



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0082727  
(43) 공개일자 2021년07월06일

(51) 국제특허분류(Int. Cl.)  
G06F 16/22 (2019.01) G06F 40/258 (2020.01)  
(52) CPC특허분류  
G06F 16/221 (2019.01)  
G06F 40/258 (2020.01)  
(21) 출원번호 10-2019-0174967  
(22) 출원일자 2019년12월26일  
심사청구일자 2019년12월26일

(71) 출원인  
포항공과대학교 산학협력단  
경상북도 포항시 남구 청암로 77 (지곡동)  
(72) 발명자  
한옥신  
경상북도 포항시 남구 청암로 77 창의IT융합공학과 (지곡동, 포항공과대학교)  
나인혁  
서울특별시 서대문구 연희로32길 48, 104동 105호(연희동, 연희동성원아파트)  
(74) 대리인  
특허법인이름리온  
(뒷면에 계속)

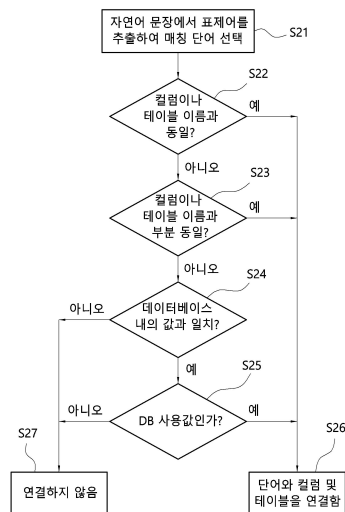
전체 청구항 수 : 총 6 항

(54) 발명의 명칭 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법

### (57) 요약

본 발명은 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법에 관한 것으로, 마이크로컴퓨터를 포함하는 컴퓨팅 장치를 이용하여 자연어 문장의 단어와 데이터베이스의 컬럼 및 테이블을 매칭시키는 방법으로서, a) 자연어 단어가 데이터베이스의 컬럼 또는 테이블 이름과 일치하는지 확인하고, 일치하는 컬럼 또는 테이블을 해당 단어에 매칭시키는 단계와, b) 상기 a) 단계에서 매칭되지 않은 단어가 있는 경우, 컬럼 또는 테이블 이름이 단어와 부분적으로 일치하는지 확인하고, 부분적으로 일치하는 경우 매칭시키는 단계와, c) 상기 b) 단계에서 부분적으로도 일치하지 않는 자연어 단어가 있는 경우, 데이터베이스의 값들과 일치여부를 확인하여 일치하면 매칭시키고, 일치하지 않으면 매칭에서 제외하는 단계를 포함한다.

대표도 - 도2



(72) 발명자

**김현지**

울산광역시 동구 월봉12길 50, C동 308호 (화정동,  
송정타워맨션3차)

**강혁규**

경상북도 포항시 남구 효자동길5번길 23, 201호(효  
자동)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711082907
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기술진흥센터
연구사업명	SW컴퓨팅산업원천기술개발
연구과제명	대화 가능하고 자동으로 튜닝하는 DBMS의 개발
기 여 율	1/1
과제수행기관명	포항공과대학교 산학협력단
연구기간	2019.02.01 ~ 2019.12.31

---

## 명세서

### 청구범위

#### 청구항 1

마이크로컴퓨터를 포함하는 컴퓨팅 장치를 이용하여 자연어 문장의 단어와 데이터베이스의 컬럼 및 테이블을 매칭시키는 방법으로서,

- a) 자연어 단어들이 데이터베이스의 컬럼 또는 테이블 이름과 일치하는지 확인하고, 일치하는 컬럼 또는 테이블을 해당 단어에 매칭시키는 단계;
- b) 상기 a) 단계에서 매칭되지 않은 단어들이 있는 경우, 컬럼 또는 테이블 이름이 단어와 부분적으로 일치하는지 확인하고, 부분적으로 일치하는 경우 매칭시키는 단계; 및
- c) 상기 b) 단계에서 부분적으로도 일치하지 않는 자연어 단어가 있는 경우, 데이터베이스의 값들과 일치여부를 확인하여 일치하면 매칭시키고, 일치하지 않으면 매칭에서 제외하는 단계를 포함하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법.

