

(11) 공개번호 10-2021-0086820
(43) 공개일자 2021년07월09일

(71) 출원인
주식회사 에비드넷
경기도 성남시 분당구 판교역로 182, 2층(삼평동, 한국반도체산업협회)
강병곤
경기도 수원시 영통구 에듀타운로 101, 108동 1009호 (이의동, 에듀하임1309오피스텔)
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
김혜진
경기도 성남시 분당구 정자일로 156번길 12, A동 602호(정자동, 타임브릿지)
김하영
서울특별시 서대문구 연세로 50, 새천년관 420호(신촌동)
(뒷면에 계속)
(74) 대리인
특허법인비엘티

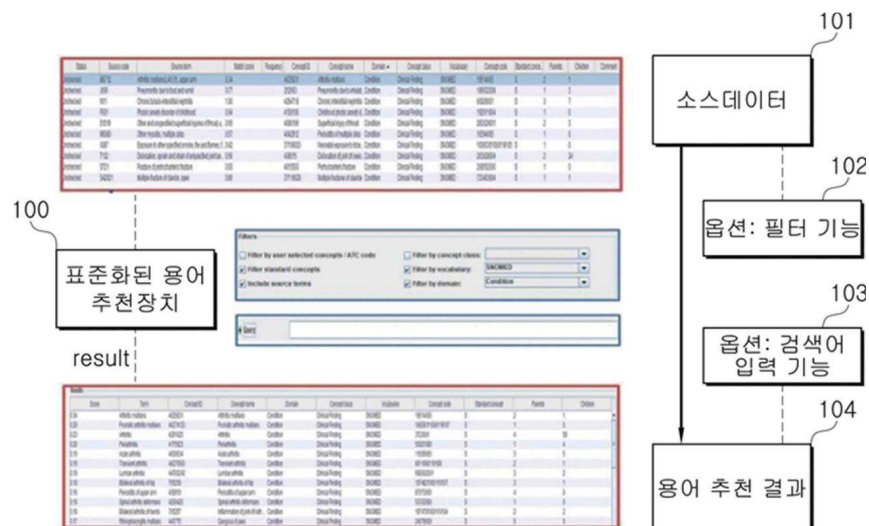
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 위계정보를 이용한 표준화된 용어 추천 방법 및 장치

(57) 요약

일 실시 예에 따른 위계정보를 이용한 표준화된 용어 추천 방법은, 입력된 소스데이터에서 단어 벡터를 추출하는 단계; 추출된 단어 벡터를 인코더를 통해 문장벡터로 변환하는 단계; 상기 문장벡터로 변환된 소스데이터와 표준 용어에 대해 기계학습을 기초로 유사도를 판단하는 단계; 및 상기 판단된 유사도를 기반으로 상기 입력된 소스데이터에 매핑되는 표준용어를 추천하는 단계를 포함하고, 상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정될 수 있다.

대표도 - 도1



(72) 발명자

강병곤

경기도 수원시 영통구 에듀타운로 101, 108동 100
9호 (이의동, 에듀하임1309오피스텔)

조수연

경기도 수원시 영통구 중부대로 271번길 27-9, 10
1동 707호 (원천동, 주공아파트)

박애란

경기도 안산시 단원구 예술대학로 105, 604동 208
호(고잔동, 주공6단지아파트)

이유림

경기도 용인시 수지구 상현로 101 111동 701호 (상현동, 상현마을수지센트럴아이파크)

명세서

청구범위

청구항 1

위계정보를 이용한 표준화된 용어 추천 방법에 있어서,
 입력된 소스데이터에서 단어 벡터를 추출하는 단계;
 추출된 단어 벡터를 인코더를 통해 문장벡터로 변환하는 단계;
 상기 문장벡터로 변환된 소스데이터와 표준용어에 대해 기계학습을 기초로 유사도를 판단하는 단계; 및
 상기 판단된 유사도를 기반으로 상기 입력된 소스데이터에 매핑되는 표준용어를 추천하는 단계를 포함하고,
 상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정되는 것인, 방법.

청구항 2

제 1항에 있어서,
 상기 보정되는 유사도는, 상기 위계정보에서 상위 계층에 존재하는 용어와 대비하여 하위 계층에 존재하는 용어에는 패널티를 부여하는 것인, 방법.

청구항 3

제 1항에 있어서,
 상기 보정되는 유사도는, 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값이 적용되는 것인, 방법.

청구항 4

제 1항에 있어서,
 상기 보정되는 유사도는, 상기 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값이 적용되는 것인, 방법.

청구항 5

제 1항에 있어서,
 상기 보정되는 유사도는, 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값 및 상기 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값에 각각 소정 가중치가 적용되어 결정되는 제3 보정값이 적용되는 것인 방법.

청구항 6

하드웨어인 컴퓨터와 결합되어, 제1항 내지 제5항 중 어느 한 항의 방법을 실행시키기 위하여 매체에 저장된, 위계정보를 이용한 표준화된 용어 추천 프로그램.

청구항 7

하나 이상의 프로세서; 및
 상기 하나 이상의 프로세서에 의한 실행 시, 상기 하나 이상의 프로세서가 연산을 수행하도록 하는 명령들이 저장된 하나 이상의 메모리를 포함하고,
 상기 하나 이상의 프로세서에 의해 수행되는 상기 연산은,
 입력된 소스데이터에서 단어 벡터를 추출하는 연산;

추출된 단어 벡터를 인코더를 통해 문장벡터로 변환하는 연산;

상기 문장벡터로 변환된 소스테이터와 표준용어에 대해 기계학습을 기초로 유사도를 판단하는 연산; 및

상기 판단된 유사도를 기반으로 상기 입력된 소스테이터의 용어에 매핑되는 표준용어를 추천하는 연산을 포함하고,

상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정되는 것인, 표준화된 용어 추천 장치.

청구항 8

제 7항에 있어서,

상기 보정되는 유사도는, 상기 위계정보에서 상위 계층에 존재하는 용어와 대비하여 하위 계층에 존재하는 용어에는 패널티를 부여하는 것인, 표준화된 용어 추천 장치.

청구항 9

제 7항에 있어서,

상기 보정되는 유사도는, 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값이 적용되는 것인, 표준화된 용어 추천 장치.

청구항 10

제 7항에 있어서,

상기 보정되는 유사도는, 상기 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값이 적용되는 것인, 표준화된 용어 추천 장치.

발명의 설명

기술 분야

[0001] 본 발명은 위계정보를 이용한 표준화된 용어 추천 방법 및 장치에 관한 것이다. 보다 상세하게는 의료용어 또는 번역용어를 표준 매핑하는 과정에서 위계정보를 기반으로 상위/하위 용어에 가중치를 적용하여 처리 정확도를 향상시킨 표준화된 용어를 추천하는 방법 및 장치에 관한 것이다.

배경 기술

[0002] 환자가 의료기관에 와서 퇴원을 할 때까지 여러 가지 단계에 따른 서비스(진료받고, 검사받고, 진단하고, 처치 받는 등)를 받게 되는데, 각각의 단계는 일일이 기록되고 데이터베이스에 저장되어 관리된다. 따라서, 이렇게 수집된 데이터를 활용하여 의료 서비스의 수준을 향상시키고자 하는 노력이 지속되고 있다. 이러한 노력의 일환으로 최근에는, 서로다른 의료기관에서 발생한 데이터들을 통합 및 교환함으로써 다양한 의료 연구활동 등을 시도하고 있다.

