성되는 교차 어텐션 결과를 획득하는 단계; 및 교차 어텐션 결과를 이용하여 구조화된 질의를 생성하는 단계를 포함할 수 있다.
- [0137] 본 발명의 일 실시예에 따른 자연어 질의 처리 방법은, 미리 학습된 언어 모델을 이용하여 자연어 질의에 기반하여 자연어 처리 결과를 획득하는 단계를 더 포함할 수 있다. 이때 자연어 처리 결과를 획득하는 단계는, 언어 모델의 내부 파라미터의 업데이트가 차단된 상태에서 수행될 수 있다.
- [0138] 도 9의 실시예에 따르는 자연어 질의 처리 방법은 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 있는 표현으로 인코딩하는 과정을 포함할 수 있다.
- [0139] 이때 방법은 종래의 기술과 동일한 방식의 인코더를 이용할 수 있지만 결과적으로 자연어 질의와 SQL 쌍 데이터를 새롭게 만들어 사전 학습하는 방식이 제안될 수 있다.
- [0140] 종래 기술에서 드러난 문제점인 자연어 질의와 스키마 쌍에 대한 학습 데이터의 부족, 자연어 처리 모델의 입력 토큰의 개수에 제한(컬럼과 테이블의 개수가 많은 스키마에 대해서는 적용 불가), 입력들의 관계 정보가 비효율적으로 이용되는 점을 해결하기 위하여, 본 발명의 실시예에서는 자연어 질의와 스키마를 인코딩하는 과정이 분리될 수 있다. 자연어 질의는 일반적인 자연어 문장과 유사하므로 사전 학습된 언어모델이 추가 학습 없이 이용



될 수 있고 자연어 요소 의미 표현이 계산될 수 있다.

- [0141] 스키마를 분석하기 위하여 각 테이블과 테이블에 속한 컬럼들에 대한 의미 표현이 먼저 계산되고, 관계 인식 트랜스포머 신경망을 이용하여 다른 테이블 정보와 상호 작용하여 최종 의미 표현이 계산될 수 있다. 이때 사용되는 테이블 간의 자세한 관계 정보는 전처리 과정에서 예측될 수 있다.
- [0142] 종래 기술에서는 하나의 모델로 자연어 질의와 스키마를 동시에 인코딩을 하였기 때문에 해당 모델을 사전학습 하려면 자연어 질의와 스키마 쌍의 학습 데이터가 필요한 문제가 있었다. 하지만, 제안하는 모델은 자연어 질의와 스키마를 인코딩하는 과정/모듈이 분리되기 때문에 스키마를 인코딩하는 모듈을 독립적으로 사전 학습할 수 있다.
- [0143] 본 발명의 실시예에서는 스키마 정보는 인터넷 등의 일반화된 데이터로부터 추출되는 테이블로부터 생성될 수 있다. 스키마 정보는 마스크 언어 모델링을 통해 비지도 학습될 수 있다. 스키마 인코더(242) 모듈만 독립적으로 사전 학습함으로써 종래 기술에서 드러난 문제점인 사전 학습된 모델과 입력 형태가 일치하지 않는 문제가 해결될 수 있다.
- [0144] 본 발명의 실시예에서는 언어 모델(244)로 자연어 질의와 테이블 정보를 각각 인코딩하기 때문에 길이 제한을 초과하지 않고 다수의 테이블로 구성된 스키마를 처리할 수 있다.
- [0145] 본 발명의 실시예에서는 테이블 단위로 먼저 인코딩하고 각 테이블의 연관성을 고려한 관계인식 트랜스포머를 활용해 스키마를 인코딩하기 때문에 종래 기술에서 드러나는 문제점인 테이블 간 연결 관계 고려 없이 상호 작용이 일어나는 문제가 발생하지 않는다.
- [0146] 본 발명의 실시예에 따르면, 자연어 질의와 스키마가 주어지면 전처리 과정을 통해서 입력 값들 간의 관계정보가 추출될 수 있다.
- [0147] 전처리 과정은 스키마 링킹 및 스키마 관계 추출을 포함할 수 있다.
- [0148] 스키마 링킹에서는 자연어 질의에 있는 단어 $w$  들과 테이블 $t$  또는 컬럼 $c$  간의 관계정보가 얻어질 수 있다.
- [0149] 스키마 관계 추출에서는 테이블과 컬럼의 포함 관계, 스키마 그래프에 있는 기본 키-외래 키(primary-foreign key), 그리고 테이블 레코드(record)를 통해서 알아내는 테이블 간의 관계가 추출될 수 있다.
- [0150] 전처리가 끝나면 자연어 질의는 사전 학습된 언어 모델(244)에 의하여 인코딩되어 자연어 처리 결과가 생성될 수 있다. 사전 학습된 스키마 인코더(242)에 의하여 스키마 관계가 얻어질 수 있다. 자연어 처리 결과는 자연어 요소 의미 표현으로서 계산되고, 스키마 관계는 스키마 의미 표현으로서 계산될 수 있다.
- [0151] 스키마 의미 표현을 계산할 때 전처리 단계에서 추출된 테이블 관계가 이용될 수 있다.
- [0152] 각 입력에 대한 의미 표현이 일차적으로 계산이 완료되면, 관계 인식 트랜스포머를 통해서 자연어 질의와 스키마 간의 정보 교환을 통해서 두 입력의 최종 의미 표현이 계산될 수 있다. 이때 전처리 단계인 스키마 링킹에서 추출한 정보에 기반하여 자연어 질의의 어떤 단어가 어떤 스키마 개체와 연관이 있는지 정보가 얻어질 수 있다.
- [0153] 이때 전처리 과정 중 일부인 스키마 링킹 과정은 공지의 기술인 KR 10-2277787호 "신경망 기반 자연어로부터 SQL 질의 번역 시 사용되는 컬럼 및 테이블을 예측하는 방법" 및 KR 10-2345568호 "자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법"에 개시된 내용의 적어도 일부인 '자연어 문장과 데이터베이스를 단어 단위로 비교하여 매칭하는 구성'을 이용하여 실행될 수 있다.
- [0154] 도 10은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 자동차 관련된 자연어 질의와 스키마가 주어졌을 때의 스키마 링킹 과정의 일 실시예를 설명하기 위한 개념도이다.
- [0155] 도 10은 자동차 관련된 자연어 질의와 스키마가 주어졌을 때의 스키마 링킹 예시를 도시한다.
- [0156] 도 10을 참조하면, 본 발명의 일 실시예의 자연어 처리 장치 및/또는 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 있는 표현으로 인코딩하는 장치에서는 입력 텍스트 전처리 과정으로 스키마 링킹 단계가 수행될 수 있다.
- [0157] 스키마 링킹은 개체 링킹(entity linking)과 상당히 유사한 문제를 해결하는 과정으로서 연결할 개체를 지식 그래프(knowledge graph)가 아닌 스키마에서 찾는다. 즉, 자연어 질의에 있는 단어를 해당 단어가 의미하고 있는 테이블 또는 컬럼과 연결시키는 것이다. 도 10에서 볼 수 있듯이 자연어 질의의 단어는 하나 이상의 컬럼/테이블과 연결이 될 수 있다. 스키마 링킹은 여러 연구에서 이미 사용되고 있는 방식으로 본 발명에서는 필요에 따

