



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2008년04월02일
(11) 등록번호 10-0818628
(24) 등록일자 2008년03월26일

(51) Int. Cl.

G06F 19/00 (2006.01)

(21) 출원번호 10-2006-0089003

(22) 출원일자 2006년09월14일

심사청구일자 2006년09월14일

(65) 공개번호 10-2008-0024635

(43) 공개일자 2008년03월19일

(56) 선행기술조사문헌

JP2003296327 A

KR1020000039749 A

KR1020060067071 A

KR1020070059869 A

전체 청구항 수 : 총 10 항

(73) 특허권자

한국전자통신연구원

대전 유성구 가정동 161번지

(72) 발명자

최승권

대전 유성구 반석동 양지마을 503-1905

권오욱

대전 서구 내동 맑은아침아파트 113-1305

(뒷면에 계속)

(74) 대리인

신영무

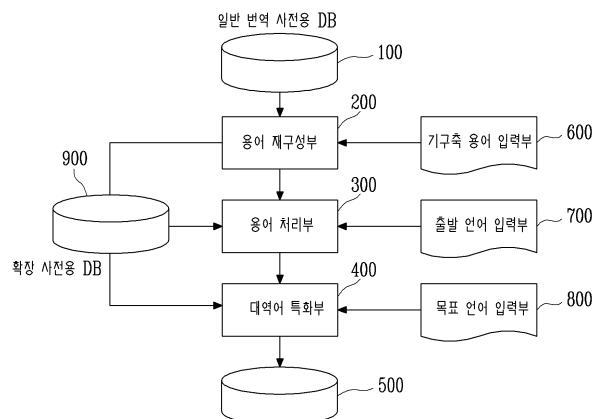
심사관 : 양희용

(54) 특허 번역 사전 구축 장치 및 방법

(57) 요약

대용량의 특허 번역 사전을 효율적이며 경제적으로 구축하는 장치 및 방법을 개시한다. 이를 위하여, 상기 장치는 기구축된 전문 용어를 입력받아 특허 번역 사전용 DB에 기록 가능한 형태로 상기 입력된 전문 용어의 포맷을 변경하여 확장 사전용 DB에 기록하는 용어 재구성부와, 출발 언어 특허 문서를 입력받아 상기 확장 사전용 DB에 기록되지 않은 미등록 전문 용어를 상기 특허 문서로부터 추출하고, 상기 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하며, 대역어가 부착되지 않은 상기 단일어 엔트리에 대해 상기 대역어를 부착한 후, 상기 확장 사전용 DB에 상기 엔트리를 추가하는 용어 처리부와, 목표 언어 특허 문서를 입력받아 상기 확장 사전용 DB에서 복수의 대역어들에 대해 특허 분야에 적합한 대역어를 선택하고, 복합어에 대해서는 특화된 단일어 대역어로 번역한 후, 상기 특허 번역 사전용 DB에 기록하는 대역어 특화부를 포함함으로써, 특허 분야별로 대역어가 구분된 대용량의 특허 번역 사전을 구축하는 시간과 비용을 절감할 수 있다.

대표도 - 도1



(72) 발명자

이기영

대전 서구 만년동 초원아파트 107-1406

노윤형

대전 유성구 신성동 208-9 301호

홍문표

대전 유성구 지족동 반석마을 102동 401호

김영길

대전 중구 오류동 삼성아파트 1-1005

서영애

대전 유성구 지족동 열매마을 대우아파트 303동 1003호

김창현

대전 유성구 송강동 송강마을아파트 204-1405

양성일

대전 유성구 신성동 두레아파트 110-1008

박상규

대전 유성구 송강동 한마을아파트 105-1305

특허청구의 범위

청구항 1

기구축된 전문 용어를 입력받아 특허 번역 사전용 DB에 기록 가능한 형태로 상기 입력된 전문 용어의 포맷을 변경하여 확장 사전용 DB에 기록하는 용어 재구성부;

출발 언어 특허 문서를 입력받고, 상기 출발 언어 특허 문서로부터 상기 확장 사전용 DB에 기록되지 않은 미등록 전문 용어를 추출하고, 상기 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하며, 대역어가 부착되지 않은 상기 단일어 엔트리에 대해 상기 대역어를 부착한 후, 상기 확장 사전용 DB에 상기 엔트리를 추가하는 용어 처리부; 및

목표 언어 특허 문서를 입력받고, 상기 목표 언어 특허 문서에 포함된 단일어에 대하여는 상기 확장 사전용 DB에 기록된 복수의 대역어 중 상기 목표 언어 특허 문서가 속하는 특허 분야에 대응하는 대역어를 선택하고, 상기 목표 언어 특허 문서에 포함된 복합어에 대하여는 상기 목표 언어 특허 문서가 속하는 특허 분야에 대응되는 단일어 대역어로 번역한 후, 상기 특허 번역 사전용 DB에 기록하는 대역어 특화부를 포함하는 특허 번역 사전 구축 장치.

청구항 2

제 1 항에 있어서, 상기 용어 재구성부는,

상기 기구축된 전문 용어를 입력받아 상기 기구축된 전문 용어의 표제어 언어 및 대역어 언어와 일반 번역 사전용 DB에 기록된 언어를 비교하는 용어 비교부;

상기 비교결과에 따른 상기 전문 용어를 상기 일반 번역 사전용 DB에 기록 가능한 포맷으로 변환하는 용어 변환부;

상기 변환된 전문 용어가 상기 일반 번역 사전용 DB에 존재하는 엔트리인 경우, 상기 일반 번역 사전용 DB의 상기 엔트리에 대응하는 표제어에 상기 전문 용어의 대역어를 추가시키고, 존재하지 않는 엔트리인 경우, 상기 전문 용어를 상기 일반 번역 사전용 DB의 새로운 엔트리로 추가시키는 용어 제거부; 및

상기 일반 번역 사전용 DB의 엔트리와 상기 미등록 전문 용어를 단일어 엔트리와 복합어 엔트리로 각각 분리하고, 상기 분리된 단일어 엔트리 중에서 대역어가 결합된 단일어 엔트리에 대해서는 확장 사전용 DB에 기록하고, 상기 대역어가 결합되지 않은 단일어 엔트리에 대해서는 해당 대역어를 부착하여 상기 확장 사전용 DB에 기록하는 전문 용어 처리부를 포함하는 것을 특징으로 하는 특허 번역 사전 구축 장치.

