(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0056812
(43) 공개일자 2021년05월20일

(51) 국제특허분류(Int. Cl.)
G06F 40/20 (2020.01) *G06F 16/93* (2019.01)
G06Q 50/18 (2012.01)
(52) CPC특허분류
G06F 40/258 (2020.01)
G06F 16/93 (2019.01)
(21) 출원번호 10-2019-0143716
(22) 출원일자 2019년11월11일
심사청구일자 2019년11월11일

(71) 출원인
한림대학교 산학협력단
강원도 춘천시 한림대학길 1, 한림대학교(옥천동)
(72) 발명자
김의직
강원도 춘천시 지석로 29, 516동 1303호 (석사동, 휴먼타운아파트)
이상우
강원도 춘천시 백령로 75길 7 신아아파트 1003호
권정혁
강원도 춘천시 스무숲1길 32-9, 202호
(74) 대리인
김남혁

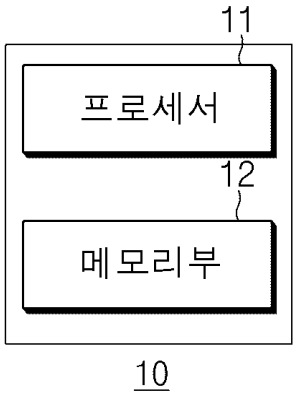
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치, 방법 및 프로그램**

(57) 요 약

본 발명에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치는 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성하고, 상기 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하고, 상기 후보 연구 카테고리 별 상기 카테고리 유사도에 기초하여 상기 연구문헌의 연구 카테고리를 선택하는 프로세서; 및 상기 연구문헌을 저장하는 메모리;를 포함한다.

대 표 도 - 도2



(52) CPC특허분류
G06F 40/216 (2020.01)
G06Q 50/18 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 NRF-2017R1D1A1B03031055
부처명 교육부
과제관리(전문)기관명 한국연구재단
연구사업명 이공학개인기초연구지원사업(기본연구)
연구과제명 산업환경 안전을 위한 사물인터넷 빅데이터 통합 프레임워크 개발
기 여 율 1/2
과제수행기관명 한림대학교 산학협력단
연구기간 2019.03.01 ~ 2020.02.29
이 발명을 지원한 국가연구개발사업
과제고유번호 2019-0-00102
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 정보통신·방송 연구개발사업
연구과제명 복합 전파환경에서의 국민건강 보호 기반 구축
기 여 율 1/2
과제수행기관명 한국전자통신연구원
연구기간 2019.01.01 ~ 2019.12.31

명 세 서

청구범위

청구항 1

후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성하고,

상기 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하고,

상기 후보 연구 카테고리 별 상기 카테고리 유사도에 기초하여 상기 연구문헌의 연구 카테고리를 선택하는 프로세서; 및

상기 연구문헌을 저장하는 메모리;를 포함하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치.

청구항 2

제1항에 있어서,

상기 프로세서는

상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치.

청구항 3

제2항에 있어서,

상기 프로세서는

상기 연구문헌의 단어수와 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어의 상기 연구문헌에서의 제1 카테고리 출현횟수 간의 제1 카테고리 출현비율을 산출하고, 상기 연구문헌의 문장수와 상기 카테고리 특징 단어가 포함된 상기 연구문헌의 문장수 간의 제2 카테고리 출현비율을 산출하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치.

청구항 4

제3항에 있어서,

상기 프로세서는

상기 제1 카테고리 출현비율과 상기 제2 카테고리 출현비율에 기초하여 상기 연구문헌 내 중요도를 산출하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치.

청구항 5

제1항에 있어서,

상기 프로세서는

상기 후보 연구 카테고리 별로 상기 산출된 연구문헌 내 중요도를 합산하여 상기 카테고리 유사도로 산출하고, 상기 카테고리 유사도가 최대인 후보 연구 카테고리를 상기 연구 카테고리로 선택하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치.

청구항 6

프로세서가 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성하는 단계;

상기 프로세서가 상기 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계; 및

상기 프로세서가 상기 후보 연구 카테고리 별 상기 카테고리 유사도에 기초하여 상기 연구문헌의 연구 카테고리를 선택하는 단계;를 포함하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법.

청구항 7

제6항에 있어서,

상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는

상기 프로세서가 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출하는 단계;를 포함하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법.

청구항 8

제7항에 있어서,

상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는

상기 프로세서가 상기 연구문헌의 단어수와 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어의 상기 연구문헌에서의 제1 카테고리 출현횟수 간의 제1 카테고리 출현비율을 산출하고, 상기 연구문헌의 문장수와 상기 카테고리 특징 단어가 포함된 상기 연구문헌의 문장수 간의 제2 카테고리 출현비율을 산출하는 단계;를 더 포함하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법.

청구항 9

제8항에 있어서,

상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는

상기 프로세서가 상기 제1 카테고리 출현비율과 상기 제2 카테고리 출현비율에 기초하여 상기 연구문헌 내 중요도를 산출하는 단계;를 더 포함하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법.

청구항 10

제6항에 있어서,

상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는

상기 프로세서가 상기 후보 연구 카테고리 별로 상기 산출된 연구문헌 내 중요도를 합산하여 상기 카테고리 유

사도로 산출하는 단계;를 더 포함하고,

상기 연구문헌의 연구 카테고리를 선택하는 단계는

상기 프로세서가 상기 카테고리 유사도가 최대인 후보 연구 카테고리를 상기 연구 카테고리로 선택하는 단계;를 포함하는 것을 특징으로 하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법.

청구항 11

하드웨어인 컴퓨터와 결합되어, 제6항의 방법을 수행할 수 있도록 컴퓨터에서 독출 가능한 기록매체에 저장된 컴퓨터 프로그램.

발명의 설명

기 술 분 야

[0001] 본 발명은 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치, 방법 및 프로그램에 관한 것으로, 더욱 상세하게는 연구 카테고리 별 카테고리 특징 단어 목록과 연구문헌에 포함된 문장을 비교하여 연구문헌의 연구 카테고리를 선택하는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치, 방법 및 프로그램에 관한 것이다.

배 경 기 술

[0003] 최근 인터넷을 중심으로 방대한 양의 정보가 유입되고 있고, 스마트 폰 등 휴대용 기기의 소지자가 증가함에 따라 많은 양의 정보 중 필요한 정보만을 획득하는 기술이 주목되고 있다. 이를 위해, 다양한 전자 문서들의 문장 중 핵심 문장만을 추출하는 기술이 연구되고 있다.