라 종래 기술의 스키마 링킹의 알고리즘을 적어도 부분적으로 이용할 수 있다.

- [0158] 스키마 링킹에서는 주어진 자연어 질의 L의 각 단어 w에 대해서 스키마에 있는 각 테이블 t들과 문자열 비교를 한다. 이때 도 10의 cylinders와 horsepower와 같이 완전히 일치하는 단어가 발견되면 이를 연결한다. 하지만 만약 완전히 일치하는 단어가 없다면 부분 일치하는 단어를 찾고 발견된 모든 개체를 연결시킨다. 이때 부분 일치하는 문자열이 아닌 단어 기준으로 연결될 수 있다. 만약 부분 일치하는 테이블이 없으면 컬럼에 대해서 같은 동작을 반복한다.
- [0159] 도 11은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 하나의 자연어 질의가 데이터베이스 내의 하위 데이터베이스(테이블) 간의 관계에 따라 서로 다른 구조화된 질의로 번역될 수 있음을 설명하기 위한 개념도이다.
- [0160] 도 11을 참조하면, 테이블 관계 추출은 두 테이블 간의 관계를 찾는 과정으로 정확한 질의 번역에 있어서 필요한 과정이다. 도 11에 도시된 바와 같이, 동일한 자연어 질의에 대해서 스키마의 테이블의 관계에 따라서 번역되는 SQL이 달라질 수 있다. SQL1의 경우는 student와 has\_pet 테이블의 관계가 one-to-one인 경우에 맞는 번역이고 one-to-many 관계일 때는 SQL2가 올바른 번역이다.
- [0161] 즉, 동일한 자연어 질의에 대해서도 student 및 pet(has\_pet) 간의 관계가 one-to-one인 경우의 구조화된 질의와 student 및 pet 간의 관계가 one-to-many인 경우의 구조화된 질의는 동일하지 않다. 따라서 올바른 구조화된 질의를 생성하기 위해서는 자연어 질의에 직접적으로 포함되지 않는 스키마 관계를 더 필요로 한다.
- [0162] 도 12는 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 데이터베이스 레코드 값을 활용하여 하위 데이터베이스(테이블) 간의 관계를 추출하는 실시예를 도시하는 개념도이다.
- [0163] 도 12를 참조하면, 본 발명의 일 실시예에서는 두 테이블 간의 관계를 추출하는 방법으로 데이터베이스 레코드 값을 활용한다.
- [0164] 기본 키로 설정된 컬럼 C1과 외래 키로 설정된 컬럼 C2가 있을 때, 테이블 레코드에서 각 컬럼에 해당하는 값들을 V1와 V2라고 한다. 이때, V2에 중복되는 값이 있으면 one-to-many, 중복되는 값이 없으면 one-to-one 관계가 추출될 수 있다.
- [0165] 또한, V1의 집합과 V2의 집합이 일치하면 mandatory 관계, 일치하지 않으면 optional 관계가 추출/예측될 수 있다.
- [0166] 도 12를 참조하면, student 테이블의 id와 has\_pet 테이블의 student\_id의 관계는 one-to-many optional인 것을 알 수 있다. has\_pet 테이블과 pet 테이블의 관계는 one-to-one mandatory이다.
- [0167] 도 13은 본 발명의 일 실시예에 따른 자연어 질의 처리 장치에서 수행되는 자연어 질의 처리 방법에서, 스키마 인코딩 예시를 도시하는 개념도이다.
- [0168] 도 13을 참조하면, 본 발명의 일 실시예에서는 의미 표현 계산 과정으로 자연어 질의 인코딩 단계와 스키마 인코딩 단계가 수행된다.
- [0169] 설명의 편의상 자연어 질의 인코딩은 사전 학습된 언어 모델 BERT를 통해서 진행하는 것으로 가정한다.
- [0170] 바이트 쌍 방식의 토큰화를 통해 자연어 문장을 토큰으로 분리한다. 이때 토큰의 어휘(vocabulary)는 BERT와 같은 것을 사용한다. 그 후, 토큰으로 변환된 자연어 질의를 BERT의 입력으로 넣어 각 토큰에 대한 의미 표현을 얻을 수 있다. 이때, 문장 L에 대해서 각 단어 w에 대한 의미 표현이 계산될 수 있다. 한 단어가 여러 개의 토큰으로 분리된 경우 각 토큰의 의미 표현을 평균하여 단어 별 의미 표현이 계산될 수 있다.
- [0171] 스키마 인코딩은 도 13과 같이 2가지 단계로 나누어진다. 첫번째로 사전 학습된 언어 모델 BERT를 사용하여 스키마에 S에 있는 각 테이블 t에 대한 의미 표현이 독립적으로 계산될 수 있다. 테이블 정보를 BERT에 넣어주기 위해서 테이블 정보를 문장으로 펼치는(flatten) 과정이 진행될 수 있다. 테이블 정보로는 테이블의 이름과 테이블에 속한 컬럼들의 이름이 활용될 수 있다. 테이블 정보를 문장으로 펼치는 과정은 다음과 같다.
- [0172] 먼저 [CLS](분류자)라는 특별 토큰(special token) 뒤에 테이블 이름을 토큰화하여 연결할 수 있다. 그 후, 테이블에 속한 모든 컬럼 이름들을 토큰화하여 테이블 이름 뒤에 연결할 수 있다. 이때 컬럼들 사이에 [SEP](분리자)라는 특별 토큰을 이용하여 컬럼들을 분리시킨다. 연결된 토큰들을 BERT에 입력하여 각 테이블과 컬럼에 대한 의미 표현을 계산하고, 전처리 단계에서 추출한 테이블 간의 관계, 및 테이블/컬럼 간의 관계 정보에 기반하

