



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0088893  
(43) 공개일자 2010년08월11일

- |   |   |
|---|---|
| <p>(51) Int. Cl.<br/>G06F 17/21 (2006.01) G06F 17/27 (2006.01)<br/>G06F 17/30 (2006.01)</p> <p>(21) 출원번호 10-2009-0008032<br/>(22) 출원일자 2009년02월02일<br/>심사청구일자 2009년02월02일</p> | <p>(71) 출원인<br/>엘지전자 주식회사<br/>서울특별시 영등포구 여의도동 20번지</p> <p>(72) 발명자<br/>차완규<br/>서울특별시 영등포구 여의도동 LG트윈타워 8층 (서관) 특허실<br/>김정중<br/>서울특별시 영등포구 여의도동 LG트윈타워 8층 (서관) 특허실<br/>(뒷면에 계속)</p> <p>(74) 대리인<br/>허용특</p> |
|---|---|

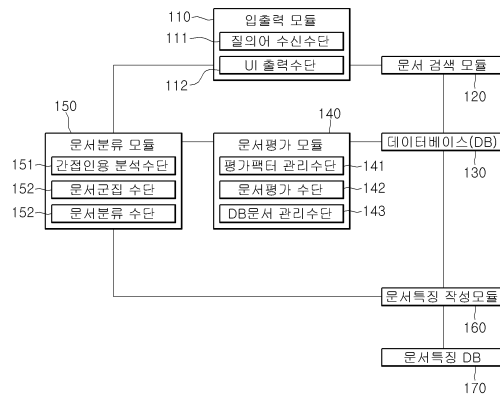
전체 청구항 수 : 총 8 항

**(54) 문서 분석 시스템**

**(57) 요약**

본 실시예의 문서 분석 시스템은 특허문서들이 저장되는 데이터베이스; 상기 특허문서들 간의 간접인용 관계를 추출하고, 추출된 간접인용 관계를 이용하여 분석대상의 특허문서들에 대해서 분류를 수행하는 문서분류 모듈; 상기 특허문서가 갖는 특성정보를 이용하여, 상기 특허문서에 대한 평가 수행 및 그 평가결과로서 평가치를 연산하는 문서평가 모듈; 및 상기의 특성정보를 기준으로 분석대상의 특허문서들에 대한 평가정보를 사용자에게 제공하는 UI출력 수단을 포함한다.

대표도 - 도1



(72) 발명자

**최성호**

서울특별시 영등포구 여의도동 LG트윈타워 8층 (서관) 특허실

**안한준**

서울특별시 영등포구 여의도동 LG트윈타워 8층 (서관) 특허실

**정미경**

서울특별시 영등포구 여의도동 LG트윈타워 8층 (서관) 특허실

## 특허청구의 범위

### 청구항 1

특허문서들이 저장되는 데이터베이스;

상기 특허문서들 간의 간접인용 관계를 추출하고, 추출된 간접인용 관계를 이용하여 분석대상의 특허문서들에 대해서 분류를 수행하는 문서분류 모듈;

상기 특허문서가 갖는 특성정보를 이용하여, 상기 특허문서에 대한 평가 수행 및 그 평가결과로서 평가치를 연산하는 문서평가 모듈; 및

상기의 특성정보를 기준으로 분석대상의 특허문서들에 대한 평가정보를 사용자에게 제공하는 UI출력 수단;을 포함하는 문서 분석 시스템.

### 청구항 2

제 1 항에 있어서,

상기 특성정보는 상기 특허문서에 기록된 사항으로부터 도출되는 대내적 특성과, 상기 특허문서가 인용하고 있는 피인용 특허문서에 기록된 사항을 고려함으로써 도출되는 대외적 특성을 포함하는 것을 특징으로 하는 문서 분석 시스템.

### 청구항 3

제 2 항에 있어서,

상기 특성정보는 상기 특허문서들간의 인용횟수 또는 피인용 횟수에 대한 정보를 이용하는 인용 인덱스를 더 포함하는 것을 특징으로 하는 문서 분석 시스템.

### 청구항 4

제 2 항에 있어서,

상기 문서평가 모듈은 특허문서의 평가결과를 기설정된 자사특허와 이외의 타사특허로 분류하고,

상기 UI출력 수단은 상기 평가정보로서, 상기 특성정보를 기준으로 자사특허에 대한 제 1 평가정보와 타사특허에 대한 제 2 평가정보를 UI로 제공하는 것을 특징으로 하는 문서 분석 시스템.

### 청구항 5

제 2 항에 있어서,

상기 문서평가 모듈은 상기 특허문서가 갖는 출원일 또는 공개일을 이용하여 상기 특허문서들에 대해 시계열적인 평가를 수행하고,

상기 UI출력 수단은 상기 평가정보로서, 시계열적으로 배열된 특성정보를 UI로 제공하는 것을 특징으로 하는 문서 분석 시스템.

### 청구항 6

제 4 항 또는 제 5 항에 있어서,

상기 UI출력 수단은 사용자가 상기 특성정보를 구성하는 항목을 수정할 수 있는 편집UI를 사용자에게 제공하는 것을 특징으로 하는 문서 분석 시스템.

### 청구항 7

제 2 항에 있어서,

상기 문서평가 모듈은 특허문서의 평가결과를 상기 특허문서에 기록된 출원인 또는 기술분류를 기준으로 분류하고,

상기 UI출력 수단은 상기 평가정보로서, 상기 특성정보를 이용한 출원인별 또는 기술분류별 평가정보를 그래프 또는 도표를 사용자에게 제공하는 것을 특징으로 하는 문서 분석 시스템.

**청구항 8**

제 2 항에 있어서,

상기 UI출력 수단은 제공된 평가정보를 이용하여 사용자가 상기 특허문서를 열람할 수 있거나 상기 특허문서의 요지리스트를 확인할 수 있는 UI를 제공하는 것을 특징으로 하는 문서 분석 시스템.

**명세서**

**발명의 상세한 설명**

**기술분야**

[0001] 본 발명은 문서들 간의 간접 인용관계를 이용하여 복수의 특허문서들을 군집화 및 자동분류가 수행되도록 하고, 이러한 분류가 수행된 문서들에 대해서 분석 및 평가를 수행하는 시스템에 대해서 개시한다.

**배경기술**

[0002] 특허출원인이 특허를 받으려고 하는 경우에는, 소정의 요건을 충족시키는 서류를 작성하고, 특허청에 제출할 필요가 있다. 특허청에 제출된 특허출원 서류는 소정의 시간경과 또는 요건이 충족된 이후에 공개되는데, 이러한 문서들을 특허문서라고 할 수 있다.