[0004] 특히, 전자기장에 의한 인체 위험을 연구한 결과를 포함하는 연구문헌들의 경우, 연구문헌의 저자 이름, 저자 이름, 제목, 출판사 이름, 출판 날짜 등과 같은 간략한 정보들만이 제공되기 때문에 전자기장에 의한 인체 위험과 관련한 연구문헌을 효과적으로 검색하고 인덱싱할 수 없다.

[0005] 특히, 연구문헌의 연구 카테고리의 경우, 연구문헌 내의 자세한 내용을 파악하여야만 확인이 가능하여, 효율적으로 전자기장에 의한 인체 위험과 관련한 연구문헌을 파악하기가 용이하지 않은 문제점이 있다.

[0006] 이에 따라, 전자기장에 의한 인체 위험과 관련한 연구문헌 중에서 특정 연구 카테고리에 해당하는 연구문헌만을 추출하여 효율적으로 연구문헌을 파악할 수 있는 기술이 요구되고 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명은 후보 연구 카테고리 중에서 연구문헌에 해당하는 연구 카테고리를 선택하여 연구 카테고리 정보를 추출할 수 있는 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치, 방법 및 프로그램을 제공할 수 있다.

[0009] 본 발명의 목적들은 이상에서 언급한 목적으로 제한되지 않으며, 언급되지 않은 본 발명의 다른 목적 및 장점들은 하기의 설명에 의해서 이해될 수 있고, 본 발명의 실시예에 의해 보다 분명하게 이해될 것이다. 또한, 본 발명의 목적 및 장점들은 특허 청구 범위에 나타낸 수단 및 그 조합에 의해 실현될 수 있음을 쉽게 알 수 있을 것이다.

과제의 해결 수단

[0011] 본 발명에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치는 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성하고, 상기 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하고, 상기 후보 연구 카테고리 별 상기 카테고리 유사도에 기초하여 상기 연구문헌의 연구 카테고리를 선택하는 프로세서; 및 상기 연구문헌을 저장하는 메모리;를 포함한다.

[0012] 상기 프로세서는 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출할 수 있다.

[0013] 상기 프로세서는 상기 연구문헌의 단어수와 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어의 상기 연구문헌에서의 제1 카테고리 출현횟수 간의 제1 카테고리 출현비율을 산출하고, 상기 연구문헌의 문장수와 상기 카테고리 특징 단어가 포함된 상기 연구문헌의 문장수 간의 제2 카테고리 출현비율을 산출할 수 있다.

[0014] 상기 프로세서는 상기 제1 카테고리 출현비율과 상기 제2 카테고리 출현비율에 기초하여 상기 연구문헌 내 중요도를 산출할 수 있다.

[0015] 상기 프로세서는 상기 후보 연구 카테고리 별로 상기 산출된 연구문헌 내 중요도를 합산하여 상기 카테고리 유사도로 산출하고, 상기 카테고리 유사도가 최대인 후보 연구 카테고리를 상기 연구 카테고리로 선택할 수 있다.

[0016] 본 발명에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법은 프로세서가 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성하는 단계; 상기 프로세서가 상기 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계; 및 상기 프로세서가 상기 후보 연구 카테고리 별 상기 카테고리 유사도에 기초하여 상기 연구문헌의 연구 카테고리를 선택하는 단계;를 포함할 수 있다.

[0017] 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는 상기 프로세서가 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출하는 단계;를 포함할 수 있다.

[0018] 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는 상기 프로세서가 상기 연구문헌의 단어수와 상기 카테고리 특징 단어 중에서 상기 연구문헌 내에 포함된 카테고리 특징 단어의 상기 연구문헌에서의 제1 카테고리 출현횟수 간의 제1 카테고리 출현비율을 산출하고, 상기 연구문헌의 문장수와 상기 카테고리 특징 단어가 포함된 상기 연구문헌의 문장수 간의 제2 카테고리 출현비율을 산출하는 단계;를 더 포함할 수 있다.

[0019] 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는 상기 프로세서가 상기 제1 카테고리 출현비율과 상기 제2 카테고리 출현비율에 기초하여 상기 연구문헌 내 중요도를 산출하는 단계;를 더 포함할 수 있다.

[0020] 상기 후보 연구 카테고리 별로 카테고리 유사도를 산출하는 단계는 상기 프로세서가 상기 후보 연구 카테고리 별로 상기 산출된 연구문헌 내 중요도를 합산하여 상기 카테고리 유사도로 산출하는 단계;를 더 포함할 수 있다.

[0021] 상기 연구문헌의 연구 카테고리를 선택하는 단계는 상기 프로세서가 상기 카테고리 유사도가 최대인 후보 연구 카테고리를 상기 연구 카테고리로 선택하는 단계;를 포함할 수 있다.

[0022] 본 발명에 따른 컴퓨터 프로그램은 하드웨어인 컴퓨터와 결합되어, 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법을 수행할 수 있도록 컴퓨터에서 독출 가능한 기록매체에 저장될 수 있다.

발명의 효과

[0024] 본 발명에 따르면, 후보 연구 카테고리 별로 카테고리 특징 단어 목록과 연구문헌 간의 카테고리 유사도를 산출하여 연구문헌의 연구 카테고리를 선택함으로써, 연구문헌에 의해 수행된 연구의 연구 카테고리 정보를 신속하게 추출할 수 있다.

도면의 간단한 설명

[0026] 도 1은 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치와 전자 장치를 도시한 도면이다.

도 2는 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치의 구성을 도시한 도면이다.

도 3은 본 발명의 일 실시 예에 따른 카테고리 특징 단어 목록의 일 예를 도시한 도면이다.

도 4는 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법의 순서도이다.

도 5는 본 발명의 다른 실시 예에 따른 연구 목적 단어 목록과 비목적 단어 목록의 일 예를 도시한 도면이다.

도 6은 본 발명의 다른 실시 예에 따른 연구 주제 단어 목록의 일 예를 도시한 도면이다.

도 7은 본 발명의 다른 실시 예에 따른 가중치 산출 과정을 설명하기 위한 도면이다.

도 8은 본 발명의 또 다른 실시 예에 따른 연구문헌 내 전자기장 주파수 대역 정보를 추출하는 과정을 설명하기 위한 도면이다.

도 9는 본 발명의 또 다른 실시 예에 따른 연구문헌의 중요도를 평가하는 과정을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0027] 본 발명의 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시예들을 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.

[0028] 본 명세서에서 사용된 용어는 실시예들을 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.

[0029] 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.

[0030] 명세서에서 사용되는 “프로세서”라는 용어는 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소를 의미하며, “프로세서”는 어떤 역할들을 수행한다. 그렇지만 “프로세서”는 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. “프로세서”는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수 있다. 따라서, 일 예로서 “프로세서”는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세서들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 “프로세서”들 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 “프로세서”들로 결합되거나 추가적인 구성요소들과 “프로세서”들로 더 분리될 수 있다.

