(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2011-0094880
(43) 공개일자 2011년08월24일

(51) Int. Cl.
G06Q 50/00 (2006.01)
(21) 출원번호 10-2010-0014593
(22) 출원일자 2010년02월18일
심사청구일자 2010년02월18일

(71) 출원인
고려대학교 산학협력단
서울 성북구 안암동5가 1
(72) 발명자
박상성
서울시 노원구 상계8동 주공아파트 1602-701
신영근
서울시 성북구 안암동5가 134-68번지 305호
(뒷면에 계속)
(74) 대리인
특허법인남촌

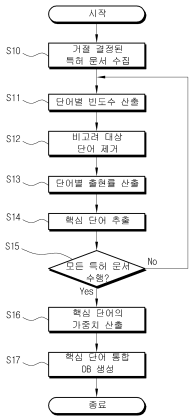
전체 청구항 수 : 총 3 항

(54) 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법

(57) 요 약

본 발명은 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법에 관한 것으로, (a1) 다수의 거절 결정된 특허 문서에 포함된 단어들을 추출하여, 상기 거절 결정된 특허 문서별로 상기 추출된 단어의 빈도수를 산출하는 단계와, (a2) 상기 산출된 빈도수에 기초하여, 상기 거절 결정된 특허 문서별로 단어들의 출현률을 산출하는 단계와, (a3) 상기 산출된 출현률이 기 설정된 추출 기준값보다 큰 값을 갖는 단어를 상기 각 거절 결정된 특허 문서의 핵심 단어로 추출하는 단계와, (a4) 상기 (a3) 단계에서 상기 다수의 거절 결정된 특허 문서별로 추출된 상기 핵심 단어에 대해 상기 (a2) 단계에서 산출된 상기 출현률의 평균값을 단어별로 산출하여 가중치로 결정하는 단계를 갖는 가중치 산출 단계와; (b1) 상기 (a1) 단계에서 사용되지 않은 다수의 특허 문서에 대해 상기 (a1) 단계, 상기 (a2) 단계 및 상기 (a3) 단계를 수행하여 상기 사용되지 않은 다수의 특허 문서의 핵심 단어를 추출하는 단계와, (b2) 상기 (a3) 단계에서 추출된 상기 핵심 단어와 상기 (b1) 단계에서 추출된 상기 핵심 단어 중 중복되는 중복 단어를 추출하는 단계와, (b3) 상기 사용되지 않은 다수의 특허 문서 중 상기 중복 단어를 포함하는 특허 문서별로 상기 중복 단어에 대한 상기 가중치와 상기 중복 단어에 대해 상기 (b1) 단계를 통해 산출된 상기 출현률을 곱한 값들의 합을 유사치로 산출하는 단계를 갖는 유사치 산출 단계와; (c) 상기 중복 단어를 포함하는 상기 특허 문서의 특허 등록 또는 거절 여부와 상기 유사치를 소정의 군집화 알고리즘에 적용하여 거절 기준값을 산출하는 단계를 갖는 거절 기준값 산출 단계와; (d1) 예측 대상 특허 문서에 대해 상기 가중치 산출 단계를 수행하여 상기 예측 대상 특허 문서의 단어별로 가중치를 산출하는 단계와, (d2) 상기 예측 대상 특허 문서에 대해 상기 유사치 산출 단계를 수행하여 상기 예측 대상 특허 문서에 대한 유사치를 산출하는 단계와, (d3) 상기 예측 대상 특허 문서의 상기 유사치와 상기 거절 기준값을 비교하여 상기 예측 대상 특허 문서의 특허 등록 가능성을 예측하는 단계를 갖는 예측 단계를 포함하는 것을 특징으로 한다.