#### 청구항 2

제1항에 있어서,

상기 a) 단계에서 자연어 단어들은,

자연어 문장에서 표제어 추출을 통해 추출된 것을 특징으로 하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법.

#### 청구항 3

제1항에 있어서,

상기 a) 단계에서 컬럼 또는 테이블과 매칭된 단어는,

상기 b) 단계의 부분적 동일성 판단을 수행하지 않는 것을 특징으로 하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법.

#### 청구항 4

제1항에 있어서,

상기 a) 단계와 상기 b) 단계에서 컬럼 또는 테이블과 매칭된 단어는,

상기 c) 단계의 데이터베이스 내부 값과의 비교를 수행하지 않는 것을 특징으로 하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법.

#### 청구항 5

제1항에 있어서,

상기 c) 단계는,

c-1) 단어와 데이터베이스의 내부 값의 일치 여부를 판단하는 과정; 및

c-2) 상기 c-1) 과정의 판단결과 일치하는 경우, 해당 단어가 사전 정의된 데이터베이스 내부 값으로 적당하지 않은 단어이면 매칭하지 않는 과정을 포함하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법.

#### 청구항 6

제5항에 있어서,

상기 c-2) 과정은,

데이터베이스 내부의 값과 일치하는 단어의 품사 분석을 통해 사전 정의된 데이터베이스 내부 값으로 적당하지 않은 단어인지 판단하는 것을 특징으로 하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법에 관한 것으로, 더 상세하게는 자연어에 대한 데이터베이스의 인터페이스 실현이 용이하도록 하는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법에 관한 것이다.

### 배경 기술

[0002] 일반적으로, 자연어를 SQL 질의로 번역하는 문제는 전문 지식이 없는 사람이 관계형 데이터베이스에 질의할 수 있다는 점에서 중요한 문제이다. 최근 SQL 질의 번역 문제를 푸는 많은 규칙 및 신경망 기반 방법들이 제안되었다.

[0003] "Diptikalyan Saha, Avrilia Floratou, Karthik Sankaranarayanan, Umar Farooq Minhas, Ashish R Mittal, Fatma Özcan : ATHENA: An Ontology-Driven System for Natural Language Querying over Relational Data Stores PVLDB 9(12): 1209-1220 (2016)(선행문헌1)"에서는 자연어를 SQL 질의로 변환하는 규칙 기반 기술 ATHENA를 제안하였다. ATHENA는 전문가가 사전 정의한 온톨로지(ontology) 및 텍스트의 단어와 온톨로지의 개체를 매핑하는 사전을 이용해 자연어의 단어들을 온톨로지의 개체들에 매핑한다. 이렇게 매핑된 개체들 및 자연어를 이용해 OQL(Ontology Query Language) 질의를 생성하고, 다시 이를 SQL 질의로 변환하는 방법으로 SQL 질의 번역을 수행한다.

[0004] 이러한 방식은 온톨로지 및 사전을 전문가가 미리 만들어 주어야만 동작하는 문제가 있으며, 데이터베이스 스키마가 바뀌면 온톨로지 및 사전을 수정해야 하는 문제가 있다.

[0005] 신경망 기술의 발전과 SQL 질의 번역 데이터가 발표되어 신경망 기반 SQL 질의 번역 기술이 활발히 연구되고 있다. 최근의 신경망 기반 연구들은 크게 슬롯 채우기(slot filling)와 문맥 자유 문법 기반 방식으로 나눌 수 있다.