[0003] 일반적으로, 특허를 출원하려고 하는 자등은 선행기술의 존재여부를 확인하기 위하여 이러한 특허문서들을 검색/서치하는 과정을 거치게되는데, 대부분의 특허문서 검색은 키워드를 입력한 형태로 이루어지고 있다.

[0004] 즉, 근래의 기술 발달에 수반하여 특허출원의 건수가 방대해지고 있으며, 이에 수반하여 특허문서의 양도 방대해지고 있다. 이에 따라, 중복 연구를 방지하거나 권리침해의 여부를 확인하거나 특허출원 전의 선행기술 조사를 하거나 타사의 기술 개발 동향을 파악하거나 연구 개발 향상 등을 위하여 특허문서를 조사하는 작업이 용이하지 않은 실정이다.

[0005] 이러한 특허문서를 서치하거나 조사하기 위한 종래의 검색 시스템에서는, 키워드를 잘못 선택하면, 불필요한 정보가 방대히 포함되는 경우가 발생하곤 한다. 그리고, 이와 같은 경우에는 조사 그 자체의 시간도 방대해지는 문제점이 있다.

**발명의 내용**

**해결하고자하는 과제**

[0006] 본 발명의 실시에는 복수의 특허문서 간의 참조 또는 인용 관계를 도출하거나 직접 인용 관계가 아니더라도 간접 인용 관계를 도출함으로써, 특허문서 간의 보다 효율적인 분류 및 군집화를 수행하고, 이러한 문서 분류 및 군집화의 결과를 사용자에게 보다 효과적으로 제공할 수 있는 문서 분석 시스템을 제공하고자 한다.

**과제 해결수단**

[0007] 본 실시예의 문서 분석 시스템은 특허문서들이 저장되는 데이터베이스; 상기 특허문서들 간의 간접인용 관계를 도출하고, 도출된 간접인용 관계를 이용하여 분석대상의 특허문서들에 대해서 분류를 수행하는 문서분류 모듈; 상기 특허문서가 갖는 특성정보를 이용하여, 상기 특허문서에 대한 평가 수행 및 그 평가결과로서 평가치를 연산하는 문서평가 모듈; 및 상기의 특성정보를 기준으로 분석대상의 특허문서들에 대한 평가정보를 사용자에게 제공하는 UI출력 수단;을 포함한다.

[0008] 그리고, 상기 특성정보는 상기 특허문서에 기록된 사항으로부터 도출되는 대내적 특성과, 상기 특허문서가 인용하고 있는 피인용 특허문서에 기록된 사항을 고려함으로써 도출되는 대외적 특성을 포함한다.

[0009] 그리고, 상기 특성정보는 상기 특허문서들간의 인용횟수 또는 피인용 횟수에 대한 정보를 이용하는 인용 인덱스를 더 포함한다.

[0010] 그리고, 상기 문서평가 모듈은 특허문서의 평가결과를 기설정된 자사특허와 이외의 타사특허로 분류하고, 상기 UI출력 수단은 상기 평가정보로서, 상기 특성정보를 기준으로 자사특허에 대한 제 1 평가정보와 타사특허에 대한 제 2 평가정보를 UI로 제공한다.

**효 과**

[0011] 제안되는 바와 같은 실시예에 의해서, 복수의 특허문서 간의 참조 또는 인용 관계를 도출하거나 직접 인용 관계가 아니더라도 간접 인용 관계를 도출함으로써, 특허문서 간의 보다 효율적인 분류를 수행할 수 있는 장점이 있다.

[0012] 그리고, 효율적인 문서의 분류 및 군집화의 결과를 다양한 UI를 통하여 사용자에게 정보를 제공함으로써, 사용자가 특허문서의 분석을 용이하게 수행할 수 있도록 하는 장점이 있다.

**발명의 실시를 위한 구체적인 내용**

[0013] 이하에서는, 본 실시예에 대하여 첨부되는 도면을 참조하여 상세하게 살펴보도록 한다. 다만, 본 실시예가 개시하는 사항으로부터 본 실시예가 갖는 발명의 사상의 범위가 정해질 수 있을 것이며, 본 실시예가 갖는 발명의 사상은 제안되는 실시예에 대하여 구성요소의 추가, 삭제, 변경등의 실시변형을 포함한다고 할 것이다.

[0014] 그리고, 이하의 설명에서, 단어 '포함하는'은 열거된 것과 다른 구성요소들 또는 단계들의 존재를 배제하지 않는다.

[0015] 도 1은 본 실시예에 따른 문서 분석 시스템의 구성을 보여주는 일례이다.

[0016] 도 1에 도시된 바와 예와 같은 시스템은 특허문서가 저장되는 데이터베이스(130)와, 상기 데이터베이스(130)에 저장된 특허문서 또는 네트워크를 통하여 접속가능한 다른 특허문서들에 대해서 기설정된 평가팩터(사용자에 의해 변경가능함)를 이용해서 평가치를 부여하는 문서평가 모듈(140)과, 사용자에 의해 지정된 또는 데이터베이스에 격납된 특허문서들에 대한 직접 및 간접 인용관계를 도출하여 특허문서들의 분류 및 군집화가 이루어지도록 하는 문서분류 모듈(150)을 포함한다.

[0017] 또한, 이러한 문서 분석 시스템은 특허문서들 간의 간접 인용관계를 이용하여 특허문서들에 분류를 수행하고, 수행된 분류의 결과(분류된 특허문서들내에서 대표화 문서)를 이용해서 미분류의 특허문서들에 대해서 군집화를 수행할 수 있는 것으로서, 서버 장치나 컴퓨터 등에 의하여 실현될 수 있으며, 입출력 모듈(110), 문서검색 모듈(120), 문서특징 작성모듈(160) 및 문서특징 DB(170)를 더 포함할 수 있다.

[0018] 도 2는 상기 문서 분석 시스템은 특허 검색(S101), 검색결과를 이용한 특허문서의 분석(S102), 분석대상의 특허 문서들에 대한 문서 분류(S103), 문서 분류를 이용한 문서 군집의 결과의 UI를 제공(S104)할 수 있으며, 각각의 스텝에 대해서는 문서 분석 시스템의 각 구성을 이용해서 보다 상세히 설명하여 보기로 한다.

[0019] 먼저, 문서 분석 시스템을 이용한 특허문서의 검색(S101) 동작에 대해서 설명하여 보기로 한다.