대 표 도 - 도1



(72) 발명자

정원교

경기도 성남시 분당구 정자동 동양파라곤A 106동 604호

구정민

경상북도 경산시 중방동 842-13

장동식

서울시 성북구 안암동3가 삼익아파트 2동 1102호

특허청구의 범위

청구항 1

단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법에 있어서,

(a1) 다수의 거절 결정된 특허 문서에 포함된 단어들을 추출하여, 상기 거절 결정된 특허 문서별로 상기 추출된 단어의 빈도수를 산출하는 단계와,

(a2) 상기 산출된 빈도수에 기초하여, 상기 거절 결정된 특허 문서별로 단어들의 출현률을 산출하는 단계와,

(a3) 상기 산출된 출현률이 기 설정된 추출 기준값보다 큰 값을 갖는 단어를 상기 각 거절 결정된 특허 문서의 핵심 단어로 추출하는 단계와,

(a4) 상기 (a3) 단계에서 상기 다수의 거절 결정된 특허 문서별로 추출된 상기 핵심 단어에 대해 상기 (a2) 단계에서 산출된 상기 출현률의 평균값을 단어별로 산출하여 가중치로 결정하는 단계를 갖는 가중치 산출 단계와;

(b1) 상기 (a1) 단계에서 사용되지 않은 다수의 특허 문서에 대해 상기 (a1) 단계, 상기 (a2) 단계 및 상기 (a3) 단계를 수행하여 상기 사용되지 않은 다수의 특허 문서의 핵심 단어를 추출하는 단계와,

(b2) 상기 (a3) 단계에서 추출된 상기 핵심 단어와 상기 (b1) 단계에서 추출된 상기 핵심 단어 중 중복되는 중복 단어를 추출하는 단계와,

(b3) 상기 사용되지 않은 다수의 특허 문서 중 상기 중복 단어를 포함하는 특허 문서별로 상기 중복 단어에 대한 상기 가중치와 상기 중복 단어에 대해 상기 (b1) 단계를 통해 산출된 상기 출현률을 곱한 값들의 합을 유사치로 산출하는 단계를 갖는 유사치 산출 단계와;

(c) 상기 중복 단어를 포함하는 상기 특허 문서의 특허 등록 또는 거절 여부와 상기 유사치를 소정의 군집화 알고리즘에 적용하여 거절 기준값을 산출하는 단계를 갖는 거절 기준값 산출 단계와;

(d1) 예측 대상 특허 문서에 대해 상기 가중치 산출 단계를 수행하여 상기 예측 대상 특허 문서의 단어별로 가중치를 산출하는 단계와,

(d2) 상기 예측 대상 특허 문서에 대해 상기 유사치 산출 단계를 수행하여 상기 예측 대상 특허 문서에 대한 유사치를 산출하는 단계와,

(d3) 상기 예측 대상 특허 문서의 상기 유사치와 상기 거절 기준값을 비교하여 상기 예측 대상 특허 문서의 특허 등록 가능성을 예측하는 단계를 갖는 예측 단계를 포함하는 것을 특징으로 하는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법.

청구항 2

제1항에 있어서,

상기 가중치를 산출하는 단계에서는 기 설정된 비고려 대상 단어가 제외된 상태에서 수행되는 것을 특징으로 하는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법.

청구항 3

제1항 또는 제2항에 있어서,

상기 (c) 단계의 상기 군집화 알고리즘으로는 k-means 알고리즘이 적용되는 것을 특징으로 하는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법에 관한 것으로서, 보다 상세하게는 거절 결정된 특허 문서들에 기재된 단어들의 빈도수를 분석하고 이를 수치화하여 분석 대상이 되는 특허 문서의 특허 등

록 가능성을 확률적으로 예측해 볼 수 있는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법에 관한 것이다.

배 경 기 술

[0002] 세계 대부분의 나라에서는 산업 발전에 기여할 수 있는 개인의 발명을 특허제도를 이용하여 이를 보호하고 있다. 즉, 개인의 발명을 공개함으로서 산업 발전에 기여할 수 있도록 하고, 그 발명을 발명한 개인에게는 그 발명을 실시에 대한 독점적이고 배타적인 권리를 일정 기간 동안 부여하고 있다.

[0003] 이와 같은 특허 제도에 의해 보호를 받는 특허권은 일 개인에게 특정 기술의 실시에 대하여 독점적이고 배타적인 권리를 부여하기 때문에, 각 나라에서는 일정한 특허 요건을 갖춘 발명에 대해서만 특허 등록을 허여하고 있다.