[0031] 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 위

해 사용될 수 있다. 공간적으로 상대적인 용어는 도면에 도시되어 있는 방향에 더하여 사용시 또는 동작 시 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. 구성요소는 다른 방향으로도 배향될 수 있으며, 이에 따라 공간적으로 상대적인 용어들은 배향에 따라 해석될 수 있다.

[0032] 본 명세서에서, 컴퓨터는 적어도 하나의 프로세서를 포함하는 모든 종류의 하드웨어 장치를 의미하는 것이고, 실시 예에 따라 해당 하드웨어 장치에서 동작하는 소프트웨어적 구성도 포괄하는 의미로서 이해될 수 있다. 예를 들어, 컴퓨터는 스마트폰, 태블릿 PC, 데스크톱, 노트북 및 각 장치에서 구동되는 사용자 클라이언트 및 애플리케이션을 모두 포함하는 의미로서 이해될 수 있으며, 또한 이에 제한되는 것은 아니다.

[0033] 이하, 첨부된 도면을 참조하여 본 발명의 실시예를 상세하게 설명한다.

[0034] 본 명세서에서 설명되는 각 단계들은 프로세서에 의하여 수행되는 것으로 설명되나, 각 단계의 주체는 이에 제한되는 것은 아니며, 실시 예에 따라 각 단계들의 적어도 일부가 서로 다른 장치에서 수행될 수도 있다.

[0035] 도 1은 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치(10, 이하 '추출 장치'라 함)와 전자 장치(20)를 도시한 도면이다.

[0036] 도 1을 참조하면, 추출 장치(10)는 전자 장치(20)로부터 연구문헌을 입력 받고, 연구문헌의 연구 카테고리 정보를 추출할 수 있다.

[0037] 한편, 추출 장치(10)는 연구문헌에 포함된 모든 문장에 대해서 텍스트 마이닝을 수행할 수 있으나, 연산 속도 및 추출 효율을 고려하여 연구문헌의 초록에 대해서만 텍스트 마이닝을 수행할 수 있다.

[0038] 이하에서, 추출 장치(10)가 연구문헌에 대해 각종 수치를 산출하거나 텍스트를 변형하는 것은 연구문헌 전체가 아닌 연구문헌의 초록에 대해서 수행되는 것을 의미할 수 있다.

[0039] 한편, 본 명세서에서의 연구문헌은 연구 카테고리, 저자, 발행처가 한정되지 않으며, 공개된 모든 연구문헌 중 일부일 수 있다.

[0040] 일 실시 예에서, 추출 장치(10)는 전자기장에 따른 인체 건강 위험에 대한 연구 내용이 포함된 연구문헌을 전자 장치(20)로부터 입력 받을 수 있다.

[0041] 전자 장치(20)는 추출 장치(10)로 연구문헌을 제공하기 위한 구성이다. 본 발명에 따른 전자 장치(200)는 서버 또는 스마트폰으로 구현될 수 있으나, 이는 일 실시예에 불과할 뿐, 서버 또는 스마트폰 외에 태블릿 PC(tablet personal computer), 이동 전화기(mobile phone), 영상 전화기, 전자책 리더기(e-book reader), 데스크탑 PC(desktop PC), 랩탑 PC(laptop PC), 넷북 컴퓨터(netbook computer), 워크스테이션(workstation), 서버, PDA(personal digital assistant), PMP(portable multimedia player) 또는 웨어러블 장치(wearable device) 중 어느 하나일 수 있다.

[0042] 도 2는 본 발명의 일 실시 예에 따른 추출 장치(10)의 구성을 도시한 도면이고, 도 3은 본 발명의 일 실시 예에 따른 카테고리 특징 단어 목록의 일 예를 도시한 도면이다.

[0043] 본 발명의 일 실시 예에 따른 추출 장치(10)는 연구 카테고리 별 카테고리 특징 단어 목록을 이용하여 연구문헌의 연구 카테고리 정보를 추출할 수 있다.

[0044] 이를 위한, 본 발명의 일 실시 예에 따른 추출 장치(10)는 프로세서(11) 및 메모리부(12)를 포함할 수 있다.

[0045] 프로세서(11)는 연구문헌에 대해 텍스트 마이닝을 수행하기 위해 연구문헌을 전처리할 수 있다.

[0046] 구체적으로, 프로세서(11)는 전처리 과정으로써, 연구문헌의 텍스트 중 대문자를 소문자 변경하는 제1 전처리 과정, 연구문헌의 문장 부호, 공백 및 be 동사와 같은 중지 단어를 제거하는 제2 전처리 과정, 연구문헌의 단어가 변형 단어인 경우 원형 단어를 추출하는 제3 전처리 과정 및 연구문헌의 단어 또는 문장을 토큰화하는 제4 전처리 과정을 수행할 수 있다.

[0047] 일 실시 예에서, 프로세서(11)는 Python 라이브러리인 NLTK(Natural Language Toolkit)를 이용하여 상술된 전처리 과정을 구현할 수 있다.

[0048] 이후, 프로세서(11)는 도 3에 도시된 바와 같이, 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성할

수 있다.

[0049] 구체적으로, 프로세서(11)는 후보 연구 카테고리인 역학적 연구 카테고리, 동물 실험 연구 카테고리 및 세포 실험 연구 카테고리 각각의 카테고리 특징 단어 목록을 생성할 수 있다.

[0050] 프로세서(11)는 전자 장치(20)로부터 입력되는 카테고리 특징 단어를 분류하여 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성할 수 있다.

[0051] 여기서, 카테고리 특징 단어 목록은 카테고리의 특징을 나타내는 카테고리 특징 단어가 포함된 단어 목록일 수 있다.

[0052] 이후, 프로세서(11)는 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 후보 연구 카테고리 별로 카테고리 유사도를 산출할 수 있다.

[0053] 구체적으로, 프로세서(11)는 카테고리 특징 단어 중에서 연구문헌 내에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출할 수 있다.

[0054] 이때, 프로세서(11)는 연구문헌의 단어수와 카테고리 특징 단어 중에서 연구문헌 내에 포함된 카테고리 특징 단어의 연구문헌에서의 제1 카테고리 출현횟수 간의 제1 카테고리 출현비율을 산출하고, 연구문헌의 문장수와 카테고리 특징 단어가 포함된 연구문헌의 문장수 간의 제2 카테고리 출현비율을 산출할 수 있다.

[0055] 이어서, 프로세서(11)는 제1 카테고리 출현비율과 제2 카테고리 출현비율에 기초하여 연구문헌 내 중요도를 산출할 수 있다.