[0020] 입출력 모듈(110)의 질의어 수신수단(111)는 사용자가 문서 검색 또는 분석등의 행위를 수행하기 위하여 키보드나 마우스등을 이용해서 입력한 질의어를 수신하는 것으로서, 사용자가 입력하는 질의어는 상기 데이터베이스(130)에 저장되어 있는(또는 네트워크 연결이 가능한) 특허문서에 기록된 키워드가 될 수 있다. 그리고, 상기 키워드는 문자 이외에 상기 특허문서를 구성하는 출원번호, 공개번호등의 숫자도 포함한다.

[0021] 그리고, 입출력 모듈(110)의 UI(User Interface) 출력수단(112)은, 상기 문서검색 모듈(120) 또는 문서분류 모듈(150) 또는 문서평가 모듈(140)에 의하여 연산 내지는 추출되는 정보를 사용자측에 제공하는 역할을 수행하며, 후술되는 다양한 UI를 제공하는 장치로 기술되어 있지만, 실시예에 따라 당연히 평가 시스템의 다른 구성요소 내에 마련되는 것도 가능하다.

[0022] 또한, 실시예의 데이터베이스(130)에는 특허문서 데이터들이 저장되며, 특허문서 데이터군은 전자화되는 특허출원 또는 특허에 관계되는 명세서의 문서 데이터를 격납하도록 구성되어 있는 데이터베이스이다. 이 특허문서 데이터는, 문자 코드에 의하여 명세서의 내용을 기술한 텍스트 데이터를 포함하는 데이터이다. 플레인 텍스트 데이터의 다른 곳, 예를 들면, SGML(Standard Generalized Markup Language), HTML(HyperText Markup Language), XML(eXtensible Markup Language)등의 범용 태그 언어에 의한 기술을 포함하는 문서 데이터도 가능하다. 그리고, 텍스트 데이터의 추출이 가능하다면, PDF(Portable Document Format)이나 범용의 워드 프로세서(word processor)의 문서 포맷(format), RTF(Rich TextFormat) 포맷등의 다른 포맷도 가능하다.

- [0023] 특허문서 데이터베이스(130)는, 특허문서 평가 시스템의 외부에 마련되어 있는 것도 가능하며, 그 경우에는, 네트워크를 이용하여 특허문서 평가 시스템이 데이터베이스에 접속하고, 특허문서의 문서 데이터를 취득할 것이다.
- [0024] 문서검색 모듈(120)은 사용자가 입력한 질의어를 바탕으로 상기 데이터베이스(130)에 저장되어 있는 특허문서들 중에서 호출대상의 특허문서들을 검색한다. 상기 문서검색 모듈(120)에 의한 특허문서의 검색에 있어서는, 상기 문서특징 작성모듈(160) 및 문서특징 DB(170)가 이용될 수 있다.
- [0025] 문서특징 작성모듈(160)은 상기 데이터베이스(130)에 저장되어 있는 문서들로부터 텍스트를 취득하여 각 키워드별 빈도수에 대한 인덱스 정보를 문서특징 DB(170)에 제공할 수 있다. 그리고, 상기 문서검색 모듈(120)은 질의어 수신수단(111)에 의하여 소정의 질의어가 수신되는 경우에 상기 문서특징 DB(170)에 저장된 각 문서의 인덱스 파일을 이용하여 질의어가 포함된 문서들을 검색할 수 있다.
- [0026] 상기 문서검색 모듈(120)에 의해 검색된 결과의 문서는 UI 출력수단(112)을 통하여 도 4에 도시된 바와 같은 UI가 사용자측에 제공될 수 있다.
- [0027] 상기 문서특징 작성모듈(160)은 질의어 수신수단(111)을 통하여 소정의 질의어가 수신되는 경우 또는 웹 로봇에 의하여 상기 데이터베이스(130)에 신규의 문서가 격납되는 경우에 해당 문서들에 대한 인덱스 파일을 작성하고, 이를 이용하여 각 문서에 대한 특징 벡터를 결정할 수 있다. 이에 대한 설명을 위하여 도 9를 참조하여 본다.
- [0028] 도 9는 각 문서의 속성 정보를 나타낸 도면이고, 도 9에 도시된 문서들의 속성정보는 문서특징 작성모듈(160)에 의해 인덱스 파일의 형태로 작성될 수 있으며, 작성된 인덱스 파일은 상기 문서특징 DB(170)에 저장된다.
- [0029] 그리고, 문서특징 DB(170)에 저장된 인덱스 파일을 이용하여 상기 문서특징 작성모듈(160)은 각 문서의 특징 벡터를 결정할 수 있으며, 상기 특징 벡터 역시 문서특징 DB(170)에 저장될 수 있다.
- [0030] 도 9에는 각 문서마다 키워드(A,B,C,D,M,I,K,O,P,Q,Z)별 발생빈도에 대한 정보가 도시되어 있으며, 예를 들면, 제 1 문서에는 키워드 A(여기서, A는 명사, 고유명사, 복합명사등의 단어를 의미하는 것이며, 알파벳 A를 의미하는 것이 아님)가 35번, 키워드 B가 19번, 키워드 C가 15번, 키워드 D가 13번이 포함되어 있음을 나타낸다.
- [0031] 그리고, 각 문서에 포함되어 있는 키워드별 발생빈도 테이블은 도 9에 도시된 바와 같이 가장 높은 빈도수를 갖는 키워드로부터 낮은 빈도수를 갖는 키워드로 순차적으로 배열되도록 작성될 수 있다.
- [0032] 예컨대, 문서 1에서는 키워드 A가 4.5%, 키워드 B가 2.4%, 키워드 C가 1.9%, 키워드 D가 1.7%가 포함되어 있음을 나타내기 위하여, 상기 문서 1에 대한 인덱스 파일은 (A,B,C,D) → (4.5%,2.4%,1.9%,1.7%)의 의미를 포함하도록 작성될 수 있다.
- [0033] 이렇게 다양한 방법에 의하여 각 문서들의 인덱스 파일이 작성되고, 작성된 인덱스 파일을 이용해서는 각 문서의 특징 벡터를 추출하는 것이 가능해진다.
- [0034] 상세히, 상기 문서특징 작성모듈(160)은 각 문서에서 키워드별 발생빈도수에 근거한 테이블을 작성하고, 이를 이용하여 각 문서의 특징 벡터를 함께 작성한다.
- [0035] 여기서, 상기 문서특징 작성모듈(160)에 의해 결정되는 특징 벡터는 각 문서에 대하여 키워드의 평가치를 요소로 하며, 예를 들어 각 문서에 포함된 키워드의 총 수가 n개인 경우, 각 문서의 특징 벡터는 n차원 공간의 벡터로서 다음 식(1)과 같이 표현될 수 있다.
- [0036] 특징 벡터 = (키워드 A의 평가치 w1, 키워드 B의 평가치 w2, . . . . . 단어 n의 평가치 wn) --- (1)
- [0037] 평가치의 연산에는, 예를 들어 문헌(Salton, G:Automatic Text Processing : The transformation, Analysis, and Retrieval of Information by Computer, Addison-Wesley)에 개시되어 있는 tf·idf법을 이용할 수 있다. tf·idf법에 따르면, 제 1 문서에 대응하는 n차원의 특징 벡터 중, 제 1 문서에 포함되는 키워드에 대응하는 요소에 대해서는, 평가치로서 0이외의 값이 산출되고, 제 1 문서에 포함되지 않은 키워드(빈도가 0인 단어)에 대응하는 요소에 대해서는 평가치로서 0이 산출된다.
- [0038] 이와 같은 견지에서, 특징 벡터의 한 요소로서 키워드의 평가치는 각각의 문서에 나타나는 각 키워드의 빈도율이 될 수 있다. 예컨대, 상기 문서검색 모듈(120)에 의해서, 제 1 문서로부터 키워드 A, 키워드 B 및 키워드 C는 유사어로 군집화될 수 있으며, 군집된 유사어는 별도의 유사어 DB에 저장될 수 있다.
- [0039] 즉, 상기 문서검색 모듈(120)에 의해서 소정의 키워드 A와 키워드 B가 군집화되고, 군집화된 키워드 A와 키워드