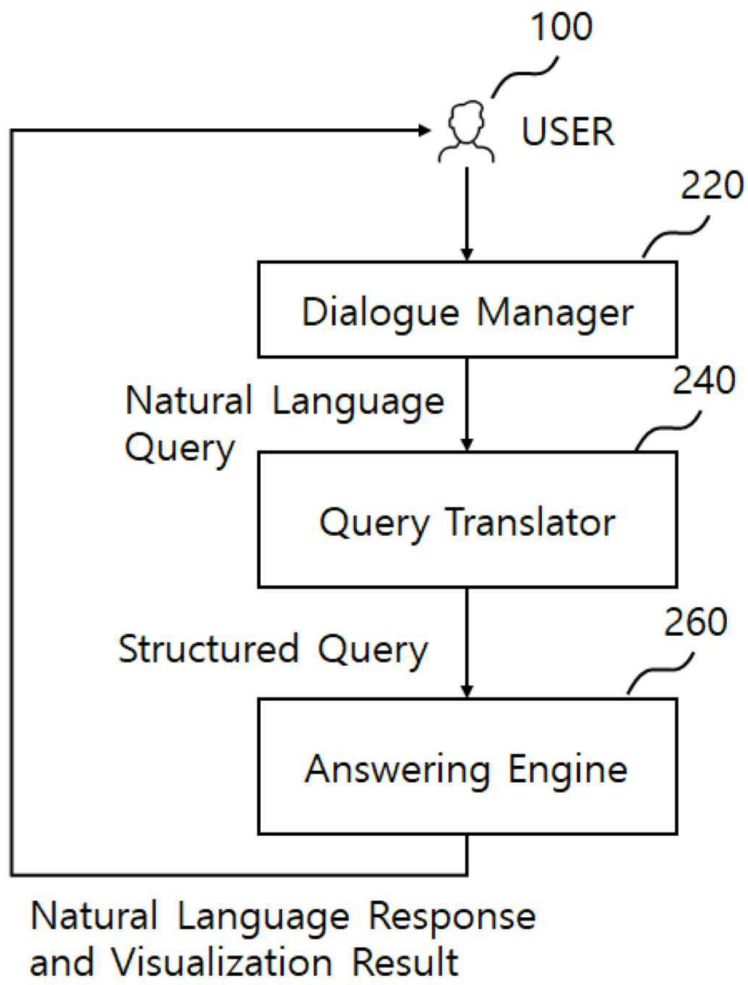
여 관계 인식 트랜스포머를 사용하여 스키마를 다시 한 번 인코딩할 수 있다.

- [0173] 본 발명의 일 실시예는 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 있는 표현으로 인코딩하는 방법, 및/또는 자연어 질의를 처리하는 방법을 구현하기 위한 컴퓨터 판독 가능한 기록매체에 저장된 컴퓨터 프로그램일 수 있다.
- [0174] 본 발명의 일 실시예는 트랜스포머 신경망 기반 자연어 질의와 관계형 데이터베이스를 의미 있는 표현으로 인코딩하는 방법, 및/또는 자연어 질의를 처리하는 방법의 프로그램을 구현하기 위한 컴퓨터 판독 가능한 기록매체일 수 있다.
- [0175] 도 1 내지 도 13의 실시예에 따르면, 테이블 간의 관계를 추출하여 올바른 구조화된 질의를 생성할 수 있다.
- [0176] 도 1 내지 도 13의 실시예에 따르면, 자연어 처리 과정의 학습과 스키마 인코딩의 학습을 분리함으로써 입력 데이터의 크기가 제한되지 않을 수 있다. 따라서 기업에서 실제로 필요로 하는 수백 개 이상의 컬럼을 이용하는 데이터베이스에 대한 자연어 질의 자동 번역 기능을 구현할 수 있다.
- [0177] 도 1 내지 도 13의 실시예에 따르면, 사전 학습된 언어 모델을 이용한 자연어 인코딩과 관계 인식 트랜스포머를 이용한 스키마 인코딩이 제안될 수 있다.
- [0178] 도 1 내지 도 13의 실시예에 따르면, 테이블 별로 언어 모델 또한 개별적으로 적용할 수 있으므로 테이블 개수에 제한 없이 인코딩이 이루어질 수 있다.
- [0179] 도 1 내지 도 13의 실시예에 따르면, 예측된 테이블 간의 관계 정보를 이용하여 구조화된 질의를 생성하기 위한 인코딩이 이루어질 수 있다.
- [0180] 도 1 내지 도 13의 실시예에 따르면, 교차 어텐션을 이용하여 스키마에서 자연어와의 관계에서 집중해야 할 하위 데이터베이스에 대한 정보를 얻을 수 있다.
- [0181] 도 1 내지 도 13의 실시예에 따르면, 교차 어텐션을 이용하여 자연어 중 어떤 단어에 집중해야 할 지에 대한 정보(어떤 단어가 어떤 스키마와 관련도가 높은 지에 대한 정보)를 얻을 수 있다.
- [0182] 도 1 내지 도 13의 실시예에 따르면, 교차 어텐션을 이용하여 자연어와 스키마 각각의 집중해야 할 요소들을 인식할 수 있으므로 구조화된 질의를 생성하는 과정을 고도화할 수 있다.
- [0183] 도 14는 본 발명의 일 실시예에 따른 도 1 내지 도 13의 과정의 적어도 일부를 수행할 수 있는 일반화된 자연어 처리 장치 또는 컴퓨팅 시스템을 도시하는 블록도이다.
- [0184] 도 1 내지 도 3의 실시예에서 대화 관리자(220), 질의 번역기(240), 및 응답 엔진(260)에 프로세서, 메모리와 전자적으로 연결되고 프로세서에 의하여 제어되거나 관리될 수 있음은 앞에서 설명한 바와 같다.
- [0185] 도 4 내지 도 13의 실시예에서도 도면 상으로는 생략되었으나 프로세서, 및 메모리가 전자적으로 각 구성 요소와 연결되고, 프로세서에 의하여 각 구성 요소의 동작이 제어되거나 관리될 수 있다.
- [0186] 본 발명의 일 실시예에 따른 자연어 처리 방법의 적어도 일부의 과정은 도 14의 컴퓨팅 시스템(1000)에 의하여 실행될 수 있다.
- [0187] 도 14를 참조하면, 본 발명의 일 실시예에 따른 컴퓨팅 시스템(1000)은, 프로세서(1100), 메모리(1200), 통신 인터페이스(1300), 저장 장치(1400), 입력 인터페이스(1500), 출력 인터페이스(1600) 및 버스(bus)(1700)를 포함하여 구성될 수 있다.
- [0188] 본 발명의 일 실시예에 따른 컴퓨팅 시스템(1000)은, 적어도 하나의 프로세서(processor)(1100) 및 상기 적어도 하나의 프로세서(1100)가 적어도 하나의 단계를 수행하도록 지시하는 명령어들(instructions)을 저장하는 메모리(memory)(1200)를 포함할 수 있다. 본 발명의 일 실시예에 따른 방법의 적어도 일부의 단계는 상기 적어도 하나의 프로세서(1100)가 상기 메모리(1200)로부터 명령어들을 로드하여 실행함으로써 수행될 수 있다.
- [0189] 프로세서(1100)는 중앙 처리 장치(central processing unit, CPU), 그래픽 처리 장치(graphics processing unit, GPU), 또는 본 발명의 실시예들에 따른 방법들이 수행되는 전용의 프로세서를 의미할 수 있다.
- [0190] 메모리(1200) 및 저장 장치(1400) 각각은 휘발성 저장 매체 및 비휘발성 저장 매체 중에서 적어도 하나로 구성될 수 있다. 예를 들어, 메모리(1200)는 읽기 전용 메모리(read only memory, ROM) 및 랜덤 액세스 메모리(random access memory, RAM) 중에서 적어도 하나로 구성될 수 있다.