청구항 3

제 2 항에 있어서, 상기 용어 비교부는,

상기 비교결과, 상기 일반 번역 사전용 DB의 표제어 언어와 상기 전문 용어의 표제어 언어가 동일하거나, 상기 일반 번역 사전용 DB의 표제어 언어와 상기 전문 용어의 대역어 언어가 동일한 경우에만, 상기 용어 변환부로 상기 전문 용어를 전달하는 것을 특징으로 하는 특허 번역 사전 구축 장치.

청구항 4

제 1 항에 있어서, 상기 용어 처리부는,

대용량의 상기 출발 언어 특허 문서를 입력받아 IPC 분류에 따른 산업 분야로 분류하는 문서 분류부;

상기 분류된 특허 문서를 상기 확장 사전용 DB에 기록되지 않은 상기 특허 분야별 상기 출발 언어 미등록 전문 용어를 추출하는 용어 추출부;

상기 추출된 미등록 전문 용어를 상기 특허 분야별 및 빈도 순으로 정렬하는 미등록 용어 정렬부; 및

상기 미등록 전문 용어를 단일어 엔트리와 복합어 엔트리로 각각 분리하고, 상기 분리된 단일어 엔트리 중에서 대역어가 결합된 단일어 엔트리에 대해서는 확장 사전용 DB에 기록하고, 상기 대역어가 결합되지 않은 단일어 엔트리에 대해서는 해당 대역어를 부착하여 상기 확장 사전용 DB에 기록하는 전문 용어 처리부를 포함하는 것을 특징으로 하는 특허 번역 사전 구축 장치.

청구항 5

제 1 항에 있어서, 상기 대역어 특화부는,

대용량의 상기 목표 언어 특허 문서를 입력받아 IPC 분류에 따른 산업 분야로 분류하는 문서 분류부;

상기 분류된 특허 문서상의 어휘들을 상기 단일어 및 복합어로 분류하는 용어 분석부;

상기 분류된 단일어 및 복합어를 상기 특허 분야별 및 빈도순으로 정렬하는 용어 정렬부;

상기 확장 사전용 DB에서 상기 목표 언어의 대역어가 붙지 않은 복합어에 대해 해당 대역어를 추가하여 상기 확장 사전용 DB에 기록하는 복합어 자동 번역부; 및

상기 용어 정렬부에서 정렬된 빈도순 목록 및 상기 용어 처리부에서 정렬된 빈도순 목록을 토대로 상기 확장 사전용 DB에 기록된 대역어에 대해 상기 특허 분야에 대응하는 대역어를 선택하여 상기 특허 번역 사전용 DB에 기록하는 용어 선택부를 포함하는 것을 특징으로 하는 특허 번역 사전 구축 장치.

청구항 6

제 5 항에 있어서, 상기 용어 선택부는,

상기 확장 사전용 DB에 기록된 엔트리가 입력되면, 현재의 엔트리가 마지막인지 판단하는 제 1 판단 수단;

상기 엔트리에 대한 표제어의 대역어 중 하나를 디폴트 대역어로 결정하는 제 1 결정수단;

상기 엔트리의 표제어가 상기 특허 분야별 및 빈도순 목록에서 특정 누적빈도 이상인지를 판단하는 제 2 판단 수단;

상기 제 2 판단 수단의 판단 결과, 상기 특정 누적 빈도 이하인 경우에, 상기 용어 정렬부에서 정렬된 빈도순 목록에서의 특허 전 분야의 디폴트 대역어 보다 상기 용어 처리부에서 정렬된 빈도순 목록에서의 특허 특정 분야의 디폴트 대역어가 더 고빈도인 경우에만 상기 용어 처리부에서 정렬된 빈도순 목록에서 상기 특정 분야의 대역어를 결정하는 제 2 결정 수단; 및

상기 제 1 내지 제 2 결정 수단에서 결정된 대역어를 상기 특허 번역 사전용 DB에 기록하는 기록 수단을 포함하는 것을 특징으로 하는 특허 번역 사전 구축 장치.

청구항 7

a) 특허 번역 사전에 기록 가능한 형태로 기존에 구축된 전문 용어의 포맷을 변경하여 확장 사전을 구축하는 단계;

b) 상기 확장 사전에 등록되지 않은 미등록 전문 용어를 출발 언어 특허 문서로부터 추출하고, 상기 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하고 대역어가 부착되지 않은 단일어 엔트리에 대해 해당 대역어를 부착한 후에 상기 확장 사전에 상기 엔트리를 추가하는 단계; 및

c) 상기 확장 사전에서 특정 표제어에 대해 복수의 대역어들이 존재하는 경우, 목표 언어 전문 용어를 토대로 상기 복수의 대역어들 중에서 특허 분야에 대응하는 대역어를 선택하는 단계를 포함하는 것을 특징으로 하는 특허 번역 사전 구축 방법.