[0006] 슬롯 채우기 방법은 특정 템플릿을 갖는 단순한 SQL 질의를 가정하여 템플릿의 각 슬롯을 신경망을 통해 채우는 방식으로, "Victor Zhong, Caiming Xiong, Richard Socher: Seq2SQL: Generating Structured Queries from Natural Language using Reinforcement Learning CoRR abs/170900103 (2017)(선행문헌2)"와 "Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, Dragomir R Radev : TypeSQL: Knowledge-Based Type-Aware Neural Text-to-SQL Generation NAACL-HLT (2) 2018 : 588-594(선행문헌3)"이 있다.

[0007] 선행문헌2는 특정 템플릿을 갖는 단순한 SQL 질의와 자연어 및 데이터베이스의 세 쌍으로 이루어진 데이터셋인 WikiSQL을 발표하였다.

[0008] 선행문헌 3은 WikiSQL에서의 SQL 질의 번역을 처음으로 슬롯 채우기 문제로 보아 TypeSQL이라는 신경망 기반 방법을 제시했다.

[0009] 하지만, WikiSQL에서와 같이 슬롯 채우기 방식을 적용할 수 있는 SQL 질의는 매우 간단하여 현실적으로 사용되 기에는 무리가 있다. 예를 들어, 슬롯 채우기 방식들은 다양한 키워드(ORDER BY, GROUP BY 등)를 가정하지 않고, 중첩 질의를 처리할 수 없다.

[0010] 또한, 문맥 자유 문법 기반 방식의 예로는 "Wonseok Hwang, Jinyeung Yim, Seunghyun Park, Minjoon Seo: A Comprehensive Exploration on WikiSQL with Table-Aware Word Contextualization CoRR abs/190201069 (2019)(선행문헌4)"와 "Pengcheng He, Yi Mao, Kaushik Chakrabarti, Weizhu Chen : X-SQL: reinforce schema representation with context CoRR abs/190808113 (2019)(선행문헌5)"가 있다.

[0011] 위의 선행문헌4 및 선행문헌5는 많은 자연어로 사전 학습된 신경망에서의 전이 학습을 이용하는 SQLova 및 X-SQL을 제안하여 WikiSQL에서 약 90%의 높은 정확도를 얻었다.

[0012] 문맥 자유 문법 기반 방식은 SQL 질의와 같은 프로그래밍 언어의 문맥 자유 문법을 기반으로 신경망을 통해 최 좌단 유도(leftmost derivation)하는 방식으로 최종 문장을 생성해 내는 방식으로, 미리 정의한 문맥 자유 문법

이 생성할 수 있는 모든 SQL 질의를 생성할 수 있어 슬롯 채우기 방식의 한계를 극복할 수 있다.