[0003] 그런데, 각각의 의료기관의 시스템 및 데이터베이스 형태가 서로 상이하기 때문에 다기관 연구를 하기 위한 통합 데이터베이스를 구축하기 어려운 문제가 있었다. 이러한 문제를 극복하기 위해 다양한 표준용어 체계가 정의되고 활용되고 있으나 이 역시 목적에 따라 하나의 체계로 통일한 경우이나 활용이 가능하다. 따라서, 개별기관의 용어와 표준용어, 표준용어 체계간에도 1:1의 명확한 매핑이 정의되어 있지 않은 것이 용어 체계간의 매핑을 어렵게 하는 현실이다.

[0004] 따라서, 각 의료기관에서 서로 다르게 구성된 데이터를 하나의 데이터로 통합하기 위해서는 여전히 전문가의 수작업이 필요하며, 많은 시간과 노력이 들 뿐만 아니라 분류하는 사람마다 시각이 달라 표준 용어를 정하기 어려운 경우도 많다. 따라서, 이러한 기법을 발전시키기 위한 연구가 많이 필요하다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 공개특허공보 제10-1897080호, 2018.09.04

발명의 내용

해결하려는 과제

- [0006] 본 발명이 해결하고자 하는 과제는 각 의료기관에서 서로 다르게 구성된 데이터를 하나의 데이터로 통합하기 위해 표준용어를 추천하는 방법을 제공하는 것이다.
- [0007] 따라서 일 실시 예에 따른 표준화된 용어 추천 방법은, 의료용어의 위계정보를 기반으로 상위/하위 용어에 가중치를 적용하여 처리 정확도를 향상시킨 표준화된 의료용어 추천하는 방법 및 장치를 개시한다.
- [0008] 본 발명이 해결하고자 하는 과제들은 이상에서 언급된 과제로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

- [0009] 상술한 과제를 해결하기 위한 본 발명의 일 면에 따른 위계정보를 이용한 표준화된 용어 추천 방법은, 위계정보를 이용한 표준화된 용어 추천 방법에 있어서, 입력된 소스데이터에서 단어 벡터를 추출하는 단계; 추출된 단어 벡터를 인코더를 통해 문장벡터로 변환하는 단계; 상기 문장벡터로 변환된 소스데이터와 표준용어에 대해 기계학습을 기초로 유사도를 판단하는 단계; 및 상기 판단된 유사도를 기반으로 상기 입력된 소스데이터에 매핑되는 표준용어를 추천하는 단계를 포함하고, 상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정되는 것일 수 있다.
- [0010] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 상위 계층에 존재하는 용어와 대비하여 하위 계층에 존재하는 용어에는 패널티를 부여하는 것일 수 있다.
- [0011] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정 값이 적용되는 것일 수 있다.
- [0012] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정 값이 적용되는 것일 수 있다.
- [0013] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값 및 상기 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값에 각각 소정 가중치가 적용되어 결정되는 제3 보정 값이 적용되는 것일 수 있다.
- [0014] 상술한 과제를 해결하기 위한 본 발명의 다른 일 면에 따른 표준화된 용어 추천 장치는 하나 이상의 프로세서; 및 상기 하나 이상의 프로세서에 의한 실행 시, 상기 하나 이상의 프로세서가 연산을 수행하도록 하는 명령들이 저장된 하나 이상의 메모리를 포함하고, 상기 하나 이상의 프로세서에 의해 수행되는 상기 연산은, 입력된 소스데이터에서 단어 벡터를 추출하는 연산; 추출된 단어 벡터를 인코더를 통해 문장벡터로 변환하는 연산; 상기 문장벡터로 변환된 소스데이터와 표준용어에 대해 기계학습을 기초로 유사도를 판단하는 연산; 및 상기 판단된 유사도를 기반으로 상기 입력된 소스데이터의 용어에 매핑되는 표준용어를 추천하는 연산을 포함하고, 상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정되는 것일 수 있다.
- [0015] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 상위 계층에 존재하는 용어와 대비하여 하위 계층에 존재하는 용어에는 패널티를 부여하는 것일 수 있다.
- [0016] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정 값이 적용되는 것일 수 있다.
- [0017] 또한, 상기 보정되는 유사도는, 상기 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정 값이 적용되는 것일 수 있다.
- [0018] 이 외에도 본 발명의 기타 구체적인 사항들은 상세한 설명 및 도면들에 포함되어 있다.

발명의 효과

- [0019] 본 발명의 일 실시 예에 따른 표준화된 용어 추천방법은, 서로 다른 의료기관에서 생성된 데이터를 통합하기 위해, 딥러닝 기반으로 각각 표준매핑된 의료용어를 추천함으로써, 신뢰도가 높으면서도 신속하게 통합 데이터를 생성할 수 있도록 도울 수 있다.
- [0020] 또한, 본 발명의 일 실시 예에 따른 표준화된 용어 추천방법은, 의료용어뿐 아니라 기계번역 등 다양한 언어가 매핑되는 분야에서도, 위계정보를 기반으로 상위/하위 용어에 가중치를 적용하여 처리속도를 향상시킴과 동시에 신뢰도가 높은 데이터를 생성할 수 있다.
- [0021] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

- [0022] 도 1은 일 실시 예에 따른 표준화된 용어 추천 방법의 일 예를 설명하기 위한 도이다.
- 도 2는 일 실시 예에 따른 표준화된 용어 추천 장치의 동작을 설명하기 위한 도이다
- 도 3은 일 실시 예에 따른 단어벡터 및 문장벡터를 생성하는 일 예를 설명하기 위한 도이다.
- 도 4는 일 실시 예에 따른 표준화된 용어 추천 방법을 설명하기 위한 순서도이다.
- 도 5는 일 실시 예에 따른 위계정보를 설명하기 위한 도이다.
- 도 6은 일 실시 예에 따른 위계정보를 반영하지 않은 표준화된 용어를 추천 하는 화면을 설명하기 위한 도이다.
- 도 7은 일 실시 예에 따른 위계정보를 반영하는 방법의 일 예를 설명하기 위한 도이다.
- 도 8은 일 실시 예에 따른 위계정보를 반영하는 방법의 다른 예를 설명하기 위한 도이다.
- 도 9는 일 실시 예에 따른 위계정보를 반영하는 방법의 또 다른 예를 설명하기 위한 도이다.
- 도 10은 일 실시 예에 따른 표준화된 용어 추천장치의 내부 구성을 개략적으로 나타낸 블록 도이다.