청구항 8

제 7 항에 있어서, 상기 c) 단계는,

대용량의 목표 언어 특허 문서를 입력받아 IPC 분류에 따른 산업 분야로 분류하는 단계;

상기 분류된 특허 문서상의 어휘들을 단일어 및 복합어로 분류하는 단계;

상기 분류된 단일어 및 복합어를 특허 분야별 빈도순으로 정렬하는 단계;

상기 확장 사전에서 목표 언어 대역어가 붙지 않은 복합어에 대해 해당 대역어를 추가하여 상기 확장 사전에 기록하는 단계; 및

상기 b) 단계에서 정렬된 고빈도순 목록과 상기 c) 단계에서 정렬된 고빈도순 목록을 토대로 확장 사전에 기록

된 대역어 중 상기 특허 분야에 대응하는 대역어를 선택하는 단계를 포함하는 것을 특징으로 하는 특허 번역 사전 구축 방법.

청구항 9

특허 번역 사전 구축 과정에서 특정 표제어에 대해 복수의 대역어들 중 특허 분야에 대응하는 대역어를 선택하는 방법으로서,

- 1) 확장 사전에 기록된 엔트리가 입력되면, 상기 엔트리가 마지막인지 판단하는 단계;
- 2) 상기 대역어들 중 하나를 특허 전분야의 디폴트 대역어로 결정하는 단계;
- 3) 상기 엔트리의 표제어가 출발 언어 전문 용어의 빈도순 목록에서 특정 누적빈도 이상인지를 판단하는 단계; 및
- 4) 상기 판단 결과 상기 특정 누적빈도 이하인 경우, 상기 2) 단계에서 결정된 특허 전 분야의 디폴트 대역어보다 상기 출발 언어 전문 용어의 빈도순 목록에서의 특정 분야 디폴트 대역어가 더 고빈도인 경우에만 상기 특정 분야의 대역어로 결정하는 단계를 포함하는 것을 특징으로 하는 특허 번역 사전 구축 과정에서의 대역어 선택 방법.

청구항 10

특허 번역 사전을 구축하는 프로그램을 기록한 컴퓨터로 판독 가능한 기록 매체로서,

상기 프로그램은,

상기 특허 번역 사전에 기록 가능한 형태로 기존에 구축된 전문 용어의 포맷을 변경하여 확장 사전을 구축하는 과정;

상기 확장 사전에 등록되지 않은 미등록 전문 용어를 출발 언어 특허 문서로부터 추출하고, 상기 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하고 대역어가 부착되지 않은 단일어 엔트리에 대해 해당 대역어를 부착한 후에 상기 확장 사전에 상기 엔트리를 추가하는 과정; 및

상기 확장 사전에서 특정 표제어에 대해 복수의 대역어들이 존재하는 경우, 목표 언어 전문 용어를 토대로 상기 복수의 대역어들 중에서 특허 분야에 대응하는 대역어를 선택하는 과정을 포함하는 것을 특징으로 하는 컴퓨터로 판독 가능한 기록 매체.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

- <6> 본 발명은 번역 사전을 구축하는 장치 및 방법에 관한 것으로, 더욱 상세하게는 대용량 단일어 특허 문서를 기반으로 특허 번역 사전을 구축하는 장치 및 방법에 관한 것이다.
- <7> 번역 사전을 구축하는 기술에는, 출발 언어(또는 원시 언어, source language)의 전문 용어에 대해 목표 언어(또는 목적 언어, target language)의 대역어를 수동으로 단순히 부착하는 방법이 있다. 이는, 예컨대 네트워크를 이용한 번역 시스템에서 출발 언어에 대해 기계 번역을 수행한 후, 출발 언어의 전문 용어에 대해서는 목표 언어의 대역어를 번역자들에 의해 수동으로 부착하는(번역하는) 방법이다.
- <8> 또한, 미등록어를 추출하여 그 미등록어들 사이의 관계 정보를 파악하는 기술을 그 예로서 들 수 있다. 이는, 예컨대 정보 검색 시스템에서 검색어 확장과 관련하여 미등록어를 추출하고 해당 미등록어에 대해 개념 정보를 부착한 후, 해당 개념 정보를 분석하여 수동으로 번역 사전에 등재 여부(예를 들면, 복합 명사의 경우에 모호성 제거)를 결정하는 방법이다.
- <9> 이러한 방법들은 결국 번역 사전을 구축하기 위해서 인간의 개입이 결정적인 요소를 이루며, 이러한 인간의 개입은 대용량의 번역 사전을 구축하는데 그 시간과 노력이 과다하게 소비되는 문제점이 있다.

발명이 이루고자 하는 기술적 과제

- <10> 따라서, 본 발명의 목적은 인간의 개입을 최소화하는 대용량의 특허 번역 사전을 구축하는 장치 및 방법을 제공하는데 있다.
- <11> 또한, 본 발명의 다른 목적은 대용량의 특허 번역 사전을 효율적이며 경제적으로 구축하는 장치 및 방법을 제공하는데 있다.
- <12> 이를 위하여, 본 발명의 일 실시예에 따른 특허 번역 사전 구축 장치는, 기구축된 전문 용어를 입력받아 특허 번역 사전용 DB에 기록 가능한 형태로 상기 입력된 전문 용어의 포맷을 변경하여 확장 사전용 DB에 기록하는 용어 재구성부; 출발 언어 특허 문서를 입력받아 상기 확장 사전용 DB에 기록되지 않은 미등록 전문 용어를 상기 특허 문서로부터 추출하고, 상기 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하며, 대역어가 부착되지 않은 상기 단일어 엔트리에 대해 상기 대역어를 부착한 후, 상기 확장 사전용 DB에 상기 엔트리를 추가하는 용어 처리부; 및 목표 언어 특허 문서를 입력받아 상기 확장 사전용 DB에서 복수의 대역어들에 대해 특허 분야에 적합한 대역어를 선택하고, 복합어에 대해서는 특화된 단일어 대역어로 번역한 후, 상기 특허 번역 사전용 DB에 기록하는 대역어 특화부를 포함한다.
- <13> 또한, 본 발명의 일 실시예에 따른 특허 번역 사전 구축 방법은, a) 특허 번역 사전에 기록 가능한 형태로 기존에 구축된 전문 용어의 포맷을 변경하여 확장 사전을 구축하는 단계; b) 상기 확장 사전에 등록되지 않은 미등록 전문 용어를 출발 언어 특허 문서로부터 추출하고, 상기 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하고 대역어가 부착되지 않은 단일어 엔트리에 대해 해당 대역어를 부착한 후에 상기 확장 사전에 상기 엔트리를 추가하는 단계; 및 c) 상기 확장 사전에서 특정 표제어에 대해 복수의 대역어들이 존재하는 경우, 목표 언어 전문 용어를 토대로 상기 복수의 대역어들 중에서 특허 분야에 적합한 대역어를 선택하는 단계를 포함한다.
- <14> 한편, 본 발명의 일 실시예에 따라 특허 번역 사전 구축 과정에서 특정 표제어에 대해 복수의 대역어들 중에서 특허 분야에 적합한 대역어를 선택하는 방법은, 1) 확장 사전에 기록된 엔트리가 입력되면, 상기 엔트리가 마지막인지 판단하는 단계; 2) 상기 대역어들 중 하나를 특허 전분야의 디폴트 대역어로 결정하는 단계; 3) 상기 엔트리의 표제어가 출발 언어 전문 용어의 빈도순 목록에서 특정 누적빈도 이상인지를 판단하는 단계; 및 4) 상기 판단 결과 상기 특정 누적빈도 이하인 경우, 상기 2) 단계에서 결정된 특허 전 분야의 디폴트 대역어보다 상기 출발 언어 전문 용어의 빈도순 목록에서의 특정 분야 디폴트 대역어가 더 고빈도인 경우에만 상기 특정 분야의 대역어로 결정하는 단계를 포함한다.