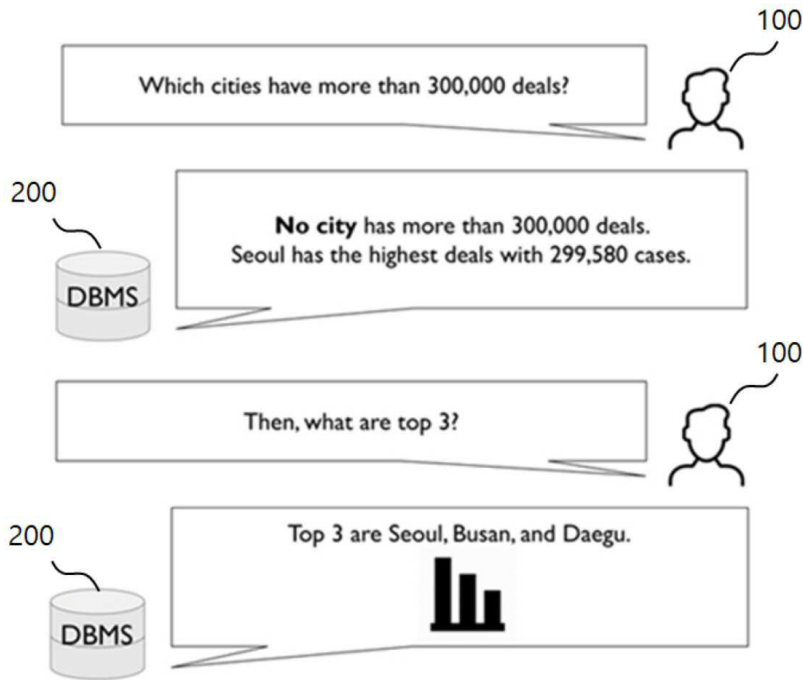
- [0191] 또한, 컴퓨팅 시스템(1000)은, 무선 네트워크를 통해 통신을 수행하는 통신 인터페이스(1300)를 포함할 수 있다.
- [0192] 또한, 컴퓨팅 시스템(1000)은, 저장 장치(1400), 입력 인터페이스(1500), 출력 인터페이스(1600) 등을 더 포함할 수 있다.
- [0193] 또한, 컴퓨팅 시스템(1000)에 포함된 각각의 구성 요소들은 버스(bus)(1700)에 의해 연결되어 서로 통신을 수행할 수 있다.
- [0194] 본 발명의 컴퓨팅 시스템(1000)의 예를 들면, 통신 가능한 데스크탑 컴퓨터(desktop computer), 랩탑 컴퓨터(laptop computer), 노트북(notebook), 스마트폰(smart phone), 태블릿 PC(tablet PC), 모바일폰(mobile phone), 스마트 워치(smart watch), 스마트 글래스(smart glass), e-book 리더기, PMP(portable multimedia player), 휴대용 게임기, 네비게이션(navigation) 장치, 디지털 카메라(digital camera), DMB(digital multimedia broadcasting) 재생기, 디지털 음성 녹음기(digital audio recorder), 디지털 음성 재생기(digital audio player), 디지털 동영상 녹화기(digital video recorder), 디지털 동영상 재생기(digital video player), PDA(Personal Digital Assistant) 등일 수 있다.
- [0195] 본 발명의 실시예에 따른 방법의 동작은 컴퓨터로 읽을 수 있는 기록매체에 컴퓨터가 읽을 수 있는 프로그램 또는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의해 읽힐 수 있는 정보가 저장되는 모든 종류의 기록장치를 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산 방식으로 컴퓨터로 읽을 수 있는 프로그램 또는 코드가 저장되고 실행될 수 있다.
- [0196] 또한, 컴퓨터가 읽을 수 있는 기록매체는 롬(rom), 램(ram), 플래시 메모리(flash memory) 등과 같이 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치를 포함할 수 있다. 프로그램 명령은 컴파일러(compiler)에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터(interpreter) 등을 사용해서 컴퓨터에 의해 실행될 수 있는 고급 언어 코드를 포함할 수 있다.
- [0197] 본 발명의 일부 측면들은 장치의 문맥에서 설명되었으나, 그것은 상응하는 방법에 따른 설명 또한 나타낼 수 있고, 여기서 블록 또는 장치는 방법 단계 또는 방법 단계의 특징에 상응한다. 유사하게, 방법의 문맥에서 설명된 측면들은 또한 상응하는 블록 또는 아이템 또는 상응하는 장치의 특징으로 나타낼 수 있다. 방법 단계들의 몇몇 또는 전부는 예를 들어, 마이크로프로세서, 프로그램 가능한 컴퓨터 또는 전자 회로와 같은 하드웨어 장치에 의해(또는 이용하여) 수행될 수 있다. 몇몇의 실시 예에서, 가장 중요한 방법 단계들의 적어도 하나 이상은 이와 같은 장치에 의해 수행될 수 있다.
- [0198] 실시예들에서, 프로그램 가능한 로직 장치(예를 들어, 필드 프로그래머블 게이트 어레이)가 여기서 설명된 방법들의 기능의 일부 또는 전부를 수행하기 위해 사용될 수 있다. 실시예들에서, 필드 프로그래머블 게이트 어레이(field-programmable gate array)는 여기서 설명된 방법들 중 하나를 수행하기 위한 마이크로프로세서(microprocessor)와 함께 작동할 수 있다. 일반적으로, 방법들은 어떤 하드웨어 장치에 의해 수행되는 것이 바람직하다.
- [0199] 이상 본 발명의 바람직한 실시 예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