발명을 실시하기 위한 구체적인 내용

- [0023] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시 예들을 참조하면 명확해질 것이다. 그러나, 본 발명은 이하에서 개시되는 실시예들에 제한되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술 분야의 통상의 기술자에게 본 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.
- [0024] 본 명세서에서 사용된 용어는 실시 예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다. 명세서에서 사용되는 "포함한다(comprises)" 및/또는 "포함하는(comprising)"은 언급된 구성요소 외에 하나 이상의 다른 구성요소의 존재 또는 추가를 배제하지 않는다. 명세서 전체에 걸쳐 동일한 도면 부호는 동일한 구성 요소를 지칭하며, "및/또는"은 언급된 구성요소들의 각각 및 하나 이상의 모든 조합을 포함한다. 비록 "제1", "제2" 등이 다양한 구성요소들을 서술하기 위해서 사용되나, 이들 구성요소들은 이들 용어에 의해 제한되지 않음은 물론이다. 이들 용어들은 단지 하나의 구성요소를 다른 구성요소와 구별하기 위하여 사용하는 것이다. 따라서, 이하에서 언급되는 제1 구성요소는 본 발명의 기술적 사상 내에서 제2 구성요소일 수도 있음은 물론이다.
- [0025] 다른 정의가 없다면, 본 명세서에서 사용되는 모든 용어(기술 및 과학적 용어를 포함)는 본 발명이 속하는 기술 분야의 통상의 기술자에게 공통적으로 이해될 수 있는 의미로 사용될 수 있을 것이다. 또한, 일반적으로 사용되는 사전에 정의되어 있는 용어들은 명백하게 특별히 정의되어 있지 않는 한 이상적으로 또는 과도하게 해석되지 않는다.
- [0026] 공간적으로 상대적인 용어인 "아래(below)", "아래(beneath)", "하부(lower)", "위(above)", "상부(upper)" 등은 도면에 도시되어 있는 바와 같이 하나의 구성요소와 다른 구성요소들과의 상관관계를 용이하게 기술하기 위해 사용될 수 있다. 공간적으로 상대적인 용어는 도면에 도시되어 있는 방향에 더하여 사용시 또는 동작시 구성요소들의 서로 다른 방향을 포함하는 용어로 이해되어야 한다. 예를 들어, 도면에 도시되어 있는 구성요소를 뒤집을 경우, 다른 구성요소의 "아래(below)"또는 "아래(beneath)"로 기술된 구성요소는 다른 구성요소의 "위

(above)"에 놓여질 수 있다. 따라서, 예시적인 용어인 "아래"는 아래와 위의 방향을 모두 포함할 수 있다. 구성 요소는 다른 방향으로도 배향될 수 있으며, 이에 따라 공간적으로 상대적인 용어들은 배향에 따라 해석될 수 있다.

- [0027] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세하게 설명한다.
- [0028] 도 1은 일 실시 예에 따른 표준화된 용어 추천 방법의 일 예를 설명하기 위한 도이다.
- [0029] 도 1을 참조하면, 일 실시 예에 따른 표준화된 용어 추천장치(100)는 새로운 표준용어로 매핑을 하고자 하는 소스데이터(101, 의료용어 테이블, 리스트 등)를 입력 받으면, 대상 도메인의 전체 표준용어 리스트에서, 일 실시 예에 따른 표준화된 용어 추천 방법의 학습모델을 기반으로 가장 높은 매칭추천점수를 갖는 용어부터 순서대로 n개의 용어를 후보로 하는 용어 추천결과(104)를 제공할 수 있다.
- [0030] 한편, 의료용어의 경우에는 서로 다른 용어를 하나의 데이터로 통합하는 과정에서 의미가 서로 다름에도 유사한 단어를 포함하는 용어로 매칭되는 경우가 많아, 신뢰도가 다른 분야에 비해 상당히 낮은 편이며, 나아가 신뢰도를 높이기 위해 전문가의 수작업이 필요하게 되어 많은 시간이 소요되기가 쉽다. 따라서, 일 실시 예에 따른 표준화된 용어 추천장치(100)는 위계정보를 기반으로 상위/하위 용어에 가중치를 적용하여 보다 신뢰도 높은 표준용어를 사용자에게 순서대로 제공한다. 따라서, 표준매핑 용어 선정 과정의 처리속도를 향상 시키고 또 신뢰도를 향상시킬 수 있다. 또한, 서로 다른 용어의 체계간 위계를 고려할 수도 있기 때문에, 제공되는 추천 매핑용어에 대한 체계 상의 정보를 추가로 제공하여 정확도를 제고할 수 있다.
- [0031] 또한, 일 실시 예에 따른 표준화된 용어 추천 장치(100)는 사용자의 편의를 향상시키기 위해 제1 옵션인 필터기능(102) 및 제2 옵션인 검색어 입력 기능(103)을 통해 각 체계에 맞게 커스터마이징 된 추천 결과를 확인할 수도 있다.
- [0032] 한편, 일 실시 예에 따른 표준화된 용어 추천 장치(100)가 입력 소스데이터에 매칭되는 표준화된 용어를 추천하는 방법을 이하, 도 2를 참조하여 상술한다.
- [0033] 도 2는 일 실시 예에 따른 표준화된 용어 추천 장치(100)의 내부 동작을 설명한다.
- [0034] 도 2를 참조하면, 소스용어(201)와 표준용어(202)는 기계학습을 위해 입력되는 데이터를 의미할 수 있다. 예컨대, 표준용어와 매핑되기 전 데이터가 소스용어(201)로 입력되고 표준매핑이 완료된 데이터가 표준용어(202)로 각각 입력되어 기계학습을 하는데 이용될 수 있다. 물론, 기계학습이 완료된 이후에는, 소스데이터 입력에 대해 기계학습된 표준 용어들이 표준매핑되어 출력되게 된다.
- [0035] 우선, 일 실시 예에 따른 표준화된 용어 추천장치(100)는 입력된 소스데이터(203)에서 단어를 추출한다. 여기서 입력되는 소스데이터는 자연어 형태일수도 있고 테이블 형태일 수도 있다.
- [0036] 다음으로, 일 실시 예에 따른 표준화된 용어 추천장치(100)는 소스데이터(203)에서 단어 벡터(204)를 추출한다.
- [0037] 예를 들면, 도 3의(a)에 도시된 것과 같이, 입력되는 소스용어(201)들은 쪼개어 벡터로 변환될 수 있다. 이때, 용어를 벡터화 하게되면 비슷한 용어일수록 벡터간의 거리가 가깝게 나타날 수 있다.
- [0038] 다시 도 2를 참조하면, 변환된 단어벡터(204)들은 인코더(205)를 통해 문장벡터(206)로 생성될 수 있다. 그리고 생성된 문장벡터(206)의 컨볼루션/풀링(207)동작과 완전연결 레이어(Fully-Connected Layer, 208)를 통해 최종적으로 소스용어(201)와 표준용어(202)간의 유사도(p, 209)를 획득할 수 있다.
- [0039] 예를 들면, 도 3의(b)는 벡터로 변환된 용어들의 조합인 문장 전체를 벡터로 변환하여 유사도를 판단하는 일례를 도시한다. 이때, 기계학습을 위해 표준 매핑하기 전의 소스용어(201)와 매핑 후의 표준용어(202)를 모두 입력으로 넣어 라벨(label)을 예측할 수 있다. 특히 라벨은 실제 매핑 타겟 여부에 따라 2개의 클래스(class)로 나뉘질 수 있다.
- [0040] 한편, 도2에 도시되듯이 소스용어(201)에 매핑되는 표준용어(202)는 각각의 위계정보(200)를 포함하고 있다. 일 실시 예에 따른 위계정보(200) 개념(concept)끼리의 거리에 대한 정보를 포함하고 있다. 따라서 거리가 가까울수록 의미가 가까워 진다. 일 실시 예에 따른 표준화된 용어 추천방법은 이러한 위계정보(200)를 함께 기계학습 시킴으로서 정확도를 높일 수 있다.
- [0041] 이러한 위계정보(200)와 유사도의 상관관계는 다음의 수학적 식 1과 같이 나타날 수 있다.