발명의 구성 및 작용

- <15> 이하에서는, 첨부 도면 및 실시예를 참조하여 본 발명을 상세히 설명한다. 도면상에서 동일 또는 유사한 구성 요소에 대하여는 동일한 참조번호를 부여하였다.
- <16> 도 1은 본 발명의 일 실시예에 따른 특허 번역 사전 구축 장치를 나타내는 구성도이다.
- <17> 도 1에 도시한 바와 같이, 이 장치는, 기존에 구축된 전문 용어를 토대로 확장된 일반 번역 사전용 DB(900, 이하 확장 사전용 DB라 함)를 구축하는 용어 재구성부(200)와, 이 확장 사전용 DB에 등록되지 않은 미등록 전문 용어를 부가하는 용어 처리부(300)와, 이 확장 사전용 DB를 토대로 특허 분야에 맞는 대역어로 번역하여 본 발명에 따른 특허 번역 사전용 DB(500)를 반자동으로 구축하는 대역어 특화부(400)를 포함한다.
- <18> 또한, 상기 장치는 본 발명에 따른 특허 번역 사전용 DB(500)를 구축하기 위해서 일반 번역 사전용 DB(100)와, 기존의 전문 용어를 입력받는 기구축 용어 입력부(600)와, 출발 언어 특허 문서를 입력받는 출발 언어 입력부(700)와, 목표 언어 특허 문서를 입력받는 목표 언어 입력부(800)를 더 포함한다.
- <19> 이와 같이 구성된 본 발명의 일 실시예에 따른 특허 번역 사전 구축 장치의 동작을 간략하게 설명한다.
- <20> 먼저, 용어 재구성부(200)는 기구축 용어 입력부(600)에서 기존에 구축된 전문 용어를 입력받아 일반 번역 사전용 DB(100)를 토대로 특허 번역 사전용 DB(500)에 기록 가능한 형태로 그 포맷을 변경하여 확장 사전용 DB(900)를 구축한다.
- <21> 이어, 용어 처리부(300)는 출발 언어 입력부(700)로부터 특정 출발 언어 특허 문서를 입력받아 새로 구축된 확장 사전용 DB(900)에 등록되지 않은 미등록 전문 용어를 이 특허 문서로부터 추출한다. 또한, 용어 처리부

(300)는 이 추출된 미등록 전문 용어 중에서 단일어 엔트리를 분류하고 대역어가 부착되지 않은 단일어 엔트리에 대해서는 수동으로 부착한 후에 확장 사전용 DB(900)에 그 엔트리를 추가한다.

<22> 이후, 대역어 특화부(400)는 목표 언어 입력부(800)로부터 목표 언어 특허 문서를 입력받아 미등록 전문 용어가 추가된 확장 사전용 DB(900)에서 복수의 대역어들에 대해 특허 분야에 적합한 대역어를 선택하고, 복합어에 대해서는 특화된 단일어 대역어로 번역하여 특허 번역 사전용 DB(500)를 구축한다.

<23> 이하, 본 발명에 따른 특허 번역 사전용 DB의 구축 장치를 첨부된 도면을 참조하여 더욱 상세히 설명한다.

<24> 도 2는 도 1의 용어 재구성부를 나타내는 상세 구성도이다.

<25> 도 2에 도시한 바와 같이, 용어 재구성부(200)는 일반 번역 사전용 DB(100)의 일반 용어와 기존에 구축된 전문 용어의 표제어 및 대역어를 비교하는 용어 비교부(210)와, 용어 비교부(210)의 비교 결과에 따라 입력되는 기존에 구축된 전문 용어를 일반 번역 사전용 DB(100)의 포맷으로 변환시키는 용어 변환부(220)와, 변환된 전문 용어를 일반 번역 사전용 DB(100)에 그 엔트리의 존재 여부에 따라 상기 DB(100)에 삭제 또는 추가하는 용어 제거부(230)와, 추가된 전문 용어를 단일어 엔트리와 복합어 엔트리로 분리하여 확장 사전용 DB(900)에 개별적으로 추가하는 전문 용어 처리부(240)를 포함한다.

<26> 또한, 용어 재구성부(200)는 용어 비교부(210)에 참조되는 일반 번역 사전용 DB(100)와, 기존의 전문 용어를 입력받는 기구축 용어 입력부(600)를 더 포함한다.