[0056] 여기서, 프로세서(11)는 하기의 수학식 1을 이용하여 카테고리 특징 단어 중 어느 하나인 제x 단어의 연구문헌 내 중요도를 산출할 수 있다.

[0057]

[0058] <수학식 1>

$$W_{x,x} = \frac{W_x PL}{WPL} \times \log_b \frac{SPL}{S_{Wx} PL}$$

[0059]

[0060] 여기서, $W_{x,x}$은 카테고리 특징 단어 중 어느 하나인 제x 단어의 연구문헌 내 중요도이고, W_xPL은 제x 단어의 연구문헌에서의 제1 카테고리 출현횟수 이고, WPL은 연구문헌의 단어수이고, SPL은 연구문헌의 문장수이고, S_{Wx}PL은 제x 단어가 포함된 연구문헌의 문장수이고, b는 조절 상수이다.

[0062] 한편, 프로세서(11)는 출현 빈도에 기초하여 중요도를 산출하는 TF-IDF(Term Frequency-Inverse Document Frequency) 분석법을 이용하여 카테고리 특징 단어 중 어느 하나인 제x 단어의 연구문헌 내 중요도를 산출할 수 있다.

[0063] 이를 통해, 프로세서(11)는 후보 연구 카테고리 별 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출할 수 있다.

[0064] 이후, 프로세서(11)는 연구문헌 내 중요도를 후보 연구 카테고리 별로 합산하여 후보 연구 카테고리 별로 카테고리 유사도를 산출하고 카테고리 유사도가 최대인 후보 연구 카테고리를 연구 카테고리로 선택할 수 있다.

[0065] 이에 따라, 프로세서(11)는 후보 연구 카테고리 중에서 연구 문헌의 연구 카테고리와 가장 근접한 후보 연구 카테고리를 신속하고 정확하게 선택할 수 있다.

[0066] 메모리부(12)는 전자 장치(20)로부터 입력되는 연구문헌을 저장할 수 있다. 또한, 메모리부(12)는 프로세서(11)의 상술된 연산 과정에 필요한 프로그램을 저장할 수 있다.

[0067] 이러한 메모리부(12)는 비휘발성 메모리(예: 플래시 메모리, 하드 디스크 등) 및 휘발성 메모리(예: RAM(random access memory)를 포함할 수 있고, 프로그램은 비휘발성 메모리에 저장되고 휘발성 메모리로 로드되어 동작할

수 있다.

[0068] 도 4는 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법의 순서도이다.

[0069] 도 4를 참조하면, 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법은 S1 단계에서, 프로세서가 프로세서가 후보 연구 카테고리 별로 카테고리 특징 단어 목록을 생성하게 된다.

[0070] 이후, S2 단계에서, 프로세서가 프로세서가 카테고리 특징 단어 목록에 포함된 카테고리 특징 단어가 연구문헌에 포함되어 있는지 여부에 기초하여 후보 연구 카테고리 별로 카테고리 유사도를 산출하게 된다.

[0071] 이때, S2 단계에서, 프로세서가 카테고리 특징 단어 중에서 연구문헌 내에 포함된 카테고리 특징 단어 각각의 연구문헌 내 중요도를 산출하는 단계를 수행하게 된다.

[0072] 또한, S2 단계에서, 프로세서가 연구문헌의 단어수와 카테고리 특징 단어 중에서 연구문헌 내에 포함된 카테고리 특징 단어의 연구문헌에서의 제1 카테고리 출현횟수 간의 제1 카테고리 출현비율을 산출하고, 연구문헌의 문장수와 카테고리 특징 단어가 포함된 연구문헌의 문장수 간의 제2 카테고리 출현비율을 산출하는 단계를 더 수행하게 된다.

[0073] 또한, S2 단계에서, 프로세서가 제1 카테고리 출현비율과 제2 카테고리 출현비율에 기초하여 연구문헌 내 중요도를 산출하는 단계를 더 수행하게 된다.

[0074] 또한, S2 단계에서, 프로세서가 후보 연구 카테고리 별로 산출된 연구문헌 내 중요도를 합산하여 카테고리 유사도로 산출하는 단계를 더 수행하게 된다.

[0075] 이후, S3 단계에서, 프로세서가 후보 연구 카테고리 별 카테고리 유사도에 기초하여 연구문헌의 연구 카테고리를 선택하게 된다.

[0076] 이때, S3 단계에서, 프로세서가 프로세서가 카테고리 유사도가 최대인 후보 연구 카테고리를 연구 카테고리로 선택하는 단계를 더 수행하게 된다.

[0077] 한편, 본 발명의 일 실시 예에 따른 컴퓨터 프로그램은 하드웨어인 컴퓨터와 결합되어, 본 발명의 일 실시 예에 따른 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 방법을 수행할 수 있도록 컴퓨터에서 독출 가능한 기록매체에 저장될 수 있다.

[0078] 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, 탈착형 디스크, CD-ROM, 또는 본 발명이 속하는 기술 분야 에서 잘 알려진 임의의 형태의 컴퓨터 판독가능 기록매체에 상주할 수도 있다.

[0079] 도 5는 본 발명의 다른 실시 예에 따른 연구 목적 단어 목록과 비목적 단어 목록의 일 예를 도시한 도면이고, 도 6은 본 발명의 다른 실시 예에 따른 연구 주제 단어 목록의 일 예를 도시한 도면이고, 도 7은 본 발명의 다른 실시 예에 따른 가중치 산출 과정을 설명하기 위한 도면이다.

[0080] 도 5 내지 도 7을 참조하면, 추출 장치(10)는 프로세서(11)와 메모리부(12)를 포함할 수 있다.

[0081] 프로세서(11)는 연구문헌에 대해 텍스트 마이닝을 수행하기 위해 연구문헌을 전처리할 수 있다.

[0082] 구체적으로, 프로세서(11)는 전처리 과정으로써, 연구문헌의 텍스트 중 대문자를 소문자 변경하는 제1 전처리 과정, 연구문헌의 문장 부호, 공백, be 동사 및 1회 출현한 단어와 같은 중지 단어를 제거하는 제2 전처리 과정, 연구문헌의 단어가 변형 단어인 경우 원형 단어를 추출하는 제3 전처리 과정 및 연구문헌의 단어 또는 문장을 토큰화하는 제4 전치리 과정을 수행할 수 있다.

[0083] 일 실시 예에서, 프로세서(11)는 Python 라이브러리인 NLTK(Natural Language Toolkit)를 이용하여 상술된 전처리 과정을 구현할 수 있다.

[0084] 이후, 프로세서(11)는 연구문헌에 포함된 문장 각각에 연구 목적 정보가 포함되었는지 여부를 판단하여 문장 중

에서 후보 목적 전술 문장을 선택할 수 있다.