B는 유사어 DB에 저장된다.

- [0040] 그리고, 상기 문서검색 모듈(120)은 추출되는 키워드에 키워드 A와 키워드 B중 어느 하나가 포함되어 있는 경우에는, 나머지 키워드가 포함된 유사문서에 대해서도 검색을 수행한다.
- [0041] 추출된 키워드에 한정된 검색이 수행되는 것이 아니라 특허문서들의 속성에 근거하여 유사한 문서들의 검색이 수행될 수 있는 것이다.
- [0042] 상기 질의어 수신수단(111)을 통해 수신되는 질의어 중에 키워드A가 포함되어 있는 경우에는, 유사문서 검색시에 키워드 A와 함께 키워드 B 및 키워드 C가 포함되어 있는 문서의 검색이 수행될 수 있다.
- [0043] 한편, 문서 분석 시스템은 특허검색 과정이나 문서 분류 및 군집화 과정에서 기설정된 평가 팩터를 이용한 문서에 대해 평가치를 부여함으로써, 각 문서들에 대해 중요도 또는 트렌드를 파악할 수 있도록 한다.
- [0044] 상기 문서 분석 시스템에 의한 특허문서의 평가에 대해서 살펴보기로 한다.
- [0045] 상기 데이터베이스(130)에 격납되어 있는 또는 네트워크를 통하여 접속이 가능한 특허문서들에 대해서 상기 문서평가 모듈(140)은 기 설정된 평가 팩터에 따라 평가치를 부여한다.
- [0046] 즉, 문서평가 모듈(140)은 상기 데이터베이스(130)에 격납된 특허문서 또는 네트워크 연결가능한 특허문서들에 대해서 상기 특허문서가 갖는 속성정보를 이용해서 특허문서를 평가하며, 또한 그 평가의 결과가 사용자에게 보여지도록 그 결과를 상기 UI 출력수단(112)으로 제공한다. 그리고, UI 출력수단(112)은 특허문서의 검색결과 리스트와 함께 검색된 대상의 특허문서들에 대한 평가치에 대한 정보를 사용자측에 제공할 수 있으며, 상기 검색결과 리스트와는 별도의 팝업 또는 OSD로 특허문서들에 대한 평가치 정보를 제공할 수도 있다.
- [0047] 상기 문서평가 모듈(140)은 상기 데이터베이스(130)에 격납된 특허문서들 또는 네트워크 연결이 가능한 특허문서들에 대해서 설정된 평가항목을 이용해서 평가항목 테이블을 작성하며, 이러한 특허문서의 평가 작업은 상기 데이터베이스(130)에 신규의 특허문서가 저장되는 경우마다 수행될 수 있다.
- [0048] 다만, 상기 문서평가 모듈(140)에 의한 특허문서의 평가 작업은 사용자의 문서검색 요청이 있고 검색되는 문서가 존재하는 경우에 수행되는 것도 가능하며, 이하의 설명에서는 이러한 평가 작업이 수행되는 순서의 한정없이 서술한다.
- [0049] 상기 문서평가 모듈(140)에는 특허문서가 갖는 특성을 평가팩터로서 관리하는 평가팩터 관리수단(141)과, 상기 평가팩터를 이용하여 상기 데이터베이스(130)에 저장된 특허문서에 대해서 평가를 수행하는 문서평가 수단(142)과, 상기 문서평가 수단(142)에 의한 문서 평가결과인 평가치가 특허문서 각각에 대응되도록 하는 DB문서 관리수단(143)이 포함될 수 있다.
- [0050] 상기 평가팩터 관리수단(141)은 상기 데이터베이스(130)에 격납된 특허문서의 대내적 특성과 대외적 특성에 대한 항목을 관리하며, 이러한 특성들은 사용자에 의하여 편집될 수 있다.
- [0051] 즉, 상기 평가팩터 관리수단(141)에 의하여 특허문서의 대내적 특성 및 대외적 특성에 대한 평가팩터들의 구조는 도 3에 도시된다. 도 3은 특허문서의 평가팩터의 구조를 나타내는 도면이다.
- [0052] 도 3에 나타나 있듯이, 상기 평가팩터 관리수단(141)에 의하여 기술되는 특허에 관한 속성의 테이블이 국가별로 복수개 연결될 수 있으며, 하나하나의 테이블에는 특허문서 내에 기록되어 있는 사항으로부터 도출되는 대내적 특성과, 특허문서가 인용하고 있는 피인용 문서의 특성을 고려함으로써 도출될 수 있는 대외적 특성을 포함한다.
- [0053] 특허문서에 기록되어 있는 사항으로부터 도출될 수 있는 대내적 특성이라 함은, 특허문서의 기재사항에 대한 텍스트마이닝 작업을 통하여 추출될 수 있는 키워드 또는 해당 특허문서의 정보를 가리킨다.
- [0054] 예를 들어, 특허문서에 기록되어 있는 등록일자로부터 현재일자까지의 기간이 연산된 유지기간은 해당 특허문서 내에 기재된 사항으로부터 도출가능한 것이므로, 특허문서의 대내적 특성이 될 수 있다.
- [0055] 그리고, 특허문서에 기재된 출원일자로부터 현재일자까지의 기간이 연산된 경과정보, 특허문서의 독립항의 수, 특정 독립항에 대한 텍스트 마이닝결과 도출되는 키워드의 개수에 따라 결정될 수 있는 청구항 길이, '제 1 항에 있어서' 또는 'according to claim 1'과 같이 특정의 문구가 들어가기 때문에 종속항으로 식별가능한 종속항들의 개수 역시 특허문서의 대내적 특성이 될 수 있다.