<27> 이와 같이 구성된 용어 재구성부(200)를 더욱 상세히 설명한다.

<28> 용어 비교부(210)는, 기존에 구축된 전문 용어를 입력받아 이 전문 용어의 표제어 언어 및 대역어 언어와 일반 번역 사전용 DB(100)에 기록된 언어를 비교한다. 비교결과, 상기 DB(100)의 표제어 언어와 상기 전문 용어의 표제어 언어가 동일하거나, 상기 DB(100)의 표제어 언어와 상기 전문 용어의 대역어 언어가 동일하면, 용어 비교부(210)는 용어 변환부(220)로 그 전문 용어를 전달하고, 그 이외에는 그 전문 용어를 용어 변환부(220)로 전달하지 않고 그냥 반환한다.

<29> 용어 변환부(220)는 전달된 전문 용어를 일반 번역 사전용 DB(100)에 적합한 포맷으로 변환한다. 용어 제거부(230)는 이 변환된 전문 용어가 일반 번역 사전용 DB(100)에 존재하는 엔트리인 경우에 이 DB(100)의 해당 표제어에 전문 용어의 대역어를 추가시키고, 존재하지 않는 엔트리인 경우에 그 전문 용어를 이 DB(100)의 새로운 엔트리로 추가시킨다.

<30> 전문 용어 처리부(240)는 상기 DB(100)의 엔트리와 상기 전문 용어를 단일어 엔트리와 복합어 엔트리로 각각 분리한다. 또한, 전문 용어 처리부(240)는, 이렇게 분리된 단일어 엔트리 중에서 대역어가 결합된 단일어 엔트리에 대해서는 확장 사전용 DB(900)에 기록하고, 대역어가 결합되지 않은 단일어 엔트리에 대해서는 적합한 대역어를 수동으로 부착하여 확장 사전용 DB(900)에 기록한다. 또한, 전문 용어 처리부(240)는 상기 분리된 복합어 엔트리를 그대로 확장 사전용 DB(900)에 기록한다.

<31> 예컨대, 1) 일반 번역 사전용 DB(100)에는

<32> solution -> {솔루션, 용액, 설명, 해결책},

<33> fluid -> {유동체, 액체},

<34> melting -> {용해, 융해} 등과 같이 기록되고,

<35> 2) 기존에 구축된 전문 용어로는

<36> 솔루션 -> {solution},

<37> 용액 -> {solution, fluid},

<38> 용해 -> {solution, melting} 등과 같이 기록되었다고 가정하면, 용어 비교부(210)는, 상기 DB(100)의 표제어 언어와 상기 전문 용어의 대역어 언어가 일치하는 경우(예컨대, solution, fluid, melting)를 비교/검출한다.

<39> 이어, 용어 변환부(220) 및 용어 제거부(230)는, 이 전문 용어가 상기 DB(100)에 존재하는 엔트리이므로, 상기 DB(100)의 해당 표제어에 전문 용어의 대역어를 추가시킨다. 즉, solution 은 {솔루션, 용액, 설명, 해결책, 용해}, fluid 는 {유동체, 액체, 용액}가 되고, melting 은 이전과 같은 {용해, 융해}가 된다.

<40> 이후, 전문 용어 처리부(240)는, 상술한 solution, fluid, melting은 모두 단일어 엔트리이고 대역어가 부착되

있으므로 확장 사전용 DB(900)에는 다음과 같이 기록된다.

- <41> solution -> {솔루션, 용액, 설명, 해결책, 용해},
- <42> fluid -> {유동체, 액체, 용액},
- <43> melting -> {용해, 융해}
- <44> 도 3은 도 1의 용어 처리부를 나타내는 상세 구성도이다.
- <45> 도 3에 도시한 바와 같이, 용어 처리부(300)는, 출발 언어로 구성된 대용량의 특허 문서를 입력받아 IPC 분류에 따른 산업 분야로 분류하는 문서 분류부(310)와, 분류된 특허 분야별 문서로부터 일반 번역 사전용 DB(100)에 등재되지 않은 미등록 전문 용어를 추출하는 용어 추출부(330)와, 추출된 미등록 전문 용어를 출현 빈도순으로 정렬하는 미등록 용어 정렬부(350)와, 정렬된 미등록 전문 용어를 단일어 엔트리와 복합어 엔트리로 분리하여 확장 사전용 DB(900)에 개별적으로 추가하는 전문 용어 처리부(240)를 포함한다.
- <46> 또한, 용어 처리부(300)는 출발 언어 특허 문서를 입력받는 출발 언어 입력부(700)와, 특허 문서 분류를 위해서 IPC와 산업분류 사이를 맵핑하는 맵핑 테이블(320)과, 형태소 분석기(340)를 더 포함한다.
- <47> 이와 같이 구성된 용어 처리부(300)를 더욱 상세히 설명한다.
- <48> 문서 분류부(310)는 대용량의 출발 언어 특허 문서를 입력받아 맵핑 테이블(320)을 참조하여 IPC 분류에 따른 산업 분야로 분류한다. 여기서, IPC 분류에 따른 산업 분야는 12개(즉, 기계, 농림수산, 사무용품, 섬유, 음료 식품, 의료위생, 잡화, 전기전자, 채광금속, 컴퓨터, 토목건설, 화학일반)로 분류된다.
- <49> 용어 추출부(330)는 상술한 바와 같이 분류된 특허 문서를 형태소 분석기(340)를 이용하여 확장 사전용 DB(900)에 기록되지 않은 특허 분야별 출발 언어 미등록 전문 용어를 추출한다. 여기서, 형태소 분석기(340)는 이 기술분야의 숙련자에게 공지되어 있으므로, 본 발명을 명확히 하기 위해 그 상세한 설명은 생략한다. 또한, 미등록 용어 정렬부(350)는 추출된 미등록 전문 용어를 분야별로 고빈도 순으로 정렬한다.
- <50> 이후, 전문 용어 처리부(240)는 각 분야별로 고빈도 순으로 정렬된 전문 용어를 단일어 엔트리와 복합어 엔트리로 각각 분리한다. 또한, 전문 용어 처리부(240)는, 이렇게 분리된 단일어 엔트리 중에서 대역어가 결합된 단일어 엔트리에 대해서는 확장 사전용 DB(900)에 기록하고, 대역어가 결합되지 않은 단일어 엔트리에 대해서는 적합한 대역어를 수동으로 부착하여 확장 사전용 DB(900)에 기록한다. 또한, 전문 용어 처리부(240)는 상기 분리된 복합어 엔트리를 그대로 확장 사전용 DB(900)에 기록한다.
- <51> 예컨대, 도 3에서 설명한 출발 언어 특허 문서를 기반으로 미등록 전문 용어를 추출한 후, 단일어 대역어가 없는 엔트리에는 대역어를 부착하여 구축된 확장 사전용 DB(900)의 예가 다음 표 1과 같이 도시된다.