- [0013] "Tao Yu, Michihiro Yasunaga, Kai Yang, Rui Zhang, Dongxu Wang, Zifan Li, Dragomir R Radev : SyntaxSQLNet: Syntax Tree Networks for Complex and Cross-Domain Text-to-SQL Task EMNLP 2018 : 1653-1663(선행문헌6)"은 문맥 자유 문법 기반 방식으로 SQL 질의 번역을 하는 첫 연구이다.
- [0014] 또한, "Jiaqi Guo, Zecheng Zhan, Yan Gao, Yan Xiao, Jian-Guang Lou, Ting Liu, Dongmei Zhang : Towards Complex Text-to-SQL in Cross-Domain Database with Intermediate Representation ACL (1) 2019 : 4524-4535(선행문헌7)"는 규칙 기반으로 데이터베이스 스키마와 자연어의 단어들을 연결하는 방법을 제안하였고, 또한 중간 언어인 SemQL을 제안하여 자연어에서 SemQL 질의로 번역하는 신경망 방식과 SemQL 질의를 SQL 질의로 번역하는 규칙 기반 방법을 통해 SQL 질의 번역을 수행한다.
- [0015]
- [0016] 도 1은 일반적인 관계형 데이터베이스에 대한 자연어 처리 흐름도이다.
- [0017] 도 1을 참조하면 (a)와 같이 자연어 질의의 입력에 따라 (b)와 같이 SQL 질의로 변환하고, 그 변환된 내용을 (c)와 같이 관계형 데이터베이스에서 검색하여 그 결과를 자연어로 출력하게 된다.
- [0018] 데이터 베이스 시스템의 자연어 인터페이스를 만드는 문제는 비전문가들이 데이터베이스에 쉽게 질의를 할 수 있도록 한다는 점에서 중요한 문제다. 데이터베이스 시스템 중 가장 널리 쓰이는 관계형 데이터베이스에 대한 질의는 SQL 프로그래밍 언어로 이루어진다.
- [0019] 따라서 자연어 질의를 SQL 질의로 번역함으로써 관계형 데이터베이스에 대한 자연어 인터페이스를 만들 수 있다.
- [0020] 자연어 질의를 SQL 질의로 번역하는 문제의 큰 어려움 중 하나는 자연어 질의가 데이터베이스의 어떤 컬럼 및 테이블을 참조하는지 알아내는 것이다. 이것이 어려운 이유는 데이터베이스의 컬럼 및 테이블 이름이 자연어 질의에 그대로 등장하는 것이 아니라 의역되어 나타나거나 직접적으로 나타나지 않을 수 있기 때문이다. 또한, 데이터베이스 스키마에 비슷한 이름을 가진 컬럼 및 테이블들이 여러 개가 존재할 수 있어 정확하게 SQL 질의에 쓰일 컬럼 및 테이블들을 찾아내는 것은 어려운 문제이다.
- [0021] 기존의 규칙 기반 SQL 질의 번역 방법들은 전문가가 사전에 매핑 사전을 만들어 주는 방식을 사용했지만, 이는 사전에 고려되지 않은 단어가 등장하거나, 데이터베이스 스키마가 달라지는 상황 등에서 문제가 생긴다. 또한, 모든 데이터베이스마다 전문가의 큰 노력이 드는 근본적인 문제가 있다.
- [0022] 특히 실제 데이터베이스는 자연어에는 일반적으로 등장하지 않는 특정 도메인의 단어들이 많이 등장할 수 있기 때문에 모든 상황에 대하여 매칭 사전을 만들거나 기계학습을 통해 미리 학습 시키는 방법은 매우 어려운 일이다.
- [0023] 따라서, 자연어 단어와 데이터베이스 컬럼 및 테이블을 매칭시키는 새로운 방법이 요구되고 있다.

## 발명의 내용

### 해결하려는 과제

- [0024] 본 발명이 해결하고자 하는 기술적 과제는, 자연어 문장에 등장하는 단어 여러개가 매칭되거나, 단어의 일부만이 매칭되거나, 단어가 여러 컬럼 및 테이블에 매칭되는 등 단어와 컬럼 및 테이블의 의미적 연결이 어려운 케이스에 대하여 매칭을 수행할 수 있는 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법을 제공함에 있다.

### 과제의 해결 수단

- [0025] 상기와 같은 과제를 해결하기 위한 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법은, 본 발명은 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법에 관한 것으로, 마이크로컴퓨터를 포함하는 컴퓨팅 장치를 이용하여 자연어 문장의 단어와 데이터베이스의 컬럼 및 테이블을 매칭시키는 방법으로서, a) 자연어 단어들이 데이터베이스의 컬럼 또는 테이블 이름과 일치하는지 확인하고, 일치하는 컬럼 또는 테이블을 해당 단어에 매칭시키는 단계와, b) 상기 a) 단계에서 매칭되지 않은 단어들이 있는 경우, 컬럼 또는 테이블 이름이 단어와 부분적으로 일치하는지 확인하고, 부분적으로 일치하는 경우 매칭시키는 단계와, c)

상기 b) 단계에서 부분적으로도 일치하지 않는 자연어 단어가 있는 경우, 데이터베이스의 값들과 일치여부를 확인하여 일치하면 매칭시키고, 일치하지 않으면 매칭에서 제외하는 단계를 포함한다.