[0042] [수학식 1]

$$BCE(x) = -\frac{1}{N} \sum_{i=1}^N (y \log(\hat{p}) + (1 - y) \log(1 - \hat{p})) \text{penalty}$$

[0043]

[0044] 여기서, BCE(x)는 학습모델의 유사도를 판단하기 위한 비용함수이며,

[0045] P는 소스용어와 표준용어간의 유사도를 나타내는 확률값, penalty는 위계정보를 반영한 값으로, 예를들면, 위계 거리정도에 따라 특정 표준용어의 경우 penalty가 더 크도록 할수있다. 이러한, 거리는 코사인 유사도(COSINE SIMILARITY)를 기반으로 결정될 수 있다. 또한, PENALTY는 결정이 필요한 값으로 지정될 수 있고, 계산을 통해 자동으로 설정될 수도 있다.

[0046] 한편, 이러한 위계정보를 이용한 표준화된 용어 추천모델의 학습 성능을 높이기 위해서는 다음 4가지의 학습 데이터가 모두 이용될 수 있으며 이들의 조합을 활용하는 등의 변형도 가능하다.

[0047] 첫번째 학습 데이터는, 진짜로 유사한 쌍이 입력되는 것이다. 두번째 학습 데이터는 서로 상이한 쌍을 입력하는 것이다. 세번째 학습데이터는 penalty를 반영하기 위해 의도적으로 위계정보중에서 자손(DESEDENT)에 해당되는 부분을 빼고 쌍으로 입력한 뒤에, 향후 penalty 값에 해당 값을 곱한 형태로 최적화 하는 것이다. 마지막으로 위계정보에서 조상(ancestor)에 해당하는 값을 빼고 penalty 값을 입력하여 최적화 할 수도 있다.

[0048] 따라서, 일 실시예에 따른 표준화된 용어 추천장치(100)를 이용하면, 고도화된 기계학습 방식을 통해 사용자는 여러 의료기관에서 서로 다르게 구성된 데이터를 하나의 데이터로 통합하기 위해 표준 용어를 학습하고, 쉽게 선택할 수 있도록 추천 할 수 있다.

[0049] 도 4는 일 실시예에 따른 표준화된 용어 추천 방법을 설명하기위한 순서도이다.

[0050] 단계 S400에서, 일 실시예에 따른 표준화된 용어 추천 방법은 입력된 소스데이터에서 단어 벡터를 추출한다.

[0051] 단계 S410에서, 일 실시 예에 따른 표준화된 용어 추천 방법은 추출된 단어 벡터를 인코더를 통해 문장벡터로 변환한다.

[0052] 단계 S420에서, 일 실시 예에 따른 표준화된 용어 추천 방법은 상기 문장벡터로 변환된 소스데이터와 표준용어에 대해 기계학습을 기초로 유사도를 판단한다. 한편, 상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정 될 수 있다. 예를 들면, 보정되는 유사도는, 위계정보에서 상위 계층에 존재하는 용어와 대비하여 하위 계층에 존재하는 용어에는 패널티를 부여하는 형태로 보정될 수 있다.

[0053] 예를 들면, 패널티 값으로 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값이 적용될 수 있다.

[0054] 다른 예로, 패널티 값으로 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정될 수도 있다.

[0055] 또 다른예로, 패널티 값으로 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값 및 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값에 각각 소정 가중치가 적용되어 결정되는 제3 보정값이 적용될 수도 있다.

[0056] 단계 S430에서, 일 실시 예에 따른 표준화된 용어 추천 방법은 상기 판단된 유사도를 기반으로 상기 입력된 소스데이터에 매핑되는 표준용어를 추천한다.

[0057] 예를 들면, 추천될 후보 표준용어는 보정된 유사도를 기반으로 점수가 채점되며, 채점순으로 입력 용어에 대응되는 2개 이상의 표준용어가 추천될 수 있다.

[0058] 이하 도 5내지 도 9를 참조하여, 위계정보를 이용하여 패널티 값을 조정함으로써, 유사도를 보정하는 방법을 상세히 설명한다.

[0059] 도 5는 일 실시 예에 따른 위계정보를 설명하기 위한 도이다.

[0060] 도 5를 참조하면, 소스용어(501)에 대해서 하나의 표준용어(502)가 선택되면, 매핑된 표준용어(502) 위계 정보 상 상위 개념에 해당하는 용어(Ancessor)에게는 패널티가 +1이 되고, 하위개념에게는 패널티가 +2가 되도록 설정될 수 있다. 여기서, 위계정보에서 상위에 존재하는 용어는 상위개념에 해당하고, 하위에 존재하는 용어는 하위개념에 해당하므로 상위개념에 해당하는 용어가 하위개념에 대비 표준용어에 가까울 가능성이 높기 때문이다.

- [0061] 한편, 이러한 위계정보는 정리된 자료를 기반으로 사용자가 입력할 수도 있고, 기계학습을 통해 학습될 수도 있다. 그렇다면 왜 이러한 위계정보를 반영해야 하는지 도 6을 참조하여 부연설명한다.
- [0062] 도 6은 일 실시 예에 따른 위계정보를 반영하지 않은 경우 표준화된 용어를 추천 하는 화면을 설명하기 위한 도이다.
- [0063] 도 6은, 입력 소스테이터가 "Other benign neoplasms of skin of other and unspecified parts of face"이고, 정답으로 매칭되는 표준용어가 "Benign neoplasm of skin of face"일 때 위계정보가 없는 경우 추천 데이터 10개를 제공한 일 예의 화면을 도시한다.
- [0064] 랭크(501)를 참조하면, 위계정보를 반영하지 않은 경우 1위부터 10위 내에
- [0065] 정답으로 매칭되는 표준용어가 존재 하지 않는다. 즉, 1위부터 10위 내에 특정 값들은 정답과 거리가 있는 값이기에 수정이 필요하다. 따라서, 관계(502) 항목과 거리(503)항목에 도식된 값들을 이용하여 하위 개념에 해당하고 거리가 1 이상 떨어져 있는 값들(순위 2, 4, 5, 6에 해당하는 값들)에 대해 패널티를 부여하고, 위계에서 직접적으로 연결되지 않는 값들(순위 1,3, 7-10)의 패널티를 상대적으로 더 크게 부여하면, 여기에 해당하는 용어들은 10위 밖으로 밀려나고 정답에 매칭되는 표준용어가 순위 내에 배치 될 수 있다. 따라서 일 실시 예에 따른 위계정보를 이용한 표준화된 용어 추천방법을 사용하는 사용자가 보다 신뢰도 높은 데이터를 접할 수 있게 된다.
- [0066] 이하 도 7내지 도 9를 참조하여 위계정보를 반영하는 방법들을 설명한다.
- [0067] 도 7은 일 실시 예에 따른 위계정보를 반영하는 방법의 일 예를 설명하기 위한 도이다.
- [0068] 도 7을 참조하면, 일 실시 예에 따른 표준화된 용어 추천 방법은, 패널티 값으로 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값이 적용될 수 있다.
- [0069] 예를들면, 수학식 1에 정의된 penalty 값은 거리에 따른 상수(700)로 미리 정해져 있을 수 있다.
- [0070] 또한, 상수 a, b, c, d는 각각 0.5 (positive), 0.7 (negative with 거리 1), 0.9 (negative with 거리 2), 1(negative)로 정해 질 수 있다. 여기서 positive는 위계 정보상 상위에 있는 용어를 의미하며 negative는 하위에 위치하는 값일 수 있다.
- [0071] 도 8은 일 실시 예에 따른 위계정보를 반영하는 방법의 다른 예를 설명하기 위한 도이다.
- [0072] 도 8을 참조하면, 일 실시 예에 따른 표준화된 용어 추천 방법은,, 패널티 값으로 위계정보에서 매핑되는 표준용어(801)와 위계정보 상에 위치하는 용어(802)들 간의 코사인 유사도를 기반으로 결정될 수도 있다.
- [0073] 예컨대 위계 정보에서 설명(description)의 코사인 유사도와 같은 정답과 위계상에 있는 것들과의 유사도를 기반으로 패널티를 정의할 수도 있다. 이렇게 결정된 유사도를 점수(u)를 그대로 패널티로 이용할수도 있고, 나아가 함수화(800) 하여 사용할 수도 있다. 예를들면 함수화된 유사도 점수는 다음의 수학식 2와 같이 나타낼 수 있다.
- [0074] [수학식 2]
- [0075]
$$f(u) = au + b$$
- [0076] 여기서, u는 유사도 점수이며, a와 b는 실수이다.
- [0077] 도 9는 일 실시예에 따른 위계정보를 반영하는 방법의 또 다른 예를 설명하기 위한 도이다.
- [0078] 도 9를 참조하면, 일 실시 예에 따른 표준화된 용어 추천 방법은, 패널티 값으로 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값 및 위계정보에서 매핑되는 표준용어(901)와 위계정보 상에 위치하는 용어(902)들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값에 각각 소정 가중치가 적용되어 결정되는 제3 보정값(900)이 적용될 수도 있다.
- [0079] 즉, 도 9에 도식된 실시예에서는 도 7 및 도 8에서 상술한 방법이 함께 적용되어 보다 정밀하게 위계 정보를 반영할 수 있다.
- [0080] 도 10은 일 실시 예에 따른 의료용어 추천장치의 내부 구성을 개략적으로 나타낸 블록 도이다.
- [0081] 도 10을 참조하면 일 실시 예에 따른 표준화된 용어 추천장치(100)는, 하나 이상의 메모리(110)와 프로세서