[0085] 구체적으로, 프로세서(11)는 도 5에 도시된 바와 같이, 연구 목적의 설명 시 이용되는 목적 관용 단어와 연구문헌의 제목에 포함된 제목 단어를 포함하는 연구 목적 단어 목록을 생성하고, 연구의 실험 결과의 설명 시 이용되는 실험 관용 단어를 포함하는 비목적 단어 목록을 생성할 수 있다.

[0086] 즉, 프로세서(11)는 연구 목적을 전술하는 문장에서 빈번하게 쓰이는 목적 관용 단어와 추출 대상이 되는 연구문헌의 연구 목적이 함축되어 제목 단어를 합쳐 연구 목적 단어 목록으로 생성할 수 있다.

[0087] 또한, 프로세서(11)는 연구문헌의 초록에 빈번하게 쓰이지만 연구 목적과 연계성이 없는 실험 결과를 설명하는 경우 빈번하게 쓰이는 실험 관용 단어가 포함되도록 비목적 단어 목록을 생성할 수 있다.

[0088] 이후, 프로세서(11)는 연구문헌에 포함된 문장 각각에 포함된 단어가 연구 목적 단어 목록과 비목적 단어 목록에 포함되는지 여부에 기초하여 연구문헌에 포함된 문장 중에서 후보 목적 전술 문장을 선택할 수 있다.

[0089] 구체적으로, 프로세서(11)는 연구문헌에 포함된 문장의 단어 중 하나 이상이 연구 목적 단어 목록에 포함되고, 비목적 단어 목록에 포함되지 않으면, 해당 문장을 후보 목적 전술 문장으로 선택할 수 있다.

[0090] 이에 따라, 프로세서(11)는 연구문헌에 포함된 문장 중에서 연구 목적과 연관성이 높은 복수의 문장만을 후보 목적 전술 문장으로 선택할 수 있다.

[0091] 한편, 프로세서(11)는 도 6에 도시된 바와 같이, 연구문헌에 포함된 단어 각각의 연구문헌 내 중요도를 산출하고, 연구문헌 내 중요도에 기초하여 연구 주제 단어 목록을 생성할 수 있다.

[0092] 구체적으로, 프로세서(11)는 연구문헌에 포함된 단어 중 어느 하나인 제1 단어와 연구문헌에 포함된 문장 중 어느 하나인 제1 문장 간의 연구문헌 내 중요도를 산출할 수 있다.

[0093] 프로세서(11)는 제1 문장에 포함된 단어의 단어수와 제1 단어의 제1 문장에서의 제1 주제 출현횟수 간의 제1 주제 출현비율을 산출하고, 연구문헌의 문장수와 제1 단어가 포함된 연구문헌의 문장수 간의 제2 주제 출현비율을 산출할 수 있다.

[0094] 이후, 프로세서(11)는 제1 주제 출현비율과 제2 주제 출현비율에 기초하여 제1 단어와 제1 문장 간의 연구문헌 내 중요도를 산출할 수 있다.

[0095] 이때, 프로세서(11)는 하기의 수학식 2를 이용하여 연구문헌에 포함된 단어 중 어느 하나인 제n 단어와 연구문헌에 포함된 문장 중 어느 하나인 제m 문장 간의 연구문헌 내 중요도를 산출할 수 있다.

[0096]

[0097] <수학식 2>

$$W_{n,m} = \frac{W_n PS_m}{WPS_m} \times \log_a \frac{SPL}{S_{Wn} PL}$$

[0098]

[0099] 여기서, $W_{n,m}$은 연구문헌에 포함된 단어 중 어느 하나인 제n 단어와 연구문헌에 포함된 문장 중 어느 하나인 제m 문장 간의 연구문헌 내 중요도이고, W_nPS_m은 제n 단어의 제m 문장에서의 제1 주제 출현횟수이고, WPS_m은 제m 문장에 포함된 단어의 단어수이고, SPL은 연구문헌의 문장수이고, $S_{Wn}PL$은 제n 단어가 포함된 연구문헌의 문장수이고, a는 조절 상수이다.

[0101] 프로세서(11)는 상술된 방식으로 연구문헌에 포함된 모든 제n 단어와 제m 문장 각각에 간의 연구문헌 내 중요도를 산출하여, 매트릭스 형태인 TF-IDF 매트릭스 데이터를 생성할 수 있다. 즉, 프로세서(11)는 제n 단어와 제m 문장 각각에 간의 연구문헌 내 중요도 이를 위해, 프로세서(11)는 출현 빈도에 기초하여 중요도를 산출하는 TF-IDF(Term Frequency-Inverse Document Frequency) 분석법을 이용하여 제n 단어와 제m 문장 각각에 간의 연구문헌 내 중요도를 산출할 수 있다.

[0102] 이후, 프로세서(11)는 잠재 의미 분석(Latent Semantic Analysis; LSA) 툴의 특이값 분해(Singular Value

Decomposition; SVD) 기능을 이용하여 제n 단어와 제m 문장 각각에 간의 연구문헌 내 중요도 중에서 유의미한 중요도만을 추출하여 후보 연구 주제 단어를 10개씩 포함하는 후보 연구 주제 단어 목록을 6개 생성할 수 있다.

[0103] 프로세서(11)는 후보 연구 주제 단어 목록 각각과 연구문헌의 제목에 포함된 단어와의 중복 단어수를 확인하고, 중복 단어수가 최다인 후보 연구 주제 단어 목록을 연구 문헌의 연구 주제 단어 목록으로 선택할 수 있다.

[0104] 이를 통해, 프로세서(11)는 연구 문헌의 연구 주제와 관련이 높은 연구 주제 단어로 구성된 연구 주제 단어 목록을 생성할 수 있다.

[0105] 이러한, 연구 주제 단어 목록에는 도 6에 도시된 바와 같이, 연구 주제 단어 10개와 연구 주제 단어 각각에 매칭된 연구문헌 내 중요도가 포함될 수 있다.

[0106] 이후, 프로세서(11)는 도 7에 도시된 바와 같이, 후보 목적 전술 문장 각각의 임시 가중치 및 연구문헌의 제목과 후보 목적 전술 문장 각각 간의 제목 유사도를 산출하여 후보 목적 전술 문장 각각의 가중치를 산출할 수 있다.

[0107] 구체적으로, 프로세서(11)는 연구 주제 단어 중에서 후보 목적 전술 문장에 포함된 연구 주제 단어의 연구문헌 내 중요도를 합산하여 임시 가중치를 산출할 수 있다.

[0108] 이후, 프로세서(11)는 후보 목적 전술 문장에 포함된 단어 중에서 연구문헌의 제목에 포함된 단어의 단어수와 연구문헌의 제목의 단어수 간의 단어수 비율을 제목 유사도로 산출할 수 있다.