- [0056] 또한, 특허문서에 기재되어 있는 발명자들의 수 역시 특허문서의 대내적 특성이 될 수 있다.
- [0057] 다만, 제 1 특허문서에서 발명자로 기록된 A가 출원한 특허의 개수에 대해서는, 해당 발명자 A가 발명자로 기록되어 있는 다른 특허문서들을 검색하여야 하기 때문에, 특허문서의 대외적 특성이라 할 수 있다.
- [0058] 그리고, 해당 특허문서에서 인용하고 있는 다른 특허문서가 있을 경우에는, 인용하고 있는 특허문헌의 개수, 인용/피인용의 기간등은 특허문서의 대외적 특성이 된다.
- [0059] 특허문서를 점수화하기 위한 평가치 연산을 위해서는, 특허문서에 대한 평가팩터가 정의되어야 하고, 정의된 평가팩터들에 대한 각각의 가중치(weighting value)를 연산함으로써, 종국적으로 해당 특허에 대한 평가치가 연산될 수 있다.
- [0060] 이러한 견지에서, 도 3에 도시된 바와 같은 일례의 테이블을 이용하여, 상기 평가팩터 관리수단(141)은 상기 데이터베이스(130)에 격납된 특허문서들 각각에 대한 평가팩터 항목들을 작성한다. 도 2에는 대내적 특성과 대외적 특성들이 랜덤하게 배열되어 있으나, 특허문서 내에서 추출되는 정보로부터 획득가능한 대내적 특성에 대한 평가치와, 해당 특허문서와 다른 특허문서(검색결과내에서의 다른 특허문서와 데이터베이스에 저장된 동일 기술분야의 다른 특허문서도 가능)간의 관계에서 산출되는 평가치를 별도의 항목으로서 구별하여 둘 수도 있다.
- [0061] 각각의 특허문서들로부터 독출되는 특성들의 값을 도 3에 도시된 바와 같은 테이블에 기록한 다음에는, 상기 문서평가 수단(142)에 의하여 특허문서의 평가치가 연산된다.
- [0062] 예를 들면, 각각의 평가팩터들에 대해서는 미리 결정된 가중치가 부여될 수 있으며, 이 경우 특허문서로부터 추출되는 대내적 특성 및 대외적 특성의 값에 상기 가중치가 연산됨으로써, 평가팩터 각각의 점수의 합이 해당 특허문서의 평가치가 될 수 있다.
- [0063] 이렇게 연산된 특허문서에 대한 평가치들은 DB문서 관리수단(143)에 의하여 별도로 관리될 수 있으며, 특허문서 검색결과와 정보가 사용자에게 보여질 때 검색된 결과에 포함되는 특허문서마다 연산된 평가치가 함께 보여지도록 한다.
- [0064] 따라서, 상기 입출력 모듈(110)의 UI 출력수단(112)은 상기 평가팩터 관리수단(141)에 의해 관리되는 평가팩터의 항목 내지는 테이블을 사용자측에 제공하고, 사용자가 추가, 편집 및 삭제하는 평가팩터의 내용은 상기 평가팩터 관리수단(141)에 의해 저장관리된다.
- [0065] 상기와 같은 문서평가 모듈(140)에 의해 각각의 특허문서들은 평가치를 부여받을 수 있으며, 이렇게 부여된 평가치는 해당 특허문서가 검색의 결과로서 사용자측에 보여질 때 그 결과 리스트와 함께, 도 4와 같이, 보여질 수 있다.
- [0066] 참고로, 도 4에는 사용자의 컴퓨터 또는 서버에 제공되는 문서검색 결과의 일 리스트가 예시되어 있다. 예를 들어, 사용자가 입력한 질의어에 대해서 상기 문서검색 모듈(120)이 상기 데이터베이스(130)에 저장된 특허문서의 검색결과가 7건이 독출된 경우에는, 검색대상의 특허문서들에 대한 서지적인 정보(예를 들면, 특허번호, 상태, 출원일, 특허일, 발명의 명칭, IPC)의 표시와 함께 각각의 특허문서들에 대한 평가치가 함께 표시된다.
- [0067] 또한, 상기 문서평가 수단(142)은 검색된 결과의 특허문서들중에서 사용자가 가장 가치가 높은 특허와 그렇지 않은 특허를 빨리 구별할 수 있도록 특허문서에 대한 평가치를 상기 UI 출력수단(112)으로 제공한다. 그리고, 특허문서 각각에 대한 평가치와 함께, 검색된 결과의 특허문서들의 평균 평가치를 연산하여, 이러한 평균 평가치 역시 상기 UI 출력수단(112)으로 제공할 수 있다.
- [0068] 검색된 결과의 특허문서들에 대한 평균 평가치가 함께 보여질 경우에는, 사용자는 검색결과와 특허문서 각각에 대한 우열을 용이하게 결정할 수 있을 것이며, 본 실시예에 따라 사용자는 그 평가 가치가 높은 특허문서들을 먼저 확인하여 봄으로써, 검색 효율을 향상시킬 수 있다.
- [0069] 그리고, 상기 문서평가 수단(142)은 검색결과와 특허문서들이 포함되는 기술분야에서의 평균 평가치를 연산할 수 있으며, 상기 UI 출력수단(112)은 검색결과와 특허문서 각각의 평가치와 함께 해당 특허문서들이 속하는 기술분야에서의 평균 평가치를 함께 제공할 수 있다.
- [0070] 이 경우, 검색된 결과의 특허문서들이 속하는 기술분야의 공통 여부는 국제분류인 IPC분류에 의하여 수행되거나, 일본 특허청에서 분류하고 있는 F-term에 의하여 판단될 수 있다. 그리고, 서로 다른 기술분야로 분류되는 특허문서들이 검색결과로서 출력되어야 할 경우에는, 검색결과에서 다수 비율을 차지하는 특허문서들



이 속하는 기술분야에 대한 평가치의 평균값이 제공될 수 있다.