[0004] 한국에서도 특허법을 제정하고, 특허법에서 규정하는 특허 요건을 만족하는 기술에 대해서만 특허 등록을 허여하고 있다. 이러한 특허 요건 중에는 해당 기술이 특허로서 보호받을 만한 기술인지 여부, 즉 특허성과 관련된 신규성과 진보성 관련 요건이 있다.

[0005] 신규성과 진보성을 포함하는 특허성과 관련된 특허 요건은 해당 발명이 특허출원된 후 심사 단계나 심판 등의 절차에서 특허청, 특허심판원 또는 법원에서 객관적인 심사 기준 하에서 판단하게 된다.

[0006] 그러나, 특허성의 판단이 객관적인 심사 기준 하에서 판단되어지더라도, 특허성, 특히 진보성이 갖는 판단의 모호함으로 인해 이를 규칙화하기 어려운 문제점이 있다. 이는, 특허법에 대한 지식이 부족한 일반 발명자들에는 해당 기술이 특허로서 보호받을 수 있는지 여부를 판단하기 어렵게 하는 요인으로 작용한다.

[0007] 따라서, 일반 발명자들은 자신의 발명이 특허 가능한지 여부를 판단하기 위해 전문가들에게 자문을 구해야하고, 이러한 자문의 기회가 부족한 경우에는 특허 등록 가능성에 대해 전혀 알지 못한 채 특허출원을 진행하게 된다.

발명의 내용

해결하려는 과제

[0008] 이에 본 발명은 상기와 같은 문제점을 해결하기 위해 안출된 것으로서, 기존의 특허 문헌들의 단어를 분석하고, 해당 단어들의 빈도수를 산출하여 이를 분석 및 수치화하여 분석 대상이 되는 특허 문헌의 특허 등록 가능성을 확률적으로 예측해 볼 수 있는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법을 제공하는데 그 목적이 있다.

과제의 해결 수단

[0009] 상기 목적은 본 발명에 따라, 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법에 있어서, (a1) 다수의 거절 결정된 특허 문서에 포함된 단어들을 추출하여, 상기 거절 결정된 특허 문서별로 상기 추출된 단어의 빈도수를 산출하는 단계와, (a2) 상기 산출된 빈도수에 기초하여, 상기 거절 결정된 특허 문서별로 단어들의 출현률을 산출하는 단계와, (a3) 상기 산출된 출현률이 기 설정된 추출 기준값보다 큰 값을 갖는 단어를 상기 각 거절 결정된 특허 문서의 핵심 단어로 추출하는 단계와, (a4) 상기 (a3) 단계에서 상기 다수의 거절 결정된 특허 문서별로 추출된 상기 핵심 단어에 대해 상기 (a2) 단계에서 산출된 상기 출현률의 평균값을 단어별로 산출하여 가중치로 결정하는 단계를 갖는 가중치 산출 단계와; (b1) 상기 (a1) 단계에서 사용되지 않은 다수의 특허 문서에 대해 상기 (a1) 단계, 상기 (a2) 단계 및 상기 (a3) 단계를 수행하여 상기 사용되지 않은 다수의 특허 문서의 핵심 단어를 추출하는 단계와, (b2) 상기 (a3) 단계에서 추출된 상기 핵심 단어와 상기 (b1) 단계에서 추출된 상기 핵심 단어 중 중복되는 중복 단어를 추출하는 단계와, (b3) 상기 사용되지 않은 다수의 특허 문서 중 상기 중복 단어를 포함하는 특허 문서별로 상기 중복 단어에 대한 상기 가중치와 상기 중복 단어에 대해 상기 (b1) 단계를 통해 산출된 상기 출현률을 곱한 값들의 합을 유사치로 산출하는 단계를 갖는 유사치 산출 단계와; (c) 상기 중복 단어를 포함하는 상기 특허 문서의 특허 등록 또는 거절 여부와 상기 유사치를 소정의 군집화 알고리즘에 적용하여 거절 기준값을 산출하는 단계를 갖는 거절 기준값 산출 단계와; (d1) 예측 대상 특허 문서에 대해 상기 가중치 산출 단계를 수행하여 상기 예측 대상 특허 문서의 단어별로 가중치를 산출하는 단계와, (d2)

상기 예측 대상 특허 문서에 대해 상기 유사치 산출 단계를 수행하여 상기 예측 대상 특허 문서에 대한 유사치를 산출하는 단계와, (d3) 상기 예측 대상 특허 문서의 상기 유사치와 상기 거절 기준값을 비교하여 상기 예측 대상 특허 문서의 특허 등록 가능성을 예측하는 단계를 갖는 예측 단계를 포함하는 것을 특징으로 하는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법에 의해서 달성된다.