표 1

분야	추출된_미등록어	빈도수
전기전자	nitride-based	4068
	untrusted	1579
	intermediate_transfer	52747
기계	paintball	11939
	in-cylinder	5411
	pressurized_fluid	34464
화학일반	crosslinked	37606
	alkoxylated	10862
	condensable_hydrocarbon	253376

- <53> 표 1은 출발 언어 특허 문서로부터 추출된 특허 분야별 출발 언어 전문 용어 고빈도 리스트의 예를 나타낸다. 이 표 1을 토대로 구축된 확장 사전용 DB(900)에 기록된 예를 다음 표 2에 나타낸다.

표 2

<54>

표제어	대역어
solution	{솔루션, 용액, 설명, 해결책, 용해}
fluid	{유동체, 액체, 용액}
melting	{용해, 용해}
nitride_based	{질화물_기반}
untrusted	{비신뢰성}
intermediate_transfer	{ }
paintball	{페인트볼}
in-cylinder	{실린더_안}
pressurized_fluid	{ }
crosslinked	{가교결합}
alkoxylated	{알콕실화 }
condensable_hydrocarbon	{ }

<55>

도 4는 도 1의 대역어 특화부를 나타내는 상세 구성도이다.

<56>

도 4에 도시한 바와 같이, 대역어 특화부(400)는 목표 언어로 구성된 대용량의 특허 문서를 입력받아 IPC 분류에 따른 산업 분야로 분류하는 문서 분류부(310)와, 분류된 특허 문서상의 어휘들을 단일어 및 복합어로 분류하는 용어 분석부(420)와, 분류된 단일어 및 복합어를 특허 분야별 빈도순으로 정렬하는 용어 정렬부(440)와, 도 3에서 기 설명된 미등록 용어 정렬부(350)와, 용어 정렬부(440) 및 미등록 용어 정렬부(350)에서 정렬된 전문 용어 목록을 토대로 확장 사전용 DB(900)에 기록된 대역어에 대해 특화된 대역어를 선택하는 용어 선택부(450)를 포함한다.

<57>

또한, 대역어 특화부(400)는 목표 언어 특허 문서를 입력받는 목표 언어 입력부(800)와, 특허 문서 분류를 위해서 IPC와 산업분류 사이를 맵핑하는 맵핑 테이블(320)과, 형태소 분석기(340)와, 복합어 자동 번역부(460)와, 확장 사전용 DB(900)와, 특허 정보가 부착된 대역어 정보를 반영한 특허 번역 사전용 DB(500)를 더 포함한다.

<58>

이와 같이 구성된 대역어 특화부(400)를 더욱 상세히 설명한다.

<59>

문서 분류부(310)는 목표 언어로 구성된 대용량의 특허 문서를 입력받아 맵핑 테이블(320)을 이용하여 IPC 분류에 따른 산업 분야로 분류한다. 여기서, IPC 분류에 따른 산업 분야는 12개(즉, 기계, 농업수산, 사무용품, 섬유, 음식식품, 의료위생, 잡화, 전기전자, 채광금속, 컴퓨터, 토목건설, 화학일반)로 분류된다.

<60>

용어 분석부(420)는 형태소 분석기(340)를 이용하여 상기 분류된 특허 문서상의 어휘들을 단일어 및 복합어로 분류한다. 용어 정렬부(440)는 이와 같이 분류된 단일어 및 복합어를 특허 분야별 고빈도순으로 정렬한다.

<61>

복합어 자동 번역부(460)는 도 3에서 구축된 확장 사전용 DB(900)에서 목표 언어 대역어가 붙지 않은 복합어를 자동 번역하여 상술한 확장 사전용 DB(900)에 추가한다. 이 결과로서, 확장 사전용 DB(900)에 기록된 단일어 및 복합어는 모두 목표 언어 대역어를 가진다. 여기서, 복합어 자동 번역부(460)는 기계적인 복합어 번역기를 이용할 수도 있으며, 경우에 따라서는 수동으로도 가능하다.

<62>

또한, 용어 선택부(450)는 상술한 용어 정렬부(440)에서 정렬된 특허 분야별 목표 언어 전문 용어 고빈도순 목록과 미등록 용어 정렬부(350)에서 정렬된 특허 분야별 출발 언어 전문 용어 고빈도순 목록을 토대로 확장 사전용 DB(900)에 기록된 대역어에 대해 특허 분야별 특화된 대역어를 선택한다(특화된 대역어 선택에 대한 상세한 설명은 도 5를 토대로 후술한다). 이렇게 함으로써 특허 분야의 정보가 반영된 특허 번역 사전용 DB(500)를 구축할 수 있다.