- [0071] 따라서, 사용자는 검색된 결과의 특허문서들 각각에 부여된 평가치를 해당 기술분야의 특허문서들의 평균 평가치와 비교함으로써, 검색된 결과의 특허문서들이 어느 정도 중요도를 갖는 특허문서들인지를 쉽게 파악할 수 있게 된다.
- [0072] 한편, 검색결과와 리스트를 사용자가 선택적으로 다운로드할 수 있는 기능이 제공될 수 있으며, 검색결과 리스트의 다운로드시에는 상기 문서평가 모듈(140)에 의하여 수행되는 평가치에 대한 정보도 함께 사용자측 컴퓨터 또는 서버에 제공될 수 있도록 한다.
- [0073] 또한, 도 4에 도시된 바와 같은 검색결과와 UI에서, 사용자가 각각의 특허문서에 부여된 평가치의 세부 항목을 확인하기 위하여 특정의 평가치(Weighting Value)를 클릭하는 경우에는, 상기 평가치를 구성하는 평가팩터들과 각 평가팩터에 대해 해당 특허문서에 부여된 점수를 상세하게 확인할 수 있도록 하는 별도의 UI를 제공할 수 있다.
- [0074] 또한, 검색결과와 리스트를 포함하는 도 4에 도시된 바와 같은 UI에서, 사용자가 특정의 특허문서를 선택하는 경우에는, 해당 특허문서에 대한 요약내용을 보여주는 별도의 창(UI)이 생성될 수 있다. 즉, 도 5에 도시된 바와 같이 특허문서 분석 UI가 사용자측에 제공될 수 있으며, 이러한 UI에도 해당 특허문서에 대한 평가치 정보가 제공된다.
- [0075] 예를 들면, 선택된 특허문서에 대한 발명의 명칭, 대표도면 및 요약등에 대한 사항과 함께 해당 특허문서에 적용된 평가팩터의 항목과, 해당 항목마다의 점수의 정보가 제공될 수 있다. 그리고, 앞서 설명한 바와 같이, 검색된 결과의 특허문서들 또는 해당 특허와 동일한 기술분야의 특허문서들의 평균 평가팩터값들이 함께 제공될 수 있다.
- [0076] 그리고, 사용자는 자신의 서버 또는 컴퓨터등을 조작하여, 표시된 평가팩터 항목에 대해서 수정 및 편집할 수 있으며, 또한 부여된 점수에 대해서도 별도로 편집할 수 있다. 이를 위해서, 상기 문서평가 모듈(140)의 평가팩터 관리수단(141)과 DB문서 관리수단(143)들은 사용자에 의해 변경된 평가팩터의 항목 및 점수에 대응되도록 해당 특허문서의 정보를 변경한다.
- [0077] 그 다음, 데이터베이스에 격납된 문서들에 대해서 평가 작업을 수행하거나 사용자가 요청한 별도의 문서들에 대해서 한정적으로 평가 작업을 수행하는 등의 과정을 거친 다음에는, 본 실시예의 문서분류 모듈(150)에 의한 문서 분류(SI03) 작업이 수행되며, 이러한 작업의 결과는 도 11 내지 도 15와 같은 다양한 UI의 정보가 UI 출력수단(112)에 의하여 사용자에게 제공될 수 있다.
- [0078] 본 실시예의 문서검색 모듈(120), 문서평가 모듈(140) 및 문서분류 모듈(150)은 별개로 동작하기 보다는, 문서의 검색, 분류 및 군집이 보다 효과적으로 이루어지도록 하기 위하여 이들이 함께 기설정된 알고리즘에 따라 복합적으로 동작하는 것임을 알 수 있다.
- [0079] 이하에서는, 사용자가 입력한 질의어에 대해서 상기 문서검색 모듈(120) 및 문서특징 작성모듈(160)에 의하여 소정의 특허문서들이 검색되고, 그 검색의 결과가 도 3과 같은 리스트로 나타나는 경우에, 문서검색 결과의 특허문서들에 대해서 그 기술적 해결과제(종래기술의 문제점) 또는 해결방법(과제 해결 수단)이 유사한 문서들끼리 분류하는 동작에 대해서 설명하여 보기로 한다.
- [0080] 즉, 본 실시예에 따라 특허문서들 간의 간접 인용관계를 이용함으로써 문서들을 분류할 수 있고, 이러한 인용관계를 갖는 특허문서들은 그 기술적 해결과제 또는 해결방법을 공통으로 하는 경향이 있으므로, 데이터베이스(130)에 격납된 특허문서들 모두를 대상으로 분류를 수행하기 보다는 사용자가 입력한 질의어에 대한 문서검색(유사검색 포함) 결과의 특허문서들에 대해서 분류를 수행하는 것이 더욱 유리하다.
- [0081] 이러한 점에서, 상기 문서분류 모듈(150)의 동작은, 문서검색의 결과로서 소정의 유사범위에 속하는 특허문서들을 예로 들어 설명하여 보기로 한다. 다만, 문서평가 모듈(140)은 특허문서들의 분류 후 문서의 군집화에서도 동작하지만, 이러한 문서 분류와 문서 군집화 이전의 문서 검색 단계에서도 도 4와 같이 부여받은 평가치 정보가 제공될 수 있는 것이다.
- [0082] 한편, 상기 UI 출력수단(112)은 사용자가 검색결과와 특허문서 리스트중에서 일부 특허문서 또는 검색결과와 전부의 특허문서에 대해서 분류 및 군집화를 수행하는 것을 안내하기 위한 태그(34, 도 4 참조)를 제공할 수 있다.