[0010] 여기서, 상기 가중치를 산출하는 단계에서는 기 설정된 비고려 대상 단어가 제외된 상태에서 수행될 수 있다.

[0011] 그리고, 상기 (c) 단계의 상기 군집화 알고리즘으로는 k-means 알고리즘이 적용될 수 있다.

발명의 효과

[0012] 상기 구성에 따라 본 발명에 따르면, 기존의 특허 문헌들의 단어를 분석하고, 해당 단어들의 빈도수를 산출하여 이를 분석 및 수치화하여 분석 대상이 되는 특허 문헌의 특허 등록 가능성을 확률적으로 예측해 볼 수 있는 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법이 제공된다.

도면의 간단한 설명

[0013] 도 1은 본 발명에 따른 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법의 가중치 산출 단계를 설명하기 위한 도면이고,

도 2는 본 발명에 따른 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법의 유사치 산출 단계 및 거절 기준값 산출 단계를 설명하기 위한 도면이고,

도 3은 본 발명에 따른 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법의 등록 가능성 예측 단계를 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0014] 이하에서는 첨부된 도면들을 참조하여 본 발명에 따른 실시예에 대해 상세히 설명한다.

[0015] 본 발명에 따른 단어의 빈도수를 이용한 특허의 등록 가능성 예측 방법은, 도 1 내지 도 3에 도시된 바와 같이, 가중치 산출 단계, 유사치 산출 단계, 거절 기준값 산출 단계 및 특허 등록 가능성 예측 단계를 포함한다.

[0016] **[가중치 산출 단계]**

[0017] 본 발명에 따른 등록 가능성 예측 방법의 가중치 산출 단계를, 도 1을 참조하여 상세히 설명한다.

[0018] 먼저, 가중치의 산출을 위한 다수의 특허 문서가 수집된다(S10). 여기서, 가중치의 산출을 위해 사용되는 특허 문서는 거절 결정된 특허 문서가 수집된다. 이 때, 거절 결정된 특허 문서의 수집은 해당 특허 발명의 기술 분야별로 수집되어 기술 분야별로 등록 가능성 예측 방법이 적용되는 것이 바람직하다. 이는 기술 분야별로 사용되는 용어가 상이할 수 있음을 반영하게 된다.

[0019] 거절 결정된 다수의 특허 문서가 수집되면, 각 특허 문서별로 해당 특허 문서에 포함된 단어들을 추출하고, 추출된 단어들에 대해 단어별로 빈도수를 산출한다(S11). 그리고, 추출된 단어들은 산출된 빈도수에 따라 정렬된다.

[0020] 이 때, 1 번 문서부터 m번 문서까지 있을 경우 각 특허 문서를 y로 표기한다. 즉 y = 1, 2, 3, 4, ... , m이 된다. 여기서, 빈도수가 높을수록 해당 특허 문서에서 내용의 핵심이 되는 단어일 가능성이 커지게 된다.

[0021] 그런 다음, 빈도수에 따라 정렬된 단어로부터 기 설정된 비고려 대상 단어를 제거한다(S12). 즉, 단어들 중 기술적인 내용이 아닌 특별한 의미를 부여할 필요가 없는 단어, 예를 들어, 영어의 경우 'a'나 'the' 등의 관사나 'in'이나 'with' 등과 같은 전치사 , 그리고 'invention, claim, fig' 등과 같이 특허 명세서를 작성할 때 필수적으로 들어가는 단어들이 비고려 대상 단어에 속할 수 있다.