(120)를 포함할 수 있다. 표준화된 용어 추천장치 (100)의 동작은 메모리(110)에 저장된 프로그램을 프로세서 (120)를 통해서 실행시킴으로 수행될 수 있다.

[0082] 일 실시 예에 따른 프로세서(120)에서 수행되는 연산은, 입력된 소스테이터에서 단어 벡터를 추출하는 연산; 추출된 단어 벡터를 인코더를 통해 문장벡터로 변환하는 연산; 상기 문장벡터로 변환된 소스테이터와 표준용어에 대해 기계학습을 기초로 유사도를 판단하는 연산; 및 상기 판단된 유사도를 기반으로 상기 입력된 소스테이터의 용어에 매핑되는 표준용어를 추천하는 연산을 수행할 수 있다. 한편, 상기 유사도는 상기 표준용어의 위계정보를 기반으로 보정 될 수 있다. 예를 들면, 보정되는 유사도는, 위계정보에서 상위 계층에 존재하는 용어와 대비하여 하위 계층에 존재하는 용어에는 패널티를 부여하는 형태로 보정될 수 있다.

[0083] 예를 들면, 패널티 값으로 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값이 적용될 수 있다.

[0084] 다른 예로, 패널티 값으로 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정될 수도 있다.

[0085] 또 다른예로, 패널티 값으로 상기 위계정보에서 획득한 위계 거리를 기반으로 결정되는 제1 보정값 및 위계정보에서 매핑되는 표준용어와 위계정보 상에 위치하는 용어들 간의 코사인 유사도를 기반으로 결정되는 제2 보정값에 각각 소정 가중치가 적용되어 결정되는 제3 보정값이 적용될 수도 있다.

[0086] 한편, 상술한 일 실시 예에 따른 표준화된 용어 추천장치(100)는 하나 이상의 프로세서(120) 및/또는 하나 이상의 메모리(110)를 포함할 수 있다. 또한, 도 10에는 설명의 편의를 위해 일 실시 예에 따른 위계정보를 이용한 표준화된 용어 추천장치를 설명하기 위한 필수적인 구성만이 도시되어 있지만, 입력부(미도시) 및 출력부(미도시)등 발명에 해당하는 장치를 구성하기 위한 종래의 여러 구성들이 유기적으로 결합되어 도 1 내지 도 10에서 상술한 동작을 수행할 수도 있다.

[0087] 또한, 메모리(110)는, 휘발성 및/또는 비휘발성 메모리를 포함할 수 있다. 하나 이상의 메모리(110)는, 하나 이상의 프로세서(120)에 의한 실행 시, 하나 이상의 프로세서(120)가 연산을 수행하도록 하는 명령들을 저장할 수 있다. 본 개시에서, 프로그램 내지 명령은 메모리(110)에 저장되는 소프트웨어로서, 서버(100)의 리소스를 제어하기 위한 운영체제, 어플리케이션 및/또는 어플리케이션이 장치의 리소스들을 활용할 수 있도록 다양한 기능을 어플리케이션에 제공하는 미들 웨어 등을 포함할 수 있다.

[0088] 하나 이상의 프로세서(120)는, 소프트웨어(예: 프로그램, 명령)를 구동하여 프로세서(120)에 연결된 장치(100)의 적어도 하나의 구성요소를 제어할 수 있다. 또한 프로세서(120)는 본 개시와 관련된 다양한 연산, 처리, 데이터 생성, 가공 등의 동작을 수행할 수 있다. 또한 프로세서(120)는 데이터 등을 메모리(110)로부터 로드하거나, 메모리(110)에 저장할 수 있다

[0089] 일 실시예에서는, 장치(100)의 구성요소들 중 적어도 하나가 생략되거나, 다른 구성요소가 추가될 수 있다. 또한 추가적으로(additionally) 또는 대체적으로(alternatively), 일부의 구성요소들이 통합되어 구현되거나, 단수 또는 복수의 개체로 구현될 수 있다.

[0090] 본 발명의 실시예와 관련하여 설명된 방법 또는 알고리즘의 단계들은 하드웨어로 직접 구현되거나, 하드웨어에 의해 실행되는 소프트웨어 모듈로 구현되거나, 또는 이들의 결합에 의해 구현될 수 있다. 소프트웨어 모듈은 RAM(Random Access Memory), ROM(Read Only Memory), EPROM(Erasable Programmable ROM), EEPROM(Electrically Erasable Programmable ROM), 플래시 메모리(Flash Memory), 하드 디스크, 착탈형 디스크, CD-ROM, 또는 본 발명이 속하는 기술 분야에서 잘 알려진 임의의 형태의 컴퓨터 판독가능 기록매체에 상주할 수도 있다.

[0091] 이상, 첨부된 도면을 참조로 하여 본 발명의 실시예를 설명하였지만, 본 발명이 속하는 기술분야의 통상의 기술자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며, 제한적이지 않은 것으로 이해해야만 한다.