[0026] 본 발명의 실시예에서, 상기 a) 단계에서 자연어 단어들은, 자연어 문장에서 표제어 추출을 통해 추출된 것일 수 있다.

[0027] 본 발명의 실시예에서, 상기 a) 단계에서 컬럼 또는 테이블과 매칭된 단어는, 상기 b) 단계의 부분적 동일성 판단을 수행하지 않을 수 있다.

[0028] 본 발명의 실시예에서, 상기 a) 단계와 상기 b) 단계에서 컬럼 또는 테이블과 매칭된 단어는, 상기 c) 단계의 데이터베이스 내부 값과의 비교를 수행하지 않을 수 있다.

[0029] 본 발명의 실시예에서, 상기 c) 단계는, c-1) 단어와 데이터베이스의 내부 값의 일치 여부를 판단하는 과정과, c-2) 상기 c-1) 과정의 판단결과 일치하는 경우, 해당 단어가 사전 정의된 데이터베이스 내부 값으로 적당하지 않은 단어이면 매칭하지 않는 과정을 포함할 수 있다.

[0030] 본 발명의 실시예에서, 상기 c-2) 과정은, 데이터베이스 내부의 값과 일치하는 단어의 품사 분석을 통해 사전 정의된 데이터베이스 내부 값으로 적당하지 않은 단어인지 판단할 수 있다.

### 발명의 효과

[0031] 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법은, 컬럼 및 테이블의 이름과 데이터베이스의 레코드들을 종합적으로 판단하여, 여러 단어가 매칭되거나 단어의 일부만이 매칭되거나, 단어가 여러 컬럼 및 테이블에 매칭될 때에도 효과적으로 연결 가능한 효과가 있다.

### 도면의 간단한 설명

[0032] 도 1은 일반적인 관계형 데이터베이스에 대한 자연어 처리 흐름도이다.

도 2는 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법의 순서도이다.

도 3은 매칭 과정의 예시도이다.

### 발명을 실시하기 위한 구체적인 내용

[0033] 이하, 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법에 대하여 첨부한 도면을 참조하여 상세히 설명한다.

[0034] 본 발명의 실시 예들은 당해 기술 분야에서 통상의 지식을 가진 자에게 본 발명을 더욱 완전하게 설명하기 위해 제공되는 것이며, 아래에 설명되는 실시 예들은 여러 가지 다른 형태로 변형될 수 있으며, 본 발명의 범위가 아래의 실시 예들로 한정되는 것은 아니다. 오히려, 이들 실시 예는 본 발명을 더욱 충실하고 완전하게 하며 당업자에게 본 발명의 사상을 완전하게 전달하기 위하여 제공되는 것이다.

[0035] 본 명세서에서 사용된 용어는 특정 실시 예를 설명하기 위하여 사용되며, 본 발명을 제한하기 위한 것이 아니다. 본 명세서에서 사용된 바와 같이 단수 형태는 문맥상 다른 경우를 분명히 지적하는 것이 아니라면, 복수의 형태를 포함할 수 있다. 또한, 본 명세서에서 사용되는 경우 "포함한다(comprise)" 및/또는 "포함하는(comprising)"은 언급한 형상들, 숫자, 단계, 동작, 부재, 요소 및/또는 이들 그룹의 존재를 특정하는 것이며, 하나 이상의 다른 형상, 숫자, 동작, 부재, 요소 및/또는 그룹들의 존재 또는 부가를 배제하는 것이 아니다. 본 명세서에서 사용된 바와 같이, 용어 "및/또는"은 해당 열거된 항목 중 어느 하나 및 하나 이상의 모든 조합을 포함한다.