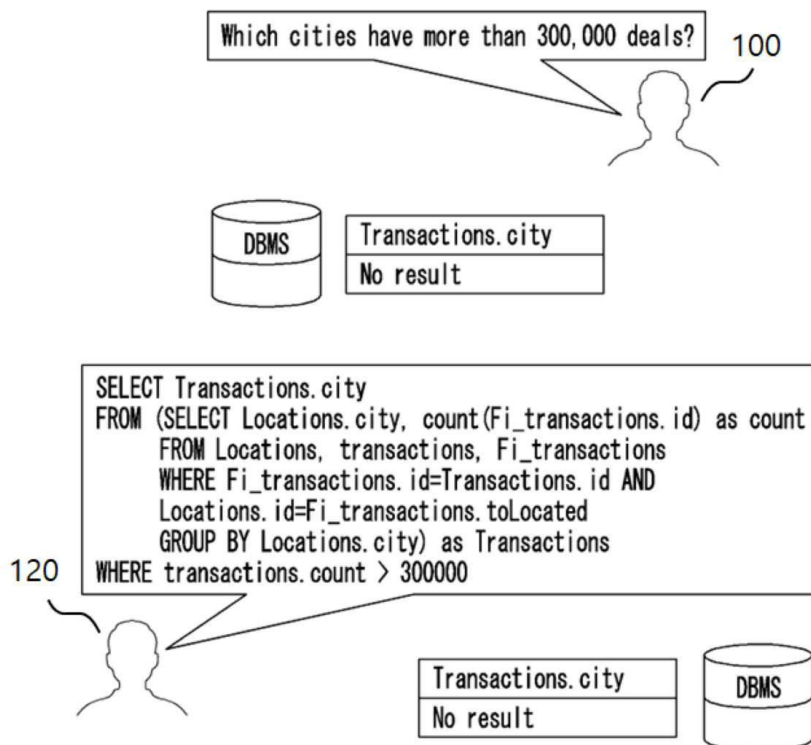
도면1



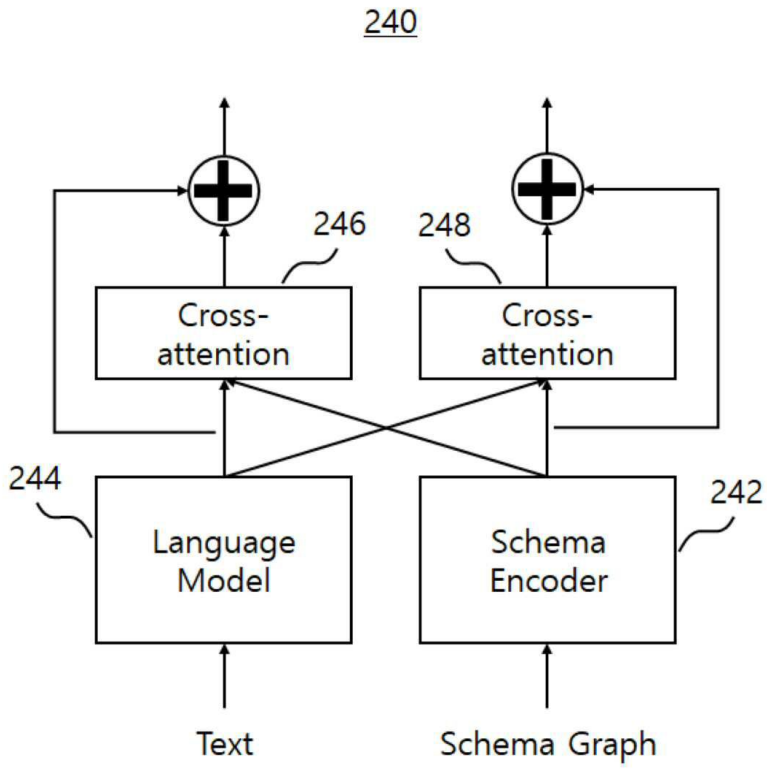
도면2



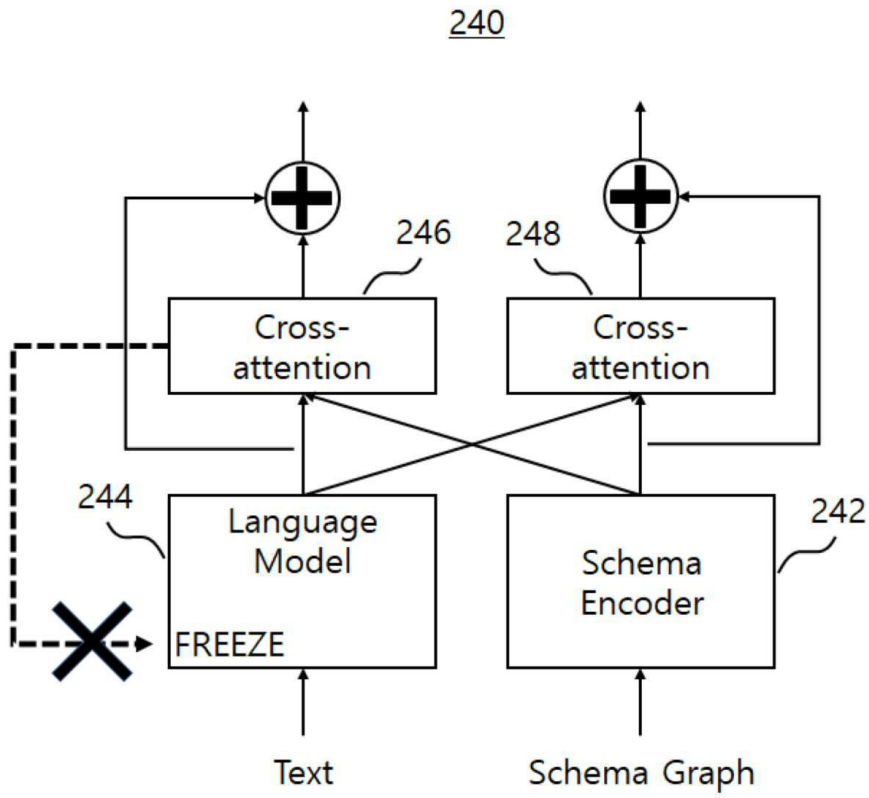
도면3



도면4

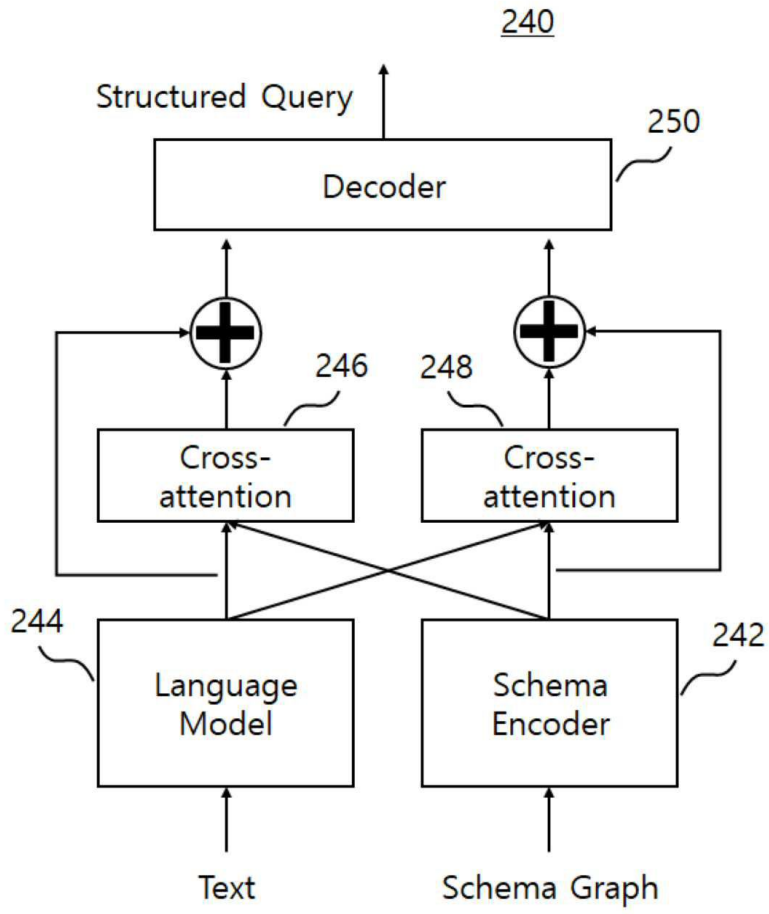


도면5