부호의 설명

[0092] 100 : 표준화된 용어 추천 장치

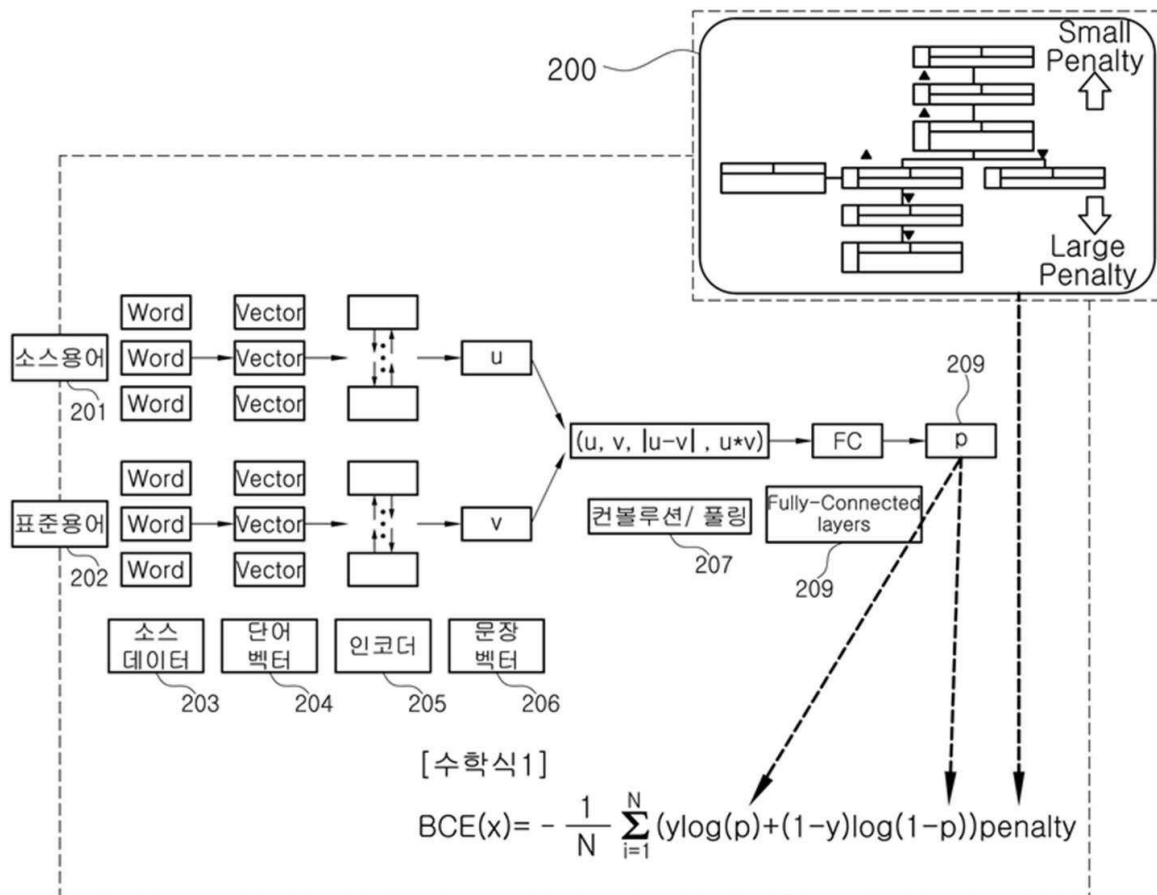
200 : 위계 정보

도면

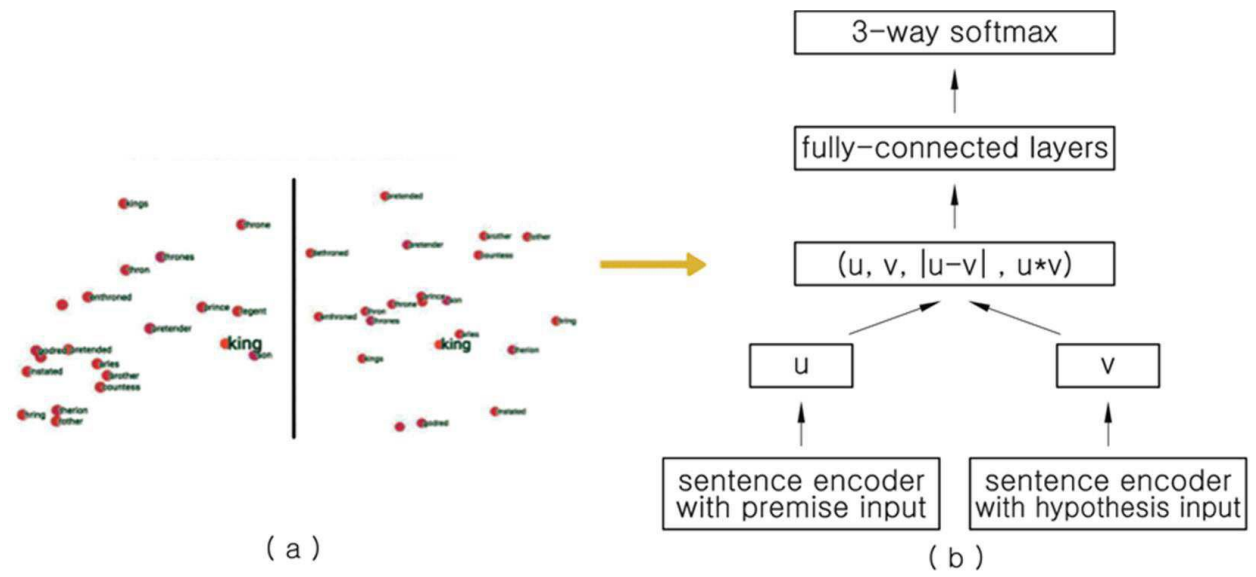
도면1



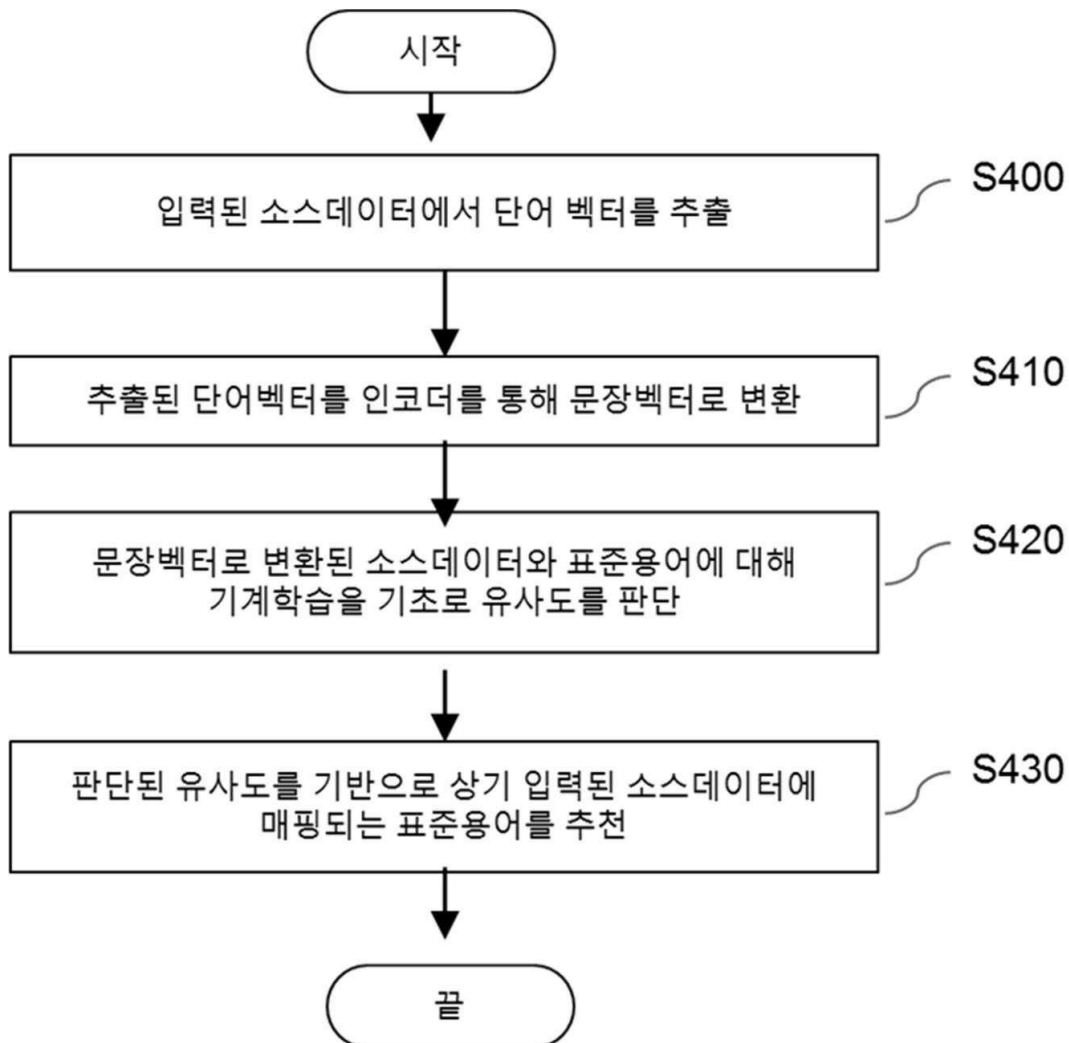
도면2



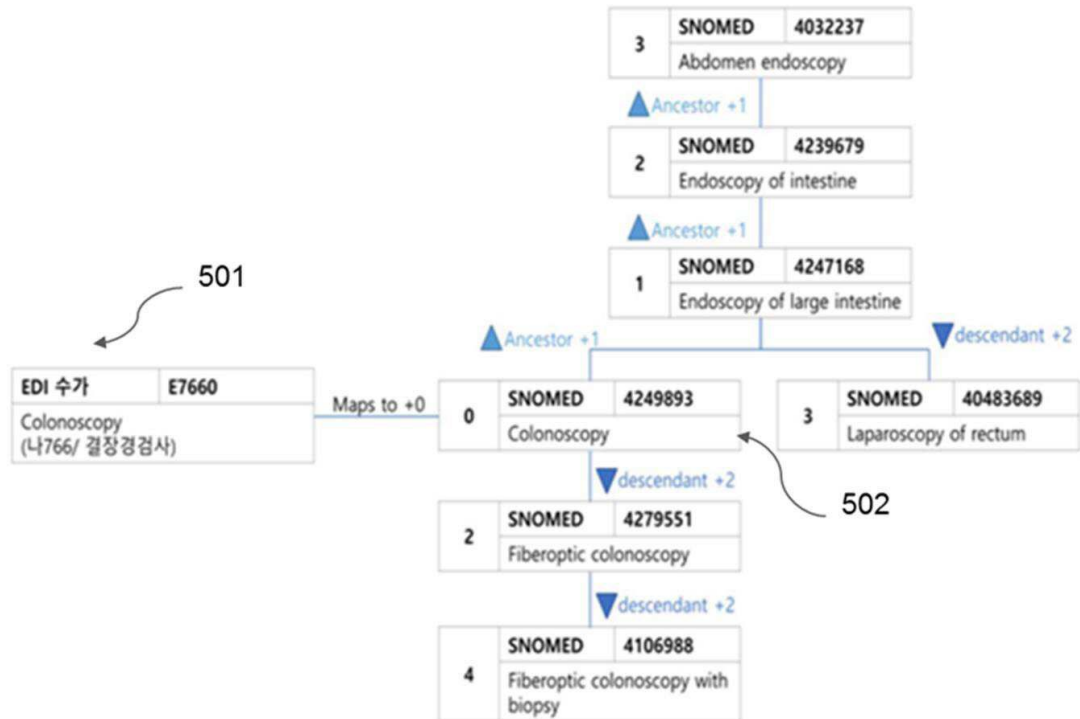
도면3



도면4



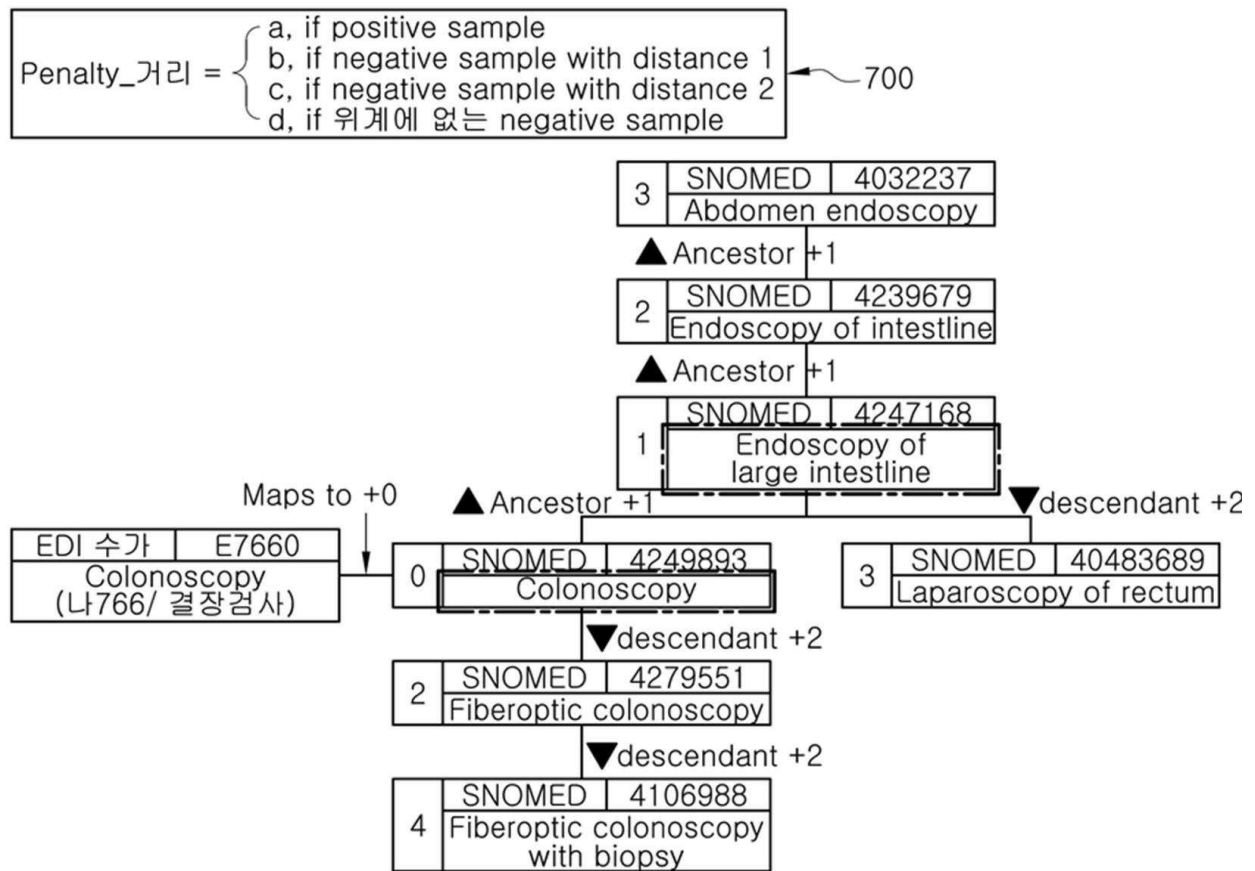
도면5



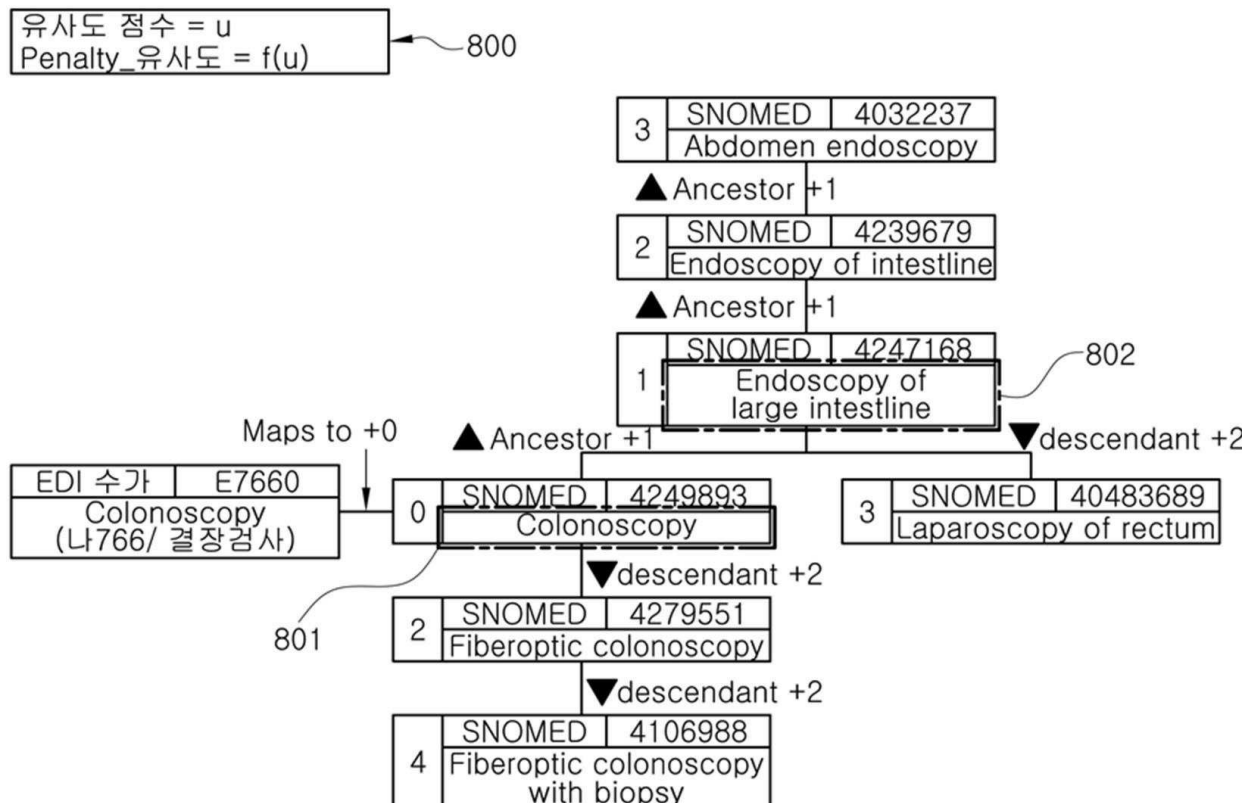
도면6

new_concept_name	rank	relationship	distance