[0036] 본 명세서에서 제1, 제2 등의 용어가 다양한 부재, 영역 및/또는 부위들을 설명하기 위하여 사용되지만, 이들 부재, 부품, 영역, 층들 및/또는 부위들은 이들 용어에 의해 한정되지 않음은 자명하다. 이들 용어는 특정 순서나 상하, 또는 우열을 의미하지 않으며, 하나의 부재, 영역 또는 부위를 다른 부재, 영역 또는 부위와 구별하기 위하여만 사용된다. 따라서, 이하 상술할 제1 부재, 영역 또는 부위는 본 발명의 가르침으로부터 벗어나지 않고서도 제2 부재, 영역 또는 부위를 지칭할 수 있다.

[0037] 이하, 본 발명의 실시 예들은 본 발명의 실시 예들을 개략적으로 도시하는 도면들을 참조하여 설명한다. 도면들에서, 예를 들면, 제조 기술 및/또는 공차에 따라, 도시된 형상의 변형들이 예상될 수 있다. 따라서, 본 발명의 실시 예는 본 명세서에 도시된 영역의 특정 형상에 제한된 것으로 해석되어서는 아니 되며, 예를 들면 제조상



초래되는 형상의 변화를 포함하여야 한다.

- [0038] 또한, 본 발명은 자연어를 SQL 질의로 번역(변환)하는 방법에서 컬럼 및 테이블과 자연어 단어를 연결(매칭)하는 방법에 관한 것으로, 이러한 방법의 동작 주체는 통상의 연산이 가능한 제어기와 저장장치 및 데이터의 임시 저장 가능한 메모리를 포함하는 컴퓨팅 장치이며, 예를 들어 퍼스널 컴퓨터 또는 서버를 사용할 수 있다.
- [0039] 도 2는 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법의 순서도이고, 도 3은 매칭 과정의 예시도이다.
- [0040] 도 2와 도 3을 각각 참조하면 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법은, 자연어 문장에서 표제어를 추출하여 매칭할 단어를 결정하는 단계(S21)와, 상기 단어와 컬럼 또는 테이블 이름과 비교하여 동일한지 판단하는 단계(S22)와, 상기 S22단계의 판단결과 동일하면 해당 단어와 컬럼 또는 테이블을 매칭시키는 단계(S26)와, 상기 S22단계의 판단결과 동일하지 않으면 해당 단어가 컬럼 또는 테이블의 이름과 부분적으로 동일한지 판단하여, 부분적으로 동일하면 상기 S26단계를 수행하는 단계(S23)와, 상기 S23단계의 판단결과 동일하지 않으면 해당 단어가 데이터베이스 내의 값과 일치하는지 확인하는 단계(S24)와, 상기 S24단계의 판단결과 일치하지 않으면 단어와 컬럼 및 테이블을 연결하지 않는 단계(S27)와, 상기 S24단계의 판단결과 일치하면, 해당 단어의 품사를 확인하여 데이터베이스에서 사용되는 단어이면 S26단계를 수행하고, 사용되는 단어가 아니면 S27단계를 수행하는 단계(S25)를 포함한다.
- [0041] 이하, 상기와 같이 구성되는 본 발명 자연어 단어를 데이터베이스의 컬럼 및 테이블과 연결하는 방법의 구성과 작용에 대하여 보다 상세히 설명한다.
- [0042] 먼저, S21 단계와 같이 자연어 문장에서 표제어를 추출하여 매칭할 단어를 결정한다.
- [0043] 이때, 자연어 문장은 텍스트 입력, 음성의 입력, 필기체 입력 등 다양한 방식의 입력될 수 있으며, 이미지 입력에서 텍스트를 추출한 결과일 수 있다.
- [0044] 도 3에는 "Show the names of republican people, and how many times they were at negative side?"라는 자연어 문장이 입력된 것을 예로 들었으며, 이 중 names, republican, people, negative를 표제어로 추출하여 매칭할 단어를 선택한 예를 나타낸다.
- [0045] 그 다음, S22단계에서는 상기 S21단계에서 선택된 단어를 데이터베이스의 컬럼 또는 테이블 이름과 비교하여 동일한 컬럼 또는 테이블이 있는지 판단한다.
- [0046] 상기 S22단계의 판단결과 특정한 단어가 컬럼 또는 테이블의 이름과 동일하면 S26단계와 같이 해당 단어와 동일한 이름의 컬럼 또는 테이블을 서로 연결한다.
- [0047] 도 3에는 자연어 문장에서 선택된 단어와 데이터베이스의 연결관계를 나타낸다.
- [0048] 좀 더 구체적으로 자연어 문장에서 선택된 단어인 'people'은 데이터베이스의 테이블 이름인 'people'과 일치하며, 따라서 해당 단어와 테이블을 서로 연결한다.
- [0049] 이때의 비교는 선택된 모든 단어와 데이터베이스의 모든 컬럼 및 테이블을 순차적으로 비교하게 되며, 단어 people에 대해서 테이블 이름이 people인 테이블이 연결된다.
- [0050] 또한, 자연어 문장에서 선택된 단어 'negative'는 테이블 이름이 debate\_people인 테이블의 컬럼 중 하나인 'negative'와 일치되며, 따라서 단어 'negative'와 컬럼 'negative'를 서로 연결한다.
- [0051] 그 다음, 상기 S22단계의 판단결과 동일하지 않으면 해당 단어가 컬럼 또는 테이블의 이름과 부분적으로 동일한지 판단한다.
- [0052] 이때, 이미 컬럼 또는 테이블과 연결된 단어인 'people'과 'negative'에 대해서는 다시 부분적 동일 여부를 판단하지 않는다.
- [0053] 상기 S22단계의 판단결과 부분적으로 동일한 컬럼 또는 테이블이 있으면 S26단계를 수행하여, 특정 단어와 부분적으로 동일한 컬럼 또는 테이블을 매칭시킨다.
- [0054] 도 3에서 단어 'names'는 이름이 people인 테이블의 컬럼 'name'과 부분적으로 동일하며, 따라서 S26단계를 수행하여 단어 'names'는 컬럼 'name'에 매칭한다.
- [0055] 그 다음, 상기 S23단계의 판단결과 동일하지 않으면 S24단계와 같이 해당 단어가 데이터베이스 내의 값과 일치