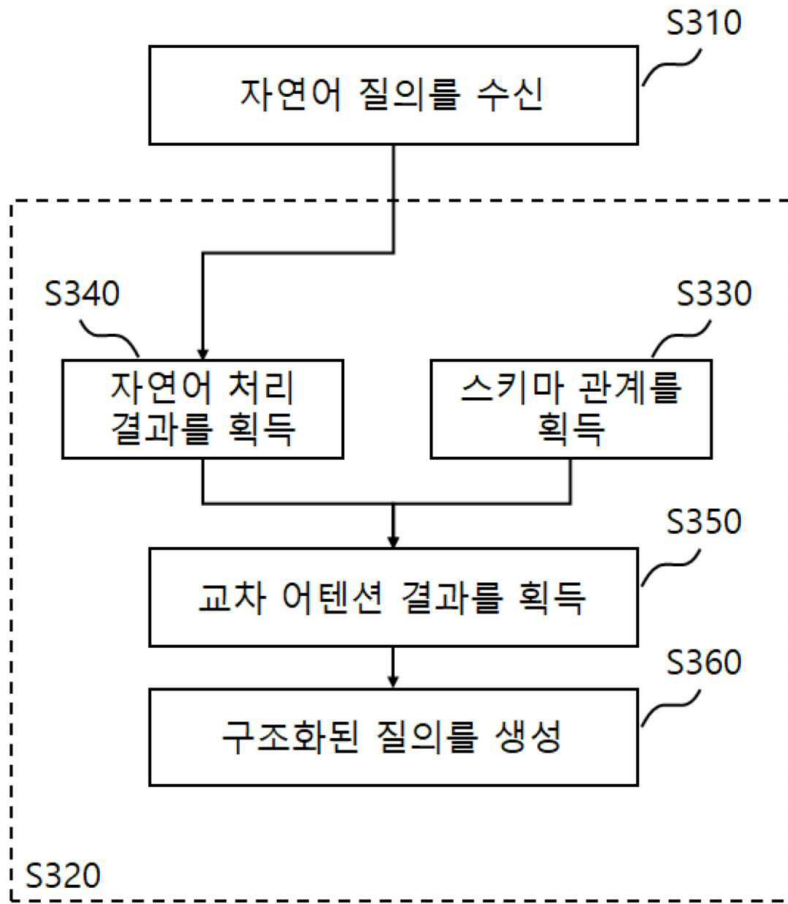




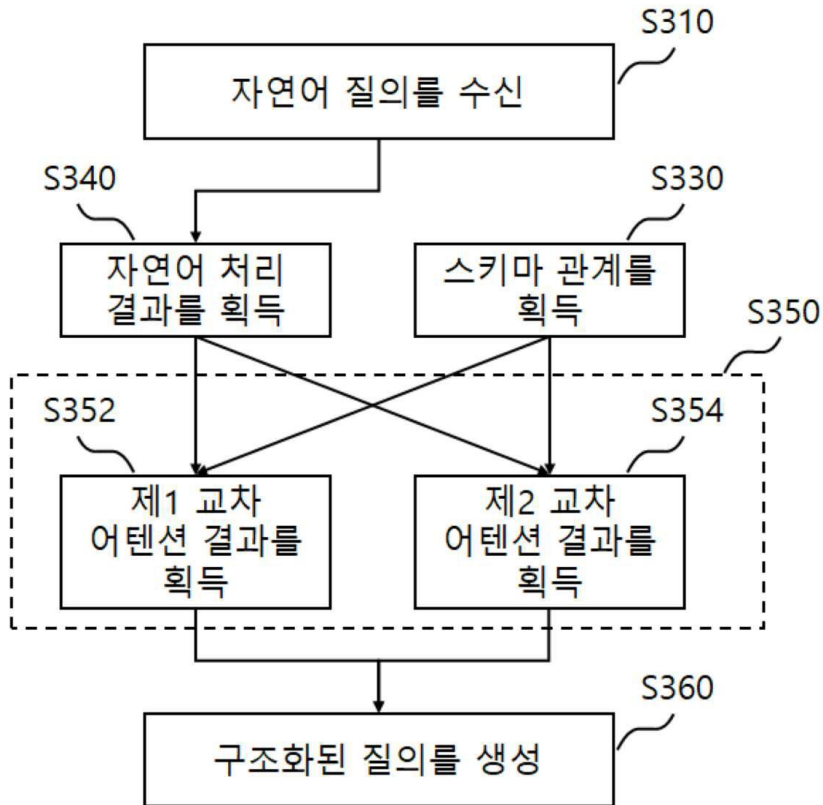
도면6



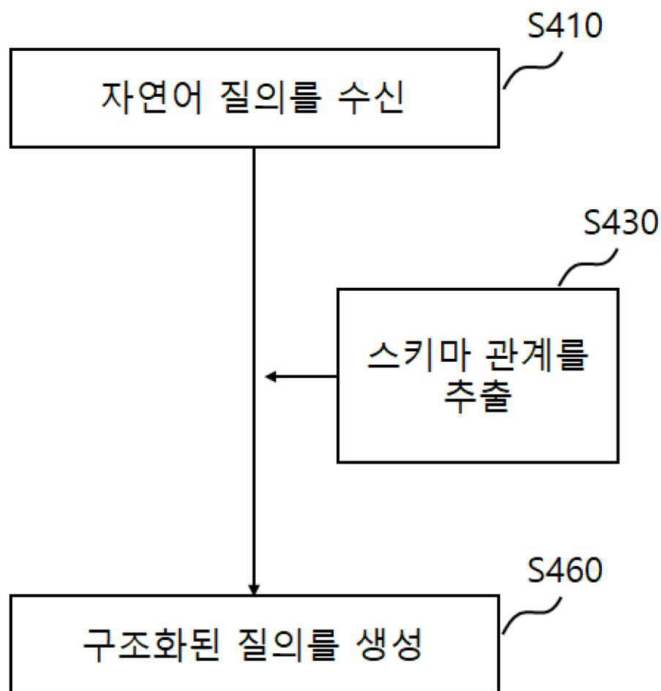
도면7



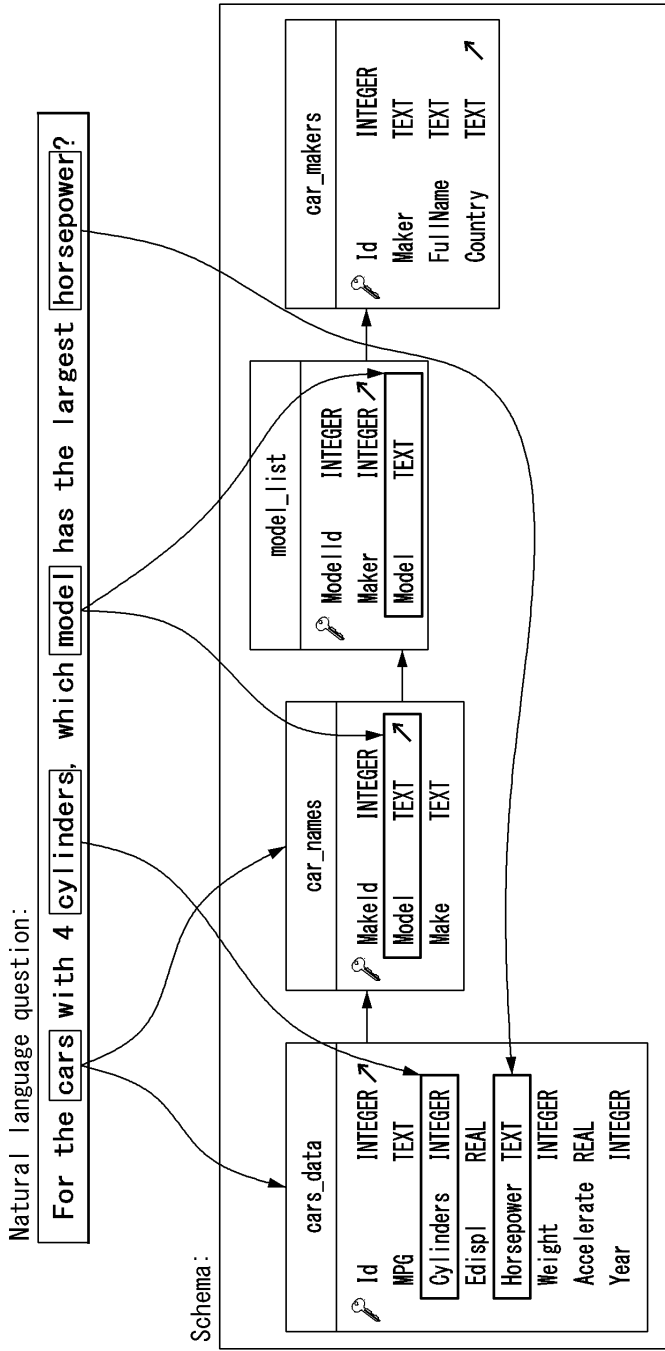
도면8



도면9