[0022] 상기와 같이 비고려 대상 단어가 제거된 상태에서, 각 단어가 해당 특허 문서에서 나타나는 비율에 해당하는 출현률을 단어별로 산출한다(S13). 그리고, 각 단어별로 산출된 출현률이 기 설정된 추출 기준값보다 큰 값을 갖

는 단어를 해당 특허 문서의 핵심 단어로 추출한다(S14).

[0023] 단어별 출현률의 산출과 핵심 단어의 추출 방법에 대해, [수학식 1] 및 [수학식 2]를 참조하여 보다 구체적으로 설명한다.

[0024] [수학식 1]

[0025] $$P_{iy} = \frac{W_{iy}}{\sum_{i=1}^{n} W_{iy}}$$

[0026] [수학식 1]의 P_{iy}는 특허 문서 y에 대한 단어 i의 출현률이고, W_{iy}는 특허 문서 y에 있는 단어 i의 빈도수이다. 여기서, y는 특허 문서의 집합에 해당하며, 상술한 바와 같이, m 개의 거절 결정된 특허 문서가 사용되는 경우, y = 1, 2, 3, 4, ... , m이 된다. 그리고, I는 특허 문서 y에 포함된 단어의 집합으로, 비고려 대상 단어가 제거된 상태를 의미한다.

[0027] 즉, [수학식 1]에서 단어 i의 출현률은 각 단어의 빈도수를 전체 단어들의 빈도수의 합으로 나눈 값으로 산출하게 된다.

[0028] [수학식 2]

[0029] $$C_y = \frac{1}{n}$$

[0030] 여기서, C_y는 특허 문서 y에 대한 추출 기준값이고, n는 특허 문서 y의 전체 단어의 수이다. 여기서, n은 상술한 비고려 대상 단어가 제거된 상태에서의 전체 단어의 수가 된다.

[0031] 상기와 같이, 특허 문서 y에 대한 추출 기준값 및 각 단어에 대한 출현률의 산출되면, $P_{iy} < C_y$인 단어를 모두 제거하고 나머지 단어를 특허 문서 y의 핵심 단어로 추출하게 된다. 즉, 추출 기준값은 특허 문서의 전체 단어 수인 n의 평균으로 하였으며, 평균 이상의 비중을 가지고 있는 단어들이 해당 특허 문서의 핵심적인 내용에 해당하는 핵심 단어로 추출된다.

[0032] 상기와 같은 방법으로 하나의 거절 결정된 특허 문서에서 핵심 단어가 추출되고 각 핵심 단어에 대한 출현률이 결정되면, 나머지 거절 결정된 특허 문서에 대해서도 S11 단계 내지 S15 단계를 거쳐 각 거절 결정된 특허 문서에 대한 핵심 단어들이 추출된다.

[0033] 상기와 같은 과정을 통해 거절 결정된 다수의 특허 문서로부터 핵심 단어들이 추출되면, 중복되는 핵심 단어가 존재하게 된다. 이 때, 중복되는 핵심 단어 각각에 대한 출현률의 평균값을 산출하여 해당 평균값을 해당 핵심 단어의 가중치로 결정하게 된다(S16). 여기서, 핵심 단어가 전체 거절 결정된 다수의 특허 문서 중 어느 하나에만 존재하는 경우 해당 핵심 단어의 출현률이 가중치로 결정된다.

[0034] 이와 같이, 다수의 특허 문서로부터 얻어진 핵심 단어들과, 각 핵심 단어에 대해 산출된 가중치가 함께 저장됨으로써, 핵심 단어 통합 DB가 생성된다(S17).

[0035] **[유사치 산출 단계 및 거절 기준값 산출 단계]**

[0036] 이하에서는, 도 2를 참조하여 본 발명에 따른 특허의 등록 가능성 예측 방법의 유사치 산출 단계와 거절 기준값 산출 단계에 대해 상세히 설명한다.

[0037] 먼저, 거절 기준값의 산출을 위한 특허 문서가 수집된다(S20). 이 때 거절 기준값의 산출을 위해 수집되는 특허 문서는 등록된 특허 문서와 거절 결정된 특허 문서 모두를 포함하게 된다. 그리고, 거절 결정된 특허 문서의 경우, 상술한 가중치 산출 단계에서 사용되지 않은 특허 문서가 수집된다.