하는지 확인한다.

- [0056] 이때, 앞서 동일성과 부분적 동일성의 판단에 의해 매칭이 이루어진 단어에 대해서는 데이터베이스 내의 값과 일치하는지 확인하지 않는다.
- [0057] 즉, 'people', 'negative' 및 'names'에 대해서는 데이터베이스 내의 값과 일치 여부를 확인하지 않는다.
- [0058] 상기 S24단계의 판단결과 일치하지 않으면 단어와 컬럼 및 테이블을 연결하지 않고 종료한다.
- [0059] 반대로 상기 S24단계의 판단결과 특정한 단어와 데이터베이스의 내부 값이 일치하면, S25단계와 같이 해당 단어의 품사를 확인하여 데이터베이스에서 사용되는 단어이면 S26단계를 수행하고, 사용되는 단어가 아니면 S27단계를 수행한다.
- [0060] 상기 S25단계는 데이터베이스 내의 값으로 적당하지 않은 단어를 사전에 정의하여, 값이 일치하더라도 연결하지 않도록 하는 것이다.
- [0061] 예를 들어 'determiner', 'preposition' 등 데이터베이스에서 사용하는 값들은 자연어 문장에서 추출한 단어와 일치되더라도 의미가 없는 것으로, 이를 단어와 데이터베이스의 연결관계로 판단하지 않도록 함을 의미한다.
- [0062] 도 3에서 자연어 문장에서 추출된 단어 'republican'은 이름이 'people'인 테이블의 'party' 컬럼의 데이터 값이며, 따라서 단어 'republican'과 데이터베이스 값 'republican'을 상호 연결한다.
- [0063] 이와 같은 과정을 통해 본 발명은 자연어 문장에서 추출한 단어가 컬럼 및 테이블의 이름과 데이터베이스의 값들과의 연관성을 종합적으로 판단할 수 있으며, 여러 단어가 매칭되는 것을 피하고, 단어의 일부만이 매칭되는 경우에도 단어와 테이블 또는 컬럼을 상호 연결할 수 있는 특징이 있다.
- [0064] 또한, 단어가 여러 컬럼 또는 테이블에 매칭되는 것을 방지하여 자연어 단어들이 의미를 정확하게 찾아낼 수 있다.
- [0065] 본 발명은 상기 실시예에 한정되지 않고 본 발명의 기술적 요지를 벗어나지 아니하는 범위 내에서 다양하게 수정, 변형되어 실시될 수 있음은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 있어서 자명한 것이다.