Plasmacytoma of skin of other and unspecified parts of face	1		
Syringoma of skin of other and unspecified parts of face	2	descendant	1.0
Neoplasm of face	3		
Eccrine spiradenoma of skin of other and unspecified parts of face	4	descendant	1.0
Eccrine poroma of skin of other and unspecified parts of face	5	descendant	1.0
Pilomatrixoma of skin of other and unspecified parts of face	6	descendant	1.0
Eccrine poroma malignant of skin of other and unspecified parts of face	7		
Hemangiopericytoma malignant of skin of other and unspecified parts of face	8		
Hemangioendothelioma malignant of skin of other and unspecified parts of face	9		
Neurilemoma malignant of skin of other and unspecified parts of face	10		

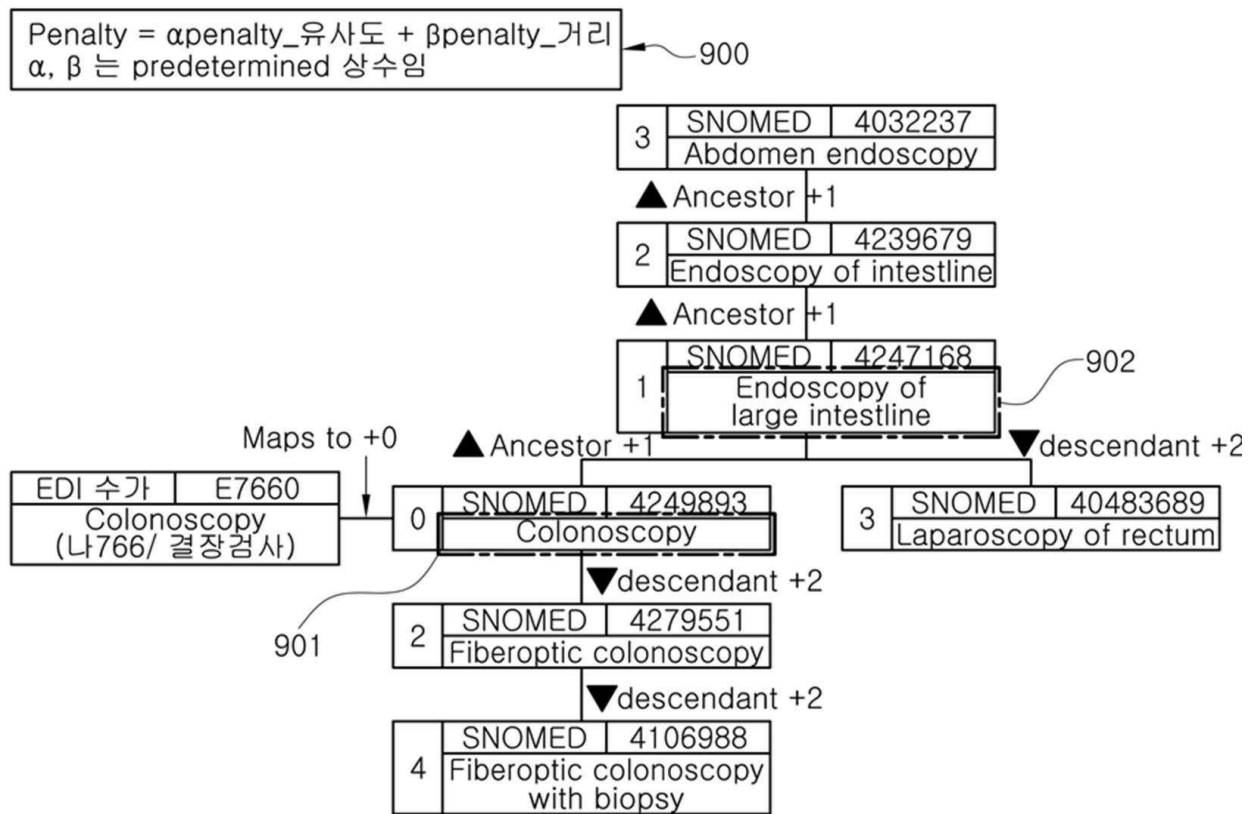
도면7



도면8



도면9



도면10