<63>

상술한 대역어 특화부(400)에 대한 이해를 돕기 위해서, 확장 사전용 DB(900)에 존재하는 대역어가 2개 이상인 엔트리에 대해 특허 분야별로 대역어를 구분하여 특허 번역 사전용 DB(500)를 구축하는 방법을 구체적인 예를 들어 설명한다.

<64>

우선, 확장 사전용 DB(900)에 기록된 예가 전술한 표 2와 같이 도시된다. 이 가운데 대역어가 2개 이상인 엔트리에 대한 표제어는 solution, fluid, melting임을 알 수가 있다.

<65>

이어, 문서 분류부(310)와 용어 분석부(420)를 거쳐 용어 정렬부(440)에서 정렬된 목표 언어에 대해 특허 분야

별 고빈도순에 따라 표제어를 정리한다. 예컨대, solution의 경우 화학일반 분야(예컨대, "PAT-화학일반"으로 표시)와 의료위생 분야(예컨대, "PAT-의료위생"으로 표시)에서는 {용액}이 그 대역어로, 특허 전 분야의 디폴트로는 {솔루션, 용액, 설명, 해결책, 용해}가 그 대역어로 정리된다. 왜냐하면, solution은 화학일반 분야와 의료위생 분야에서는 {용액}이 대역어로 쓰인 경우가 특허 전분야의 디폴트인 {솔루션}보다 그 빈도수가 높기 때문에 특정 분야의 대역어가 될 수 있다.

<66> 특허 분야별로 대역어를 구분한 예는 다음 표 3과 같이 도시된다.

표 3

<67>

표제어	대역어
solution	{PAT-화학일반 용액} {PAT-의료위생 용액} {PAT-DEFAULT 솔루션, 용액, 설명, 해결책, 용해}
fluid	{PAT-DEFAULT 유동체, 액체, 용액}
melting	{PAT-DEFAULT 용해, 융해}
nitride_based	{PAT-DEFAULT 질화물_기반}
untrusted	{PAT-DEFAULT 비신뢰성}
intermediate_transfer	{PAT-DEFAULT 중간_전송}
paintball	{PAT-DEFAULT 페인트볼}
in-cylinder	{PAT-DEFAULT 실린더_안}
pressurized_fluid	{PAT-DEFAULT 가압된_유체}
crosslinked	{PAT-DEFAULT 가교결합}
alkoxylated	{PAT-DEFAULT 알콕실화 }
condensable_hydrocarbon	{PAT-DEFAULT 응축가능_탄화수소}

<68> 한편, 상술한 설명과는 대조적으로 fluid의 경우에는 특허 전 분야의 디폴트인 {용액} 보다는 더 고빈도로 나타나는 특정 분야의 대역어가 없기 때문에 "PAT-DEFAULT"가 할당된다.

<69> 또한, 표 2에서 대역어가 존재하지 않았던 복합어는 복합어 자동 번역부(460)에서 자동 번역되어 해당 표제어에 대해 대역어가 부착된다.

<70> 도 5는 도 4의 용어 선택부의 동작을 나타내는 순서도로서, 용어 선택부(450)가 대역어 특화부(400)에서 확장 사전용 DB(900)에 기록된 대역어를 토대로 분야별 전문 용어를 선택하는 방법을 도시한다.

<71> 먼저, 확장 사전용 DB(900)에 기록된 엔트리가 입력되면, 현재의 엔트리가 마지막인지 판단한다(S501). 판단 결과, 마지막이면 전체 동작을 종료한다.

<72> 다음으로, 상기 엔트리에 대한 표제어의 대역어가 1개인가를 판단한다(S503). 판단 결과, 1개이면 특허 번역 사전용 DB(500)에 해당 표제어에 대한 대역어를 추가/기록하고(S504), 단계 S501을 다시 수행한다. 그러나, 복수이면 단계 S505를 수행한다.

<73> 그러면, 현재의 엔트리에 대한 표제어의 대역어가 복수이므로, 이들 중 하나를 특허 전분야의 디폴트 대역어로 설정한다(S505). 즉, 용어 정렬부(440)에 의해 정리된 특허 분야별 목표 언어 전문 용어의 고빈도순 목록에서 특허 전 분야에 걸쳐 가장 높은 빈도로 사용되는 대역어를 특허 전 분야의 디폴트 대역어로 결정한다.

<74> 이후, 미등록 용어 정렬부(350)에 의해 정리된 특허 분야별 출발 언어 전문 용어의 고빈도순 목록에서, 현재 엔트리의 표제어가 상기 목록에서 특정 누적빈도(예컨대, 전체 빈도 대비 상위로부터 누적 빈도가 70%) 이상인지를 판단한다(S506).

<75> 판단 결과, 그 이상인 경우에만 해당 출발 언어 표제어의 엔트리를 분리하고(S507), 번역자를 통해 그 표제어에 대한 특허 분야별 디폴트 대역어를 결정한다(S508). 이러한 결정은 경우에 따라서 수동으로 결정된다.

<76> 그 이하인 경우에는, 목표 언어 전문 용어 고빈도순 목록에서 특허 전 분야의 디폴트 대역어보다 상술한 출발 언어 전문 용어의 고빈도순 목록에서 특정 분야의 디폴트 대역어가 더 고빈도인 경우에만 상기 특정 분야의 대역어를 자동으로 설정한다(S509). 예컨대, 전술한 solution의 경우, 화학일반 분야와 의료위생 분야에서 {용액}이 대역어로 쓰인 경우가 이 경우에 해당한다.

<77> 이후, 상기 단계 S505 내지 단계 S509에서 결정 또는 선택된 대역어를 특허 번역 사전용 DB(500)에 기록하고

(S510), 다음 엔트리로 이동하여 단계 S501부터 반복적으로 수행하여 확장 사전용 DB(900)에 기록된 엔트리가 없어질 때까지 수행한다(S511).