[0038] 그런 다음, 도 1의 S11 단계 내지 S14 단계에서와 같이, 각 특허 문서로부터 핵심 단어를 추출한다(S24). 즉, 각 특허 문서별로 해당 특허 문서에 포함된 단어들을 추출하고, 추출된 단어들에 대해 단어별로 빈도수를 산출한다(S21). 그리고, 비고려 대상 단어가 제거되고(S22), 나머지 단어가 해당 특허 문서에서 나타나는 비율에 해당하는 출현률을 단어별로 산출한다(S23).

[0039] 그리고, 각 단어별로 산출된 출현률이 기 설정된 추출 기준값보다 큰 값을 갖는 단어를 해당 특허 문서의 핵심 단어로 추출한다(S24). 여기서, S21 단계 내지 S24 단계는 상술한 S11 단계 내지 S14 단계에 대응하는 바, 그 상세한 설명은 생략한다.

[0040] 상기와 같이, S20 단계에서 수집된 특허 문서에 대해 핵심 단어의 추출이 완료되면, S20 단계에서 추출된 핵심 단어와 핵심 단어 통합 DB에 저장되어 있는 핵심 단어 중 중복되는 단어를 중복 단어로 추출한다(S25).

[0041] 그리고, 해당 특허 문서에 대해 추출된 중복 단어에 대해 핵심 단어 통합 DB에 저장되어 있는 가중치와 해당 중복 단어에 대해 S23 단계에서 산출된 출현률을 곱한 값들의 합을 해당 특허 문서의 유사치로 산출한다(S26).

[0042] 여기서, 유사치의 산출은 [수학식 3]과 같이 표현될 수 있다.

[0043] [수학식 3]

[0044] $$S_y = \sum_{i=1}^{n} (V_i \times P_{iy}^*)$$

[0045] 여기서, S_y는 특허 문서 y의 유사치이고, V_i는 핵심 단어 통합 DB에 저장되어 있는 중복 단어의 가중치이고, P_{iy}는 S23 단계에서 산출된 중복 단어의 출현률이다.

[0046] 상기와 같은 과정을 통해 S20 단계에서 수집된 특허 문서 전체에 대해 S21 단계 내지 S27 단계를 수행하여, S20 단계에서 수집된 모든 특허 문서에 대한 유사치를 산출하게 된다.

[0047] 그런 다음, 중복 단어를 포함하는 특허 문서의 특허 등록 또는 거절 여부와, 해당 특허 문서의 유사치를 소정의 군집화 알고리즘에 적용하여 거절 기준값을 산출한다. 본 발명에서는 거절 기준값의 산출을 위한 군집화 알고리즘으로 k-means 알고리즘이 적용되는 것을 예로 한다(S28).

[0048] 여기서, k-means 알고리즘은 구현이 매우 쉬우며 수행이 매우 빠르다는 장점이 있다. 또한, k-means 알고리즘은 군집의 개수인 k를 직접 지정할 수 있다는 장점이 있다. 본 발명에서는 군집의 개수 k가 등록 군집과 거절 군집을 의미하는 2를 입력하며, 입력데이터는 각 특허 문서의 유사치와, 특허 등록, 거절 결정을 0, 1로 표현한 값이 된다. 출력데이터는 등록 군집과 거절 군집의 경계선이며, 이 경계선의 값이 거절 기준값으로 산출된다(S29).

[0049] **[등록 가능성 예측 단계]**

[0050] 이하에서는 상기와 같은 과정을 통해 산출된 거절 기준값을 이용하여, 예측 대상 특허 문서에 대한 특허 등록 가능성을 예측하는 방법을, 도 3을 참조하여 설명한다.

[0051] 먼저, 예측 대상 문서에 대해 상술한 가중치 산출 단계(도 1의 S11 단계 내지 S15 단계 참조)를 수행하여 예측 대상 특허 문서의 단어별 가중치를 산출한다(S30). 그런 다음, 예측 대상 특허 문서에 대해 상술한 유사치 산출 단계(도 2의 S25 단계 및 S26 단계 참조)를 수행하여 예측 대상 특허 문서에 대한 유사치를 산출한다(S31).