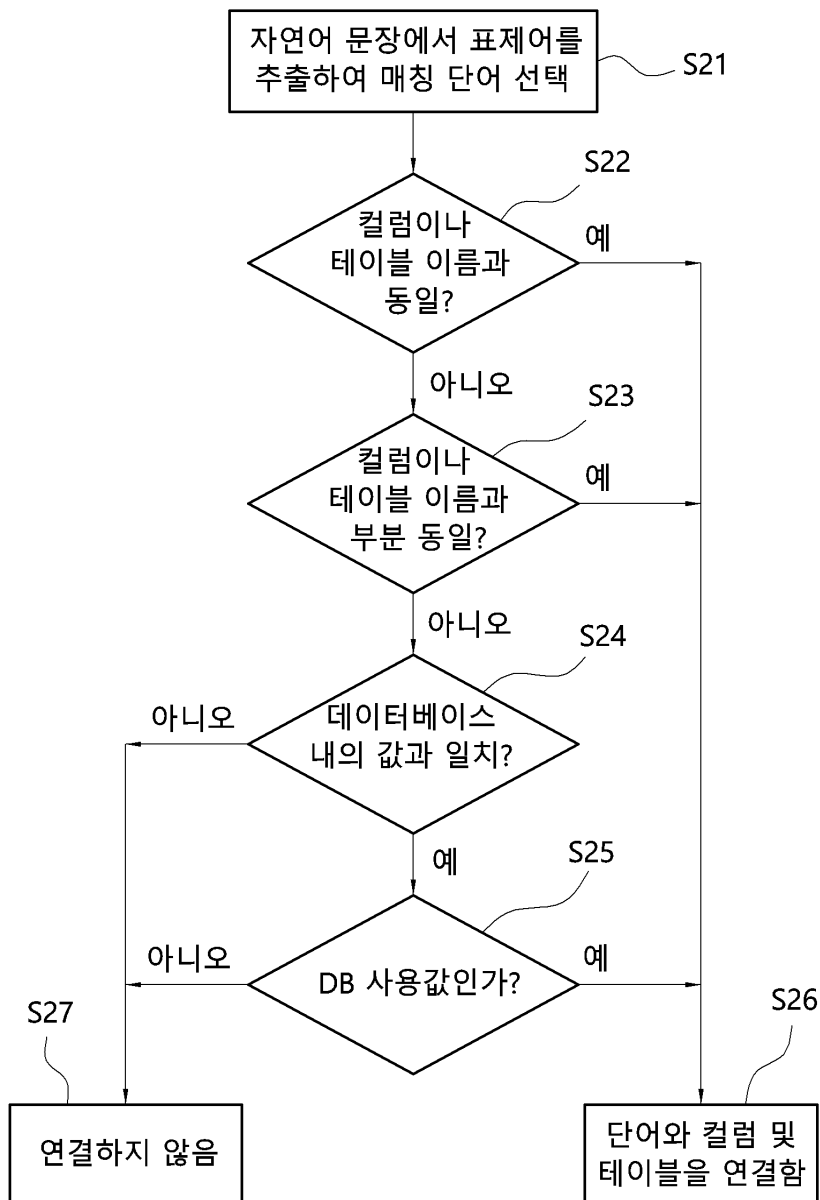
## 도면

### 도면1





도면2



도면3