[0109] 최종적으로, 프로세서(11)는 임시 가중치와 제목 유사도를 곱하여 후보 목적 전술 문장 각각의 가중치 산출할 수 있다.

[0110] 이러한, 프로세서(11)는 가중치가 최대인 후보 목적 전술 문장을 목적 전술 문장으로 선택할 수 있다.

[0111] 즉, 프로세서(11)는 연구문헌의 문장 중에서 연구 목적을 나타내는 후보 목적 전술 문장을 복수로 추출하고, 후보 목적 전술 문장 각각의 임시 가중치와 제목 유사도를 각각 산출한 후 임시 가중치와 제목 유사도를 곱하여 후보 목적 전술 문장 각각의 가중치를 산출하여 목적 전술 문장을 선택함으로써, 연구문헌의 연구 목적과 최근접한 내용을 포함하는 문장을 목적 전술 문장으로 선택할 수 있다.

[0112] 도 8은 본 발명의 또 다른 실시 예에 따른 연구문헌 내 전자기장 주파수 대역 정보를 추출하는 과정을 설명하기 위한 도면이다.

[0113] 본 발명의 또 다른 실시 예에 따른 추출 장치(10)는 연구문헌 내 전자기장 방출원의 전자기장 주파수 대역 정보를 추출할 수 있다.

[0114] 이를 위한, 본 발명의 또 다른 실시 예에 따른 추출 장치(10)는 프로세서(11) 및 연구문헌을 저장하는 메모리부(12)를 포함할 수 있다.

[0115] 프로세서(11)는 연구문헌에 대해 텍스트 마이닝을 수행하기 위해 연구문헌을 전처리할 수 있다.

[0116] 구체적으로, 프로세서(11)는 전처리 과정으로써, 연구문헌의 단어 또는 문장을 토큰화하는 제4 전치리 과정을 수행할 수 있다.

[0117] 일 실시 예에서, 프로세서(11)는 Python 라이브러리인 NLTK(Natural Language Toolkit)를 이용하여 상술된 전처리 과정을 구현할 수 있다.

[0118] 이후, 프로세서(11)는 전자기장의 주파수 대역 설명 시 이용되는 주파수 관용 단어를 포함하는 주파수 대역 단어 목록을 생성할 수 있다.

[0119] 여기서, 주파수 관용 단어는 주파수를 대역 설명 시 빈번하게 쓰이는 단어일 수 있으며, 예를 들어, 주파수 관용 단어는 "Hz", "MHz" 및 "GHz" 중 어느 하나일 수 있다.

[0120] 프로세서(11)는 주파수 관용 단어를 이용하여 연구문헌 내에서 후보 전자기장 주파수 대역 정보를 추출할 수 있다.

[0121] 구체적으로, 프로세서(11)는 도 8에 도시된 바와 같이, 연구문헌 내에서 주파수 관용 단어를 검색하고, 검색된 주파수 관용 단어와 인접한 숫자 텍스트를 검색하고, 검색된 주파수 관용 단어와 검색된 숫자 텍스트를 후보 전자기장 주파수 대역 정보로 추출할 수 있다.

[0122] 프로세서(11)는 후보 전자기장 주파수 대역 정보 중에서 전자기장 방출원 조건을 만족하는 후보 전자기장 주파수 대역 정보를 전자기장 방출원의 전자기장 주파수 대역 정보로 선택할 수 있다.

[0123] 이를 위해, 프로세서(11)는 후보 전자기장 주파수 대역 정보의 주파수 관용 단어에 대응하여 주파수 단위를 확인하고, 확인된 주파수 단위에 기초하여 후보 전자기장 주파수 대역 정보의 전자기장 주파수 대역을 확인할 수 있다.

[0124] 이후, 프로세서(11)는 후보 전자기장 주파수 대역 정보가 복수이면, 전자기장 주파수 대역이 높은 순으로 미리 설정된 개수 내에 포함되는 후보 전자기장 주파수 대역 정보가 전자기장 방출원 조건을 만족하는 것으로 판단할 수 있다.

[0125] 예를 들어, 미리 설정된 개수가 1개이고, 도 8에 도시된 바와 같이, 후보 전자기장 주파수 대역 정보 “10GHz 내지 11GHz”와 “150Mhz 내지 160Mhz” 2개인 경우, 프로세서(11)는 전자기장 주파수 대역이 가장 높은 “10GHz 내지 11GHz”가 전자기장 방출원 조건을 만족하는 것으로 판단하여 “10GHz 내지 11GHz”를 전자기장 주파수 대역 정보로 결정할 수 있다.

[0126] 다른 실시 예에 따른 프로세서(11)는 후보 전자기장 주파수 대역 정보의 전자기장 주파수 대역이 미리 설정된 주파수 대역에 포함되면 해당 후보 전자기장 주파수 대역 정보가 전자기장 방출원 조건을 만족하는 것으로 판단할 수 있다.

[0127] 예를 들어, 미리 설정된 주파수 대역이 1GHz 내지 100Ghz인 경우, 프로세서(11)는 미리 설정된 주파수 대역 1GHz 내지 100Ghz에 포함되는 전자기장 주파수 대역인 “10GHz 내지 11GHz”가 전자기장 방출원 조건을 만족하는 것으로 판단하여 “10GHz 내지 11GHz”를 전자기장 주파수 대역 정보로 결정할 수 있다.

[0128] 이를 통해, 프로세서(11)는 연구문헌에서 인체 건강에 위험한 영향일 미치는 전자기장 방출원이 방출하는 전자기장 주파수 대역 정보를 신속하고 정확하게 추출할 수 있다.

[0129] 도 9는 본 발명의 또 다른 실시 예에 따른 연구문헌의 중요도를 평가하는 과정을 설명하기 위한 도면이다.

[0130] 본 발명의 또 다른 실시 예에 따른 추출 장치(10)는 연구문헌의 연구 목적과 전자기장 주파수 대역 정보를 이용하여 연구문헌의 중요도를 평가할 수 있다.

[0131] 이를 위한, 본 발명의 또 다른 실시 예에 따른 추출 장치(10)는 프로세서(11) 및 연구문헌을 저장하는 메모리부(12)를 포함할 수 있다.

[0132] 프로세서(11)는 연구문헌의 연구 목적을 나타내는 목적 전술 문장과 연구문헌의 연구 카테고리에 대응되는 카테고리 중요 단어 목록 간의 목적 유사도를 산출할 수 있다.