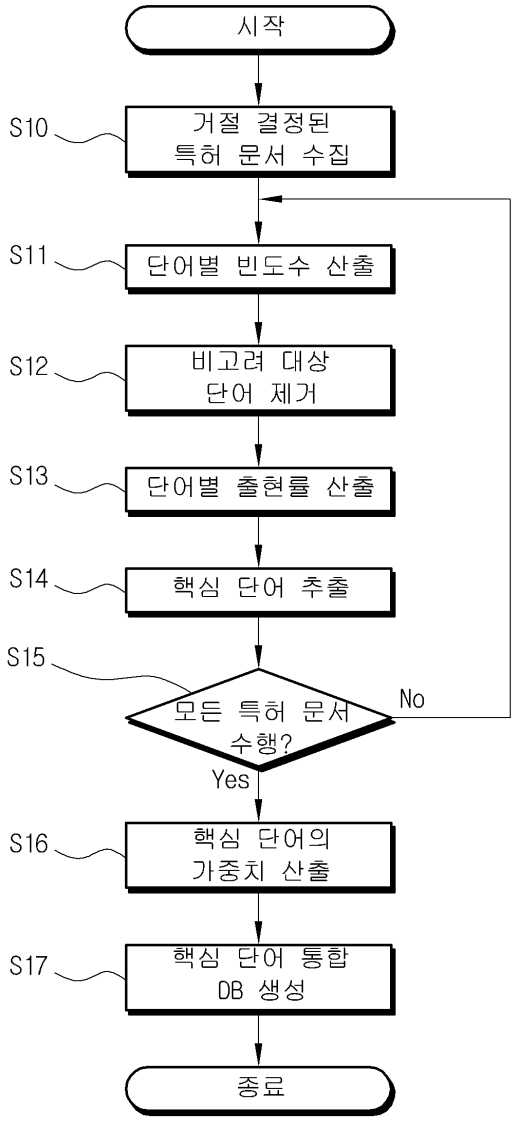
[0052] 그런 다음, 산출된 유사치가 거절 기준값보다 작은지 여부를 판단하고(S32), 산출된 유사치가 거절 기준값보다 작은 경우 예측 대상 특허 문서가 특허 등록 가능한 것으로 판단한다(S33).

[0053] 반면, 산출된 유사치가 거절 기준값보다 큰 것으로 판단되는 경우(S34) 예측 대상 특허 문서가 특허 거절 가능한 것으로 판단하게 되며(S35), 산출된 유사치가 거절 기준값과 동일한 것으로 판단되는 경우 예측 대상 특허 문서에 대한 등록 가능성 판단이 불가하다고 판단하게 된다(S36). 여기서, 산출된 유사치가 거절 기준값과 동일한 것으로 판단되는 경우에는 특허 등록 또는 특허 거절 어느 하나도 판단하도록 설정될 수 있음은 물론이다.

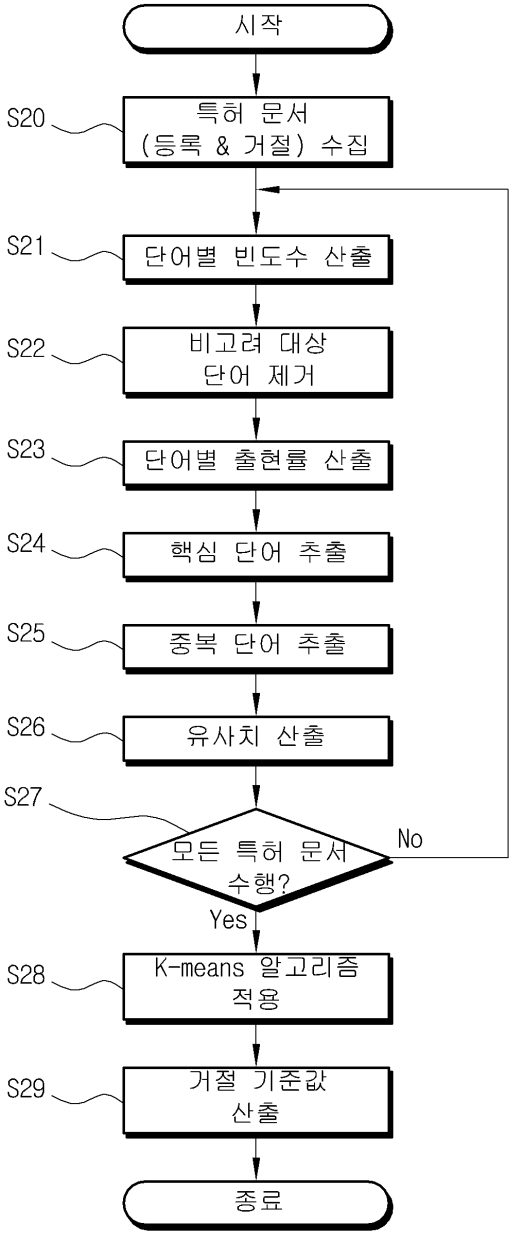
[0054] 이상에서 본 발명의 바람직한 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