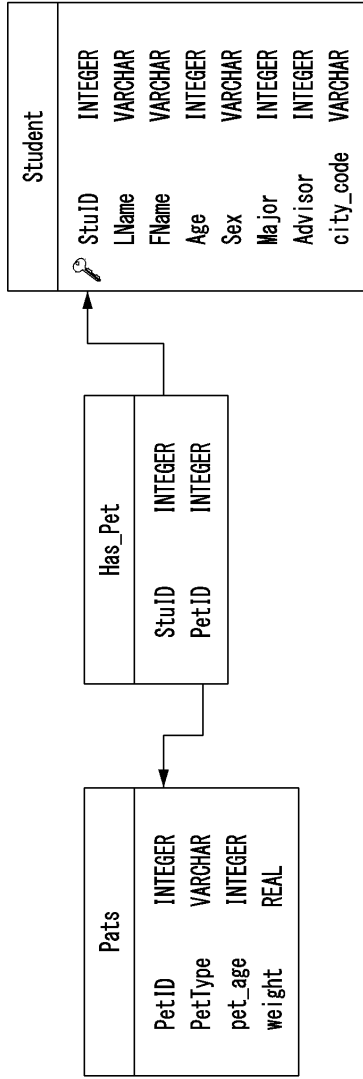


도면10



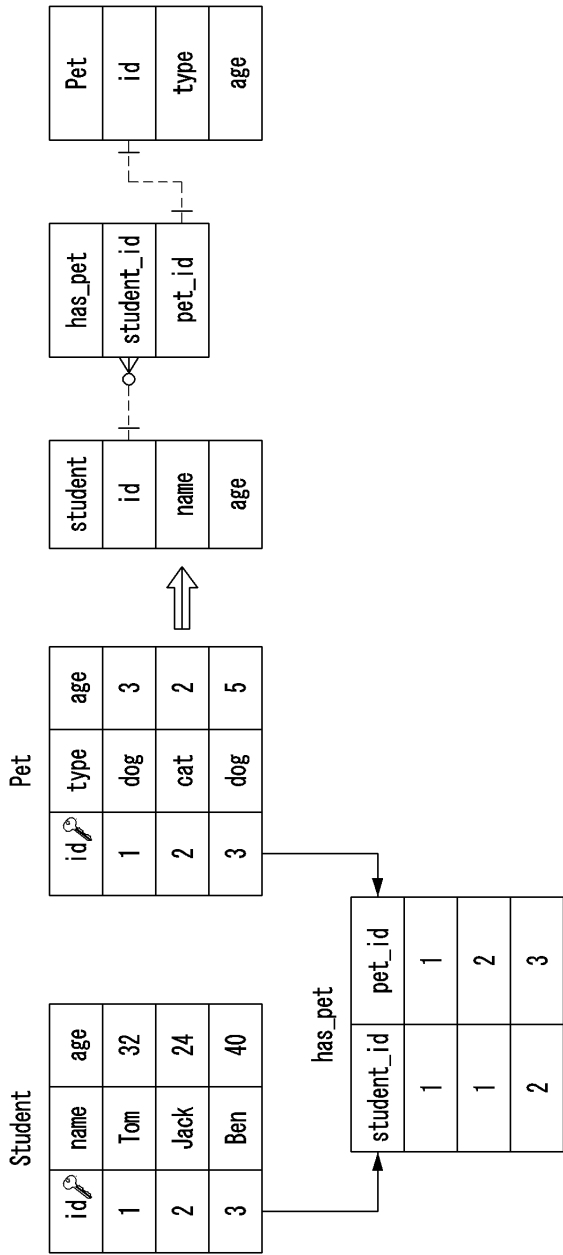
도면11

NL Question: How many students has pet?

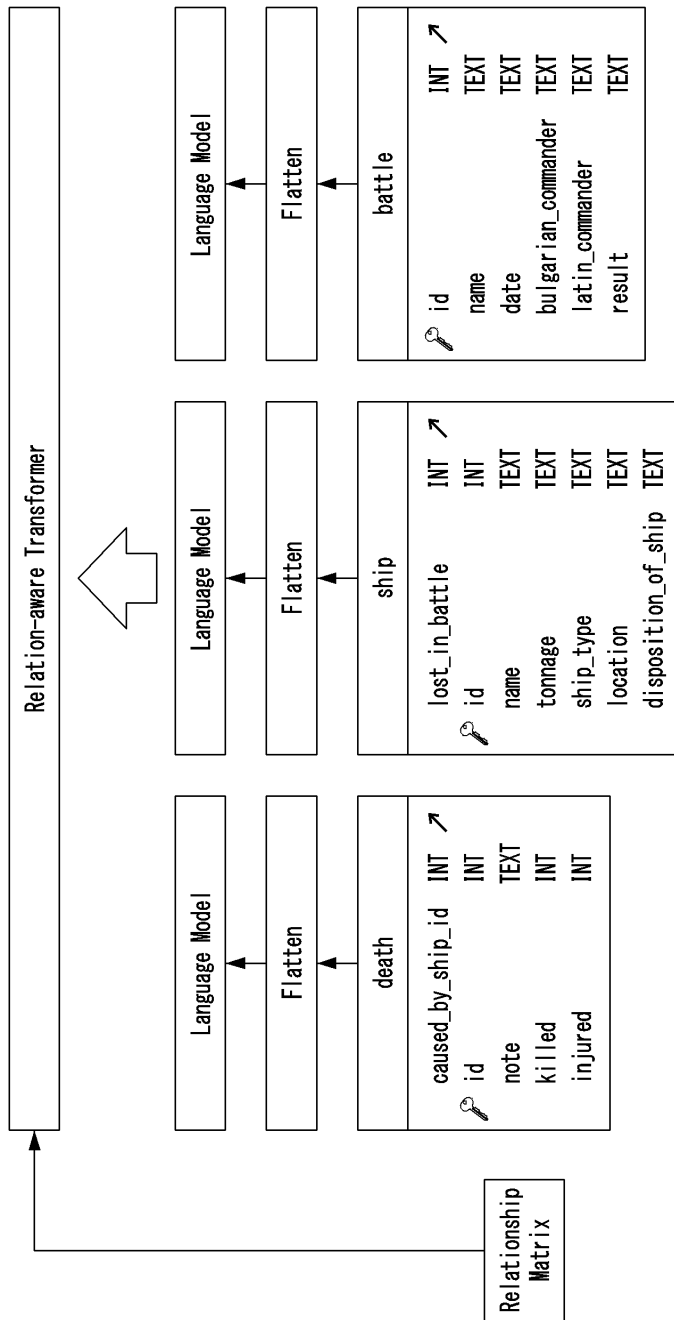


SQL1: SELECT count(\*) FROM student as T1 JOIN has\_pet as T2 on T1/id = T2.student\_id  
 SQL2: SELECT count(\*) FROM student as T1 JOIN has\_pet as T2 on T1/id = T2.student\_id GROUP BY T1.id

도면12



도면13



도면14