[0133] 프로세서(11)는 상기 카테고리 중요 단어 목록 중에서 상기 목적 전술 문장에 포함된 단어의 단어수와 상기 목적 전술 문장에 포함된 단어의 단어수 간의 단어수 비율을 상기 목적 유사도로 산출할 수 있다.

[0134] 여기서, 카테고리 중요 단어 목록은 카테고리 특징 단어 목록의 단어 중에서 연구문헌에서의 출현빈도가 상위인 일부 단어만을 포함하는 단어 목록일 수 있다.

[0135] 한편, 프로세서(11)는 상기 연구문헌 내 전자기장 방출원의 전자기장 주파수 대역 정보와 상기 연구 카테고리에 대응되는 카테고리 중요 주파수 대역 정보 간의 주파수 유사도를 산출할 수 있다.

[0136] 여기서, 연구 카테고리 주요 주파수 대역은 연구 카테고리 별로 주요하게 연구되는 주파수 대역일 수 있다.

[0137] 구체적으로, 프로세서(11)는 상기 카테고리 중요 주파수 대역 정보의 주파수 대역과 상기 전자기장 방출원의 상기 전자기장 주파수 대역 정보의 주파수 대역의 주파수 대역 비율을 상기 주파수 유사도로 산출할 수 있다.

[0138] 예를 들어, 프로세서(11)는 카테고리 중요 주파수 대역 정보의 주파수 대역이 “1Hz 내지 100Hz”이고, 상기 전자기장 방출원의 상기 전자기장 주파수 대역 정보의 주파수 대역이 “51Hz 내지 120Hz”인 경우, 카테고리 중요 주파수 대역 정보의 주파수 대역 중에서 상기 전자기장 방출원의 상기 전자기장 주파수 대역 정보의 주파수 대역과 중첩되는 주파수 대역의 비율을 주파수 대역 비율로 산출하여 “50%”로 산출할 수 있다.

[0139] 최종적으로, 프로세서(11)는 상기 목적 유사도와 상기 주파수 유사도에 기초하여 상기 연구문헌의 중요도를 평가할 수 있다.

[0140] 구체적으로, 프로세서(11)는 상기 목적 유사도와 상기 주파수 유사도를 합산하여 상기 연구문헌의 중요도로 산

출하고, 상기 중요도가 미리 설정 기준 중요도를 초과하는 경우, 상기 연구문헌의 중요도 등급을 중요 등급으로 평가할 수 있다.

[0141] 한편, 다른 실시 예에 따른 프로세서(11)는 상기 목적 전술 문장과 상기 연구문헌의 연구 카테고리와 상이한 연구 카테고리에 대응되는 카테고리 중요 단어 목록 간의 목적 비유사도를 산출할 수 있다.

[0142] 프로세서(11)는 연구문헌의 연구 카테고리와 상이한 연구 카테고리에 대응되는 카테고리 중요 단어 목록 중에서 상기 목적 전술 문장에 포함된 단어의 단어수와 상기 목적 전술 문장에 포함된 단어의 단어수 간의 단어수 비율을 상기 목적 비유사도로 산출할 수 있다.

[0143] 즉, 프로세서(11)는 목적 전술 문장에 포함된 단어와 연구문헌의 연구 카테고리와 상이한 연구 카테고리에 대응되는 카테고리 중요 단어 목록에 포함된 단어가 중복될수록 목적 비유사도가 높도록 산출할 수 있다.

[0144] 이후, 프로세서(11)는 상기 목적 유사도와 상기 목적 비유사도 간의 차이값이 미리 설정된 차이값 범위에 포함되면, 상기 연구문헌의 중요도 등급을 비중요 등급으로 평가할 수 있다.

[0145] 이제까지 본 발명에 대하여 바람직한 실시 예를 중심으로 살펴보았다. 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 본 발명을 구현할 수 있음을 이해할 것이다. 그러므로 상기 개시된 실시 예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 한다.

[0146] 이상과 같이, 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 이것에 의해 한정되지 않으며 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에 의해 본 발명의 기술사상과 아래에 기재될 특허청구범위의 균등범위 내에서 다양한 수정 및 변형이 가능함은 물론이다.