<78> 지금까지 본 발명을 바람직한 실시예를 참조하여 상세히 설명하였지만, 당업자는 본 발명의 사상 및 범위를 벗어나지 않고 다양한 변형 또는 수정이 가능하다는 것을 알 것이다.

발명의 효과

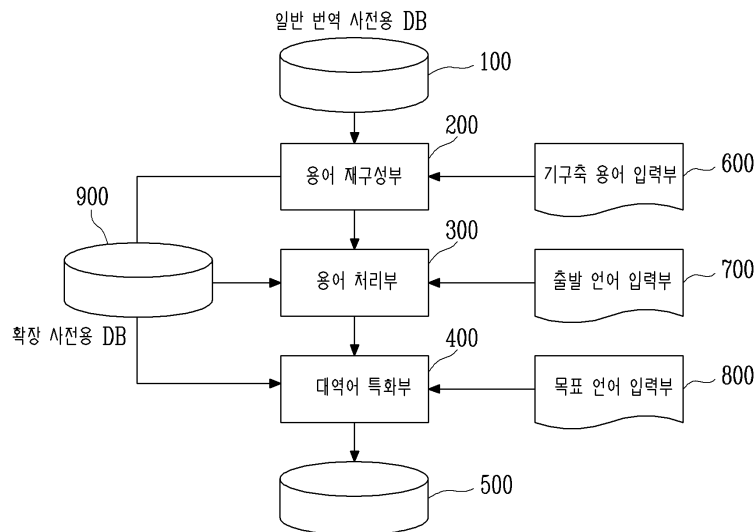
<79> 이상에서 상세히 설명한 바와 같이, 본 발명에 따르면, 종래의 일반 번역 사전과, 기존에 구축된 전문 용어와, 대용량의 특허 문서 등과 같이 쉽게 구할 수 있는 언어적 자원을 활용하여, 해당 분야의 전문가만이 알 수 있는 특허 문서 전문 용어에 대해, 특허 분야별로 대역어가 구분된 대용량의 특허 번역 사전을 구축하는 시간과 비용을 절약할 수 있는 효과가 있다.

도면의 간단한 설명

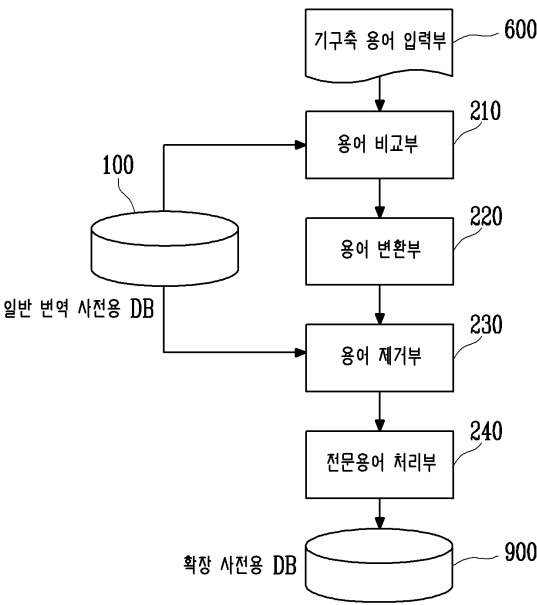
- <1> 도 1은 본 발명의 일 실시예에 따른 특허 번역 사전 구축 장치를 나타내는 구성도.
- <2> 도 2는 도 1의 용어 재구성부를 나타내는 상세 구성도.
- <3> 도 3은 도 1의 용어 처리부를 나타내는 상세 구성도.
- <4> 도 4는 도 1의 대역어 특화부를 나타내는 상세 구성도.
- <5> 도 5는 도 4의 용어 선택부의 동작을 나타내는 순서도.

도면

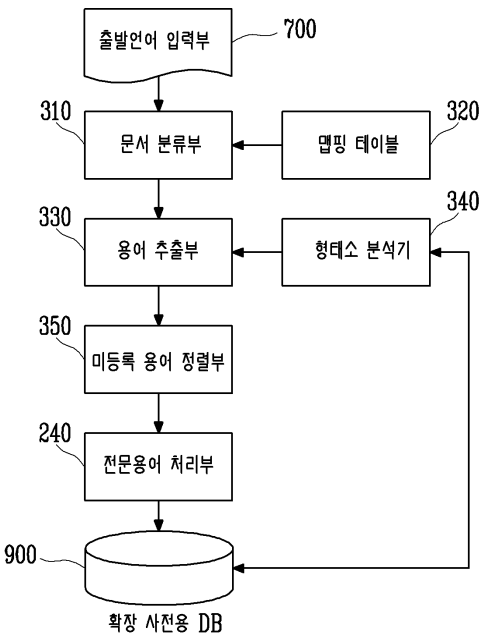
도면1



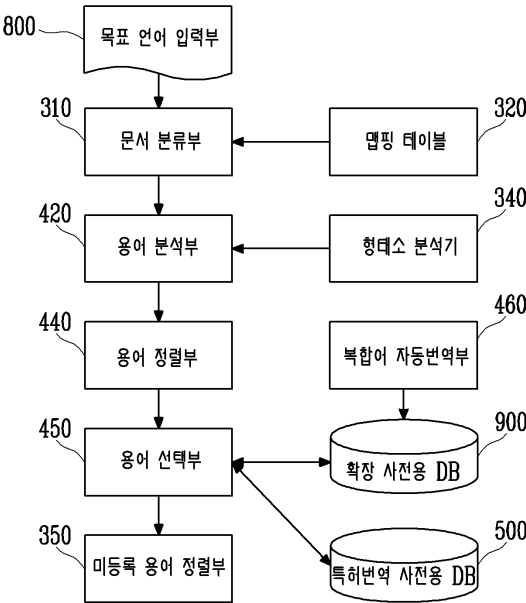
도면2



도면3



도면4



도면5