도면

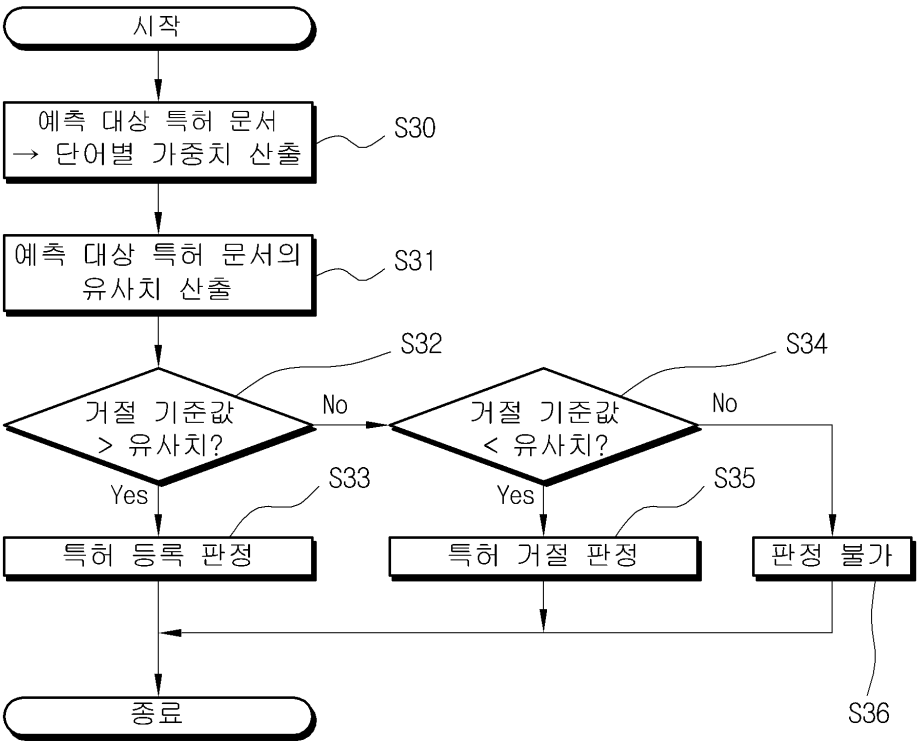
도면1



도면2


시작
S20
특허 문서
(등록 & 거절) 수집
S21
단어별 빈도수 산출
S22
비고려 대상
단어 제거
S23
단어별 출현률 산출
S24
핵심 단어 추출
S25
중복 단어 추출
S26
유사치 산출
S27
모든 특허 문서
수행?
No
Yes
S28
K-means 알고리즘
적용
S29
거절 기준값
산출
종료

도면3


시작
예측 대상 특허 문서
→ 단어별 가중치 산출
S30
예측 대상 특허 문서의
유사치 산출
S31
S32
거절 기준값
> 유사치?
No
S34
거절 기준값
< 유사치?
No
Yes
S33
Yes
S35
특허 등록 판정
특허 거절 판정
판정 불가
종료
S36