부호의 설명

[0148] 10: 연구 카테고리 별 카테고리 특징 단어 목록을 이용한 연구문헌의 연구 카테고리 정보 추출 장치

11: 프로세서

12: 메모리부

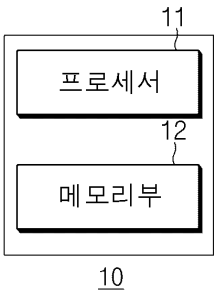
20: 전자 장치

도면

도면1



도면2



도면3

역학적 연구 카테고리 특징 단어 목록

personnel	occupation	cohort
transmitter	broadcast	industry
⋮	⋮	⋮

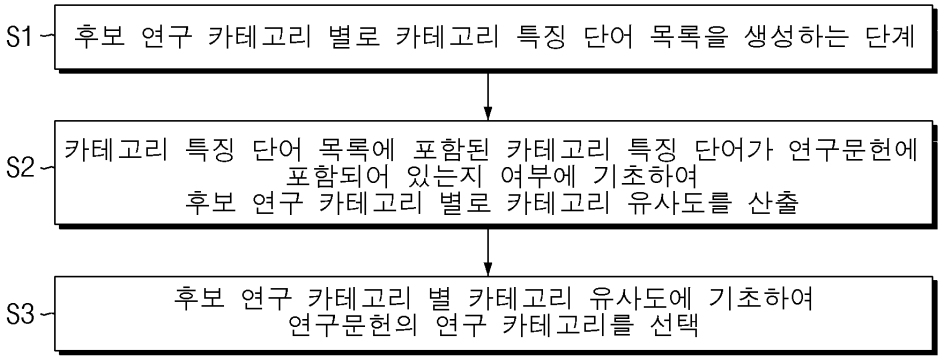
동물실험 연구 카테고리 특징 단어 목록

rat	mice	mouse
animals		
⋮	⋮	⋮

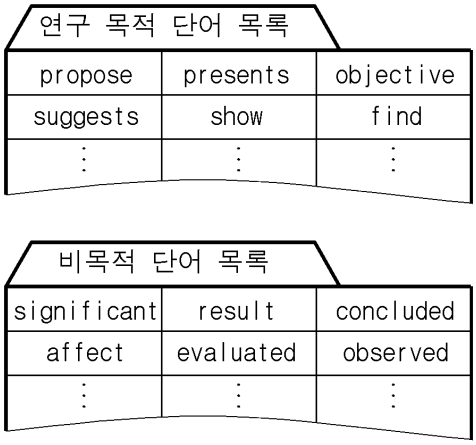
세포실험 연구 카테고리 특징 단어 목록

leukocyte	cell	DNA
RNA	protein	interferon
⋮	⋮	⋮

도면4



도면5



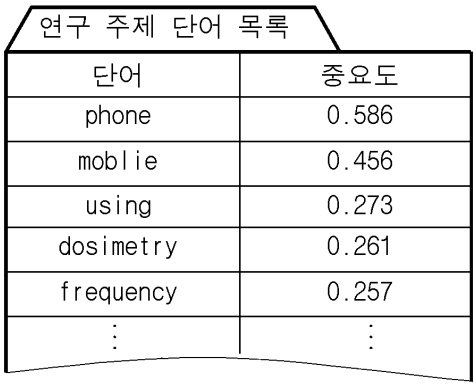
연구 목적 단어 목록

propose	presents	objective
suggests	show	find
⋮	⋮	⋮

비목적 단어 목록

significant	result	concluded
affect	evaluated	observed
⋮	⋮	⋮

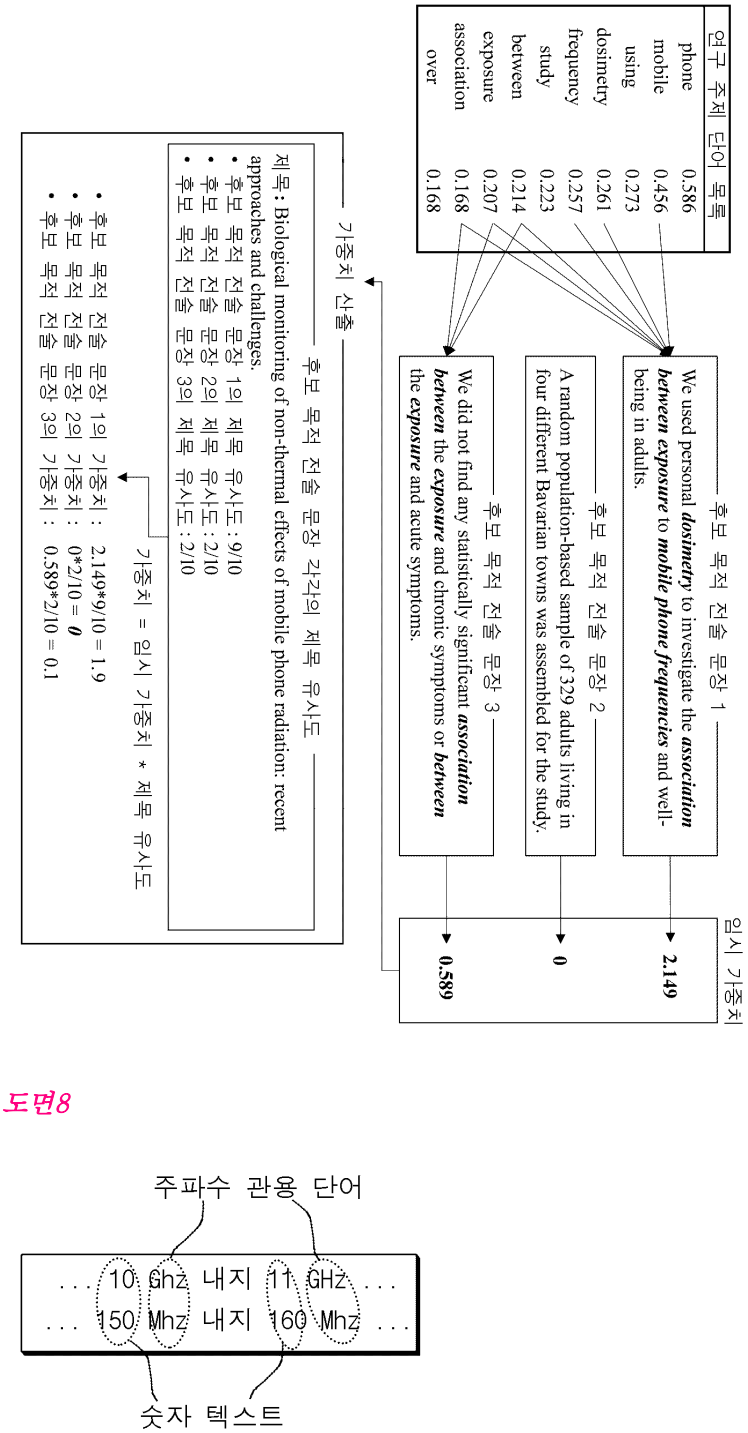
도면6



연구 주제 단어 목록

단어	중요도
phone	0.586
moblie	0.456
using	0.273
dosimetry	0.261
frequency	0.257
⋮	⋮

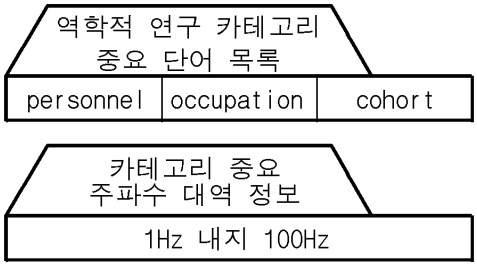
도면7


연구 주제 단어 목록
phone 0.586
mobile 0.456
using 0.273
dosimetry 0.261
frequency 0.257
study 0.223
between 0.214
exposure 0.207
association 0.168
over 0.168
후보 목적 전술 문장 1
We used personal dosimetry to investigate the association between exposure to mobile phone frequencies and well-being in adults.
후보 목적 전술 문장 2
A random population-based sample of 329 adults living in four different Bavarian towns was assembled for the study.
후보 목적 전술 문장 3
We did not find any statistically significant association between the exposure and chronic symptoms or between the exposure and acute symptoms.
임시 가중치
2.149
0
0.589
가중치 산출
후보 목적 전술 문장 각각의 제목 유사도
제목 : Biological monitoring of non-thermal effects of mobile phone radiation: recent approaches and challenges.
• 후보 목적 전술 문장 1의 제목 유사도 : 9/10
• 후보 목적 전술 문장 2의 제목 유사도 : 2/10
• 후보 목적 전술 문장 3의 제목 유사도 : 2/10
가중치 = 임시 가중치 * 제목 유사도
• 후보 목적 전술 문장 1의 가중치 : 2.149*9/10 = 1.9
• 후보 목적 전술 문장 2의 가중치 : 0*2/10 = 0
• 후보 목적 전술 문장 3의 가중치 : 0.589*2/10 = 0.1

도면8

주파수 관용 단어
... 10 Ghz 내지 11 GHz ...
... 150 Mhz 내지 160 Mhz ...
숫자 텍스트

도면9


역학적 연구 카테고리
중요 단어 목록
personnel
occupation
cohort
카테고리 중요
주파수 대역 정보
1Hz 내지 100Hz