- [0083] 이러한 문서 분류 및 군집화를 요청하는 키가 입력되면, 상기 문서분류 모듈(150)은 선택된 특허들에 대한 간접 인용관계 도출 및 이를 이용한 문서 분류를 수행하는 것이다. 예를 들어, 제 1 특허문서가 제 2 특허문서에 인용되고, 상기 제 2 특허문서가 제 3 특허문서에서 인용되는 경우에 상기 제 1 특허문서와 제 3 특허문서는 간접 인용관계에 있는 것이므로, 상기 문서분류 모듈(150)은 제 2 특허문서와 함께 제 1 및 제 3 특허문서도 동일한 카테고리 내로 분류한다.
- [0084] 여기서, 본 실시예에 따른 인용관계 즉, 간접 인용관계에 대해서 살펴보기로 한다. 인용관계라고 함은, 특허문서 내에서 종래기술의 문제점을 서술하기 위하여 기재되어 있는 다른 특허문서의 참조 문서 번호(특허출원번호, 특허공개번호, 등록번호등)가 기재되어 있다면, 인용되는 특허문서와 인용하는 특허문서의 관계가 형성될 수 있다.
- [0085] 또한, 특허문서 내에서 언급 내지는 기재하고 있는 특허문서만이 인용되는 문서라고 한정될 필요는 없으며, 해당 특허문서에 대한 심사 또는 의의신청 또는 무효심판등에서 있어서 선행기술/인용발명등으로 참조되는 문서 역시 인용관계에 있다고 할 수 있다. 따라서, 해당 특허문서 내에 다른 특허문서의 서지적 정보등에 대하여 기재되어 있는 경우 뿐만 아니라, 간접적으로 심사관 또는 다른 제 3 자등에 의하여 심사도중에 이용될 수 있는 다른 특허문서 역시 인용 관계에 있다고 할 수 있다.
- [0086] 이러한 인용관계를 확대시키기 위하여, 상기 데이터베이스(130)에는 특허문서들 각각의 인용여부에 정보가 격납되는 인용 및 참조문서 저장부가 구비될 수 있으며, 이 경우 특허문서에 기재된 사항으로부터 인용관계 여부를 도출하기 위한 독출수단 이외에 특허청 등이 제공하는 자료들로부터 심사중 또는 등록후의 절차에서 사용된 문헌들로부터 인용관계를 도출하기 위한 독출수단이 구비될 수 있다.
- [0087] 예를 들면, A 특허문서 내에서 다른 B 특허문서의 공개번호등이 기재되어 있다면, A특허문서와 B특허문서간의 직접 인용관계를 독출할 수 있는 것이다. 그리고, A특허문서에 대한 심사중에 그 인용발명으로서 C특허문서가 심사관에 의하여 제시되었다면, C특허문서 역시 A특허문서와 인용관계에 있다고 할 수 있다.
- [0088] 또한, 청구항에 기재된 사항중에는 제 1 그룹의 특허문서와 제 2 그룹의 특허문서가 있으나, 제 1 그룹은 사용자가 문서를 검색한 다음 검색결과와 특허문서들에 대해서 간접인용관계를 이용한 문서분류를 수행함으로써 형성되는 문서그룹이라 할 수 있다. 그리고, 제 2 그룹은 사용자에게 의하여 지정된 특허문서들 또는 데이터베이스(130)에 격납된 다른 특허문서들을 가리키는 것으로서, 실시예의 문서분류 모듈(150)에 의한 문서의 분류가 수행되지 않은 특허문서들의 집단을 나타낸다고 볼 수 있다.
- [0089] 따라서, 사용자가 검색결과와 특허문서들에 대해서 문서 분류를 수행할 것을 요청하는 경우에, 상기 문서분류 모듈(150)에 의하여 문서 분류가 수행된 이후에는 상기의 제 1 그룹과 같이 적어도 하나 이상의 그룹이 생성될 수 있다. 그리고, 사용자가 문서 분류 이후에 다른 특허문서들(이하, "제 2 그룹 문서"라 함)에 대해서도 문서 분류 내지는 군집화를 수행하고자 하는 경우에는, 미분류 또는 미군집화된 제 2 그룹에 속하는 문서들은 제 1 그룹의 특성(대표문서 또는 대표벡터)을 이용해서 상기 제 1 그룹에 속하는 분류들로 분류 및 군집화될 수 있다.
- [0090] 다만, 이해를 돕기 위하여, 제 1 그룹에 속하는 문서들을 간접 인용관계를 이용한 문서 분류가 수행된 것으로 정의하고, 제 2 그룹에 속하는 문서들을 아직 분류 내지는 군집화가 수행되지 않은 것으로 설명하였으나, 제 2 그룹에 속하는 문서들 역시 분류 내지는 군집화가 이미 수행된 것이어도 제 1 그룹의 분류 기준에 따라 다시 분류 및 군집화를 수행하면 되는 것으로, 반드시 이러한 정의에 한정될 필요는 없다.
- [0091] 그리고, 본 발명의 상세한 설명에 있어서, 분류와 군집이라는 용어에 대해서도 혼용될 수 있으나, 이는 상기 문서분류 모듈(150) 또는 문서검색 모듈(120)등의 동작과 관련하여서 해석하면 충분하니, 이 점 유의할 필요가 있다.
- [0092] 한편, 이러한 인용관계의 도출 이외에 본 실시예에서는 간접 인용관계를 이용하여 특허문서들을 분류할 수 있으며, 이에 대해서는 첨부되는 도 6 내지 8을 참조하여 보기로 한다.
- [0093] 도 6은 본 실시예에 따른 문서분류 모듈의 문서군집 수단에 대한 일례를 보여주는 도면이고, 도 7은 본 실시예에 따른 문서분류 모듈에 의한 간접 인용관계를 도출하는 구성을 설명하기 위한 도면이고, 도 8은 본 실시예에 따른 문서분류 모듈에 의하여 유사문서를 분류된 그룹 내로 군집화시키는 구성을 설명하기 위한 도면이다.
- [0094] 먼저, 도 7을 참조하여, 본 실시예의 문서분류 모듈(150)에 의해서 간접 인용관계를 도출하는 구성에 대해서 살펴보기로 한다.

- [0095] 사용자는 검색된 결과의 문서 또는 직접 지정한 문서들에 대해서 상기 문서분류 모듈(150)에 의한 간접 인용관계에 대한 정보를 획득하는 것이 가능하다. 도 7에 도시되어 있는 바와 같이, 사용자는 분류하고자 하는 문서들에 대해서 기간(기간 A ~ 기간 B) 설정이 가능하며, 이 경우 분류대상의 특허문서들중에서도 설정된 기간 내의 문서들에 대한 분류가 수행된다.
- [0096] 즉, 설정된 기간 내에 속하는 특허문서들 간에 직접인용관계(문서 내에 서지적 정보를 기록함으로써 형성되는 인용관계 또는 심사관등에 의하여 참조됨으로써 형성되는 인용관계)가 성립되지 않는 경우라도, 인용하는 특허문서 또는 인용되는 특허문서 간의 연관관계가 존재한다면 이러한 특허문서들에 대해서는 간접 인용관계로서 동일한 카테고리 내에 분류될 수 있다.
- [0097] 예를 들면, 문서 분석 및 분류를 위하여 사용자에게 의하여 설정된 기간이 기간A ~ 기간B이고, 이러한 기간 내에 속하는 특허문서들(Base Patent, Patent 5, Patent 6, Patent 7, Patent 8, Patent9)간에 서로 직접인용관계에 있지 않고, 설정된 기간 외의 제 1 특허문서(Patent 1)가 제 5 특허문서(Patent 5) 및 Base Patent에 인용된다면, 제 5 특허문서(Patent 5)와 베이스 특허문서(Base Patent)는 상호 간에 간접 인용관계가 성립한다.
- [0098] 다른 예를 들면, 제 3 특허문서(Patent 3)가 기간 내의 제 7 특허문서(Patent 7)과 베이스 특허문서(Base Patent)를 직접 인용하고 있다면, 상기 제 3 특허문서(Patent 3)과 제 7 특허문서(Patent 7)은 상호간에 간접 인용관계가 성립되므로, 본 실시예에 따라 동일한 카테고리로 분류된다.
- [0099] 이러한 방법을 통해서, 도 7의 경우에는, 베이스 특허문서(Base Patent)가 제 5 특허문서 내지 제 9 특허문서(Patent 5 ~ 9)에 모두 간접 인용관계를 형성하므로, 대표화 문서 또는 베이스 특허문서가 될 수 있다.
- [0100] 그리고, 사용자는 이러한 방법에 의하여 분류되는 특허문서들의 카테고리 단위들에 대해서, 그 내용을 쉽게 파악하기 위하여 분류명을 직접 작성할 수 있다. 예를 들면, 분류된 카테고리의 특허문서들이 '소음저감'이라는 해결과제(또는 해당과제)를 공통으로 하고 있는 경우에, 그 카테고리명을 '소음저감(ex.해당과제1)'로 기입할 경우에, 폴더명이 '소음저감'으로 관리될 수 있다.
- [0101] 그리고, 사용자가 분류된 카테고리들에 대해서 그 이름을 해결과제1,2,3과 해결수단 1,2,3등으로 명칭한 경우에, 각각의 해결과제와 해결수단에 대응되는 카테고리들에 대한 표시로서 이미지가 표시될 수 있다. 이 경우, 그래프 내에 표시된 이미지는 각 카테고리 내에 포함되는 특허문서들의 개수에 따라 다른 색상 또는 크기로 표현될 수 있으며, 또한 각 카테고리 내에 포함된 특허문서들의 평가치의 합(또는 평균 평가치)의 고저에 따라 다른 색상 또는 다른 크기로 표현될 수도 있다.
- [0102] 사용자에게 제공되는 문서 분류의 결과 또는 문서 군집의 결과로서 제공되는 UI에서 사용자가 특정의 카테고리(해결수단1, 해결수단2, 해결수단3, 해당과제1, 해당과제2, 해당과제3)를 선택하는 경우에는, 앞서 설명한 대표 특허문서(베이스 특허문서) 또는 문서평가 모듈에 의하여 부여된 평가치가 가장 높은 특허문서에 대한 정보를 사용자에게 제공한다.
- [0103] 이러한 과정에 의하여, 사용자는 검색된 결과의 문서들에 대해서 문서 분류를 수행할 수 있다. 나아가, 간접 인용관계를 이용한 문서 분류를 수행한 다음에는, 분류되지 않았거나 다른 간접 인용관계로 분류되어 있는 특허문서들 - 제 2 그룹에 속하는 것이라 할 수 있음 -을 분류 및 군집화할 수 있다.
- [0104] 여기서의 문서 군집화 과정은 상기 문서특징 작성모듈(160)에 의한 문서간의 유사도 판단이 이용될 수 있으며, 상기 문서분류 모듈(150)은 이미 분류된 제 1 그룹의 특허문서들을 중심으로 제 2 그룹의 특허문서들을 분류 및 군집화한다. 그리고, 상기 문서분류 모듈(150)의 문서군집 수단(152)은 제 1 그룹의 제 1 카테고리에 속하는 특허문서(제 1 카테고리의 대표문서가 될 수 있음)와 제 2 그룹의 특허문서간의 유사도를 판단함으로써, 상기 제 2 그룹에 속하는 특허문서를 제 1 그룹의 어느 카테고리로 분류할 지 여부를 결정한다.
- [0105] 상기 문서군집 수단(152)은 분류된 카테고리 내의 대표 문서를 이용하거나 해당 카테고리에 속하는 복수의 문서들을 이용해서 군집화에 필요한 대표 벡터를 산출하는 대표벡터 산출부(1521)를 포함할 수 있다.
- [0106] 그리고, 상기 문서군집 수단(152)은 특허문서를 구성하는 필드별(또는 식별항목별)로 유사한 문서들을 군집하기 위한 필드별 군집화부(1522)를 포함할 수 있다.
- [0107] 상기 대표벡터 산출부(1521)는 이미 형성된 카테고리 내에서의 대표 문서(베이스 특허문서 또는 평가치를 이용함으로써 선출되는 특허문서) 또는 동일 카테고리 내에 속하는 문서들로부터 키워드별 발생빈도에 근거하여 상기 문서특징 작성모듈(160)에 의해 작성된 인덱스 파일을 이용한다. 예를 들어, 상기 대표벡터 산출부(1521)는 각 문서에서 나타나는 키워드들 중에서 높은 빈도수를 갖는 대표 키워드들을 추출할 수 있으며, 각 문서의 인덱

스 파일로부터 발생빈도가 높은 순서로 상위 몇개의 키워드들을 선정할 수 있다.

- [0108] 도 9에 도시된 바와 같은 키워드 분포도에 대한 이러한 선정작업에 의하여, 도 10에 도시된 바와 같은 각 문서의 특징 벡터들이 형성될 수 있다.
- [0109] 그리고, 대표벡터 산출부(1521)는 발생빈도가 높은 순서로 선택된 키워드들에 대하여 각 문서에서 차지하는 백분율을 계산할 수 있으며, 예컨대 Documents 1에서 키워드 A는 4.5%, 키워드 B가 2.4%, 키워드 E가 1.9%, 키워드 D가 1.7%로 각 키워드별 발생빈도의 백분율을 계산할 수 있다.
- [0110] 이러한 과정으로, 해당 카테고리 내의 문서들 또는 대표 문서(이하에서는, '카테고리 문서'라고 함)에 대하여 키워드별 발생빈도의 백분율을 계산한다.
- [0111] 도 9 및 도 10을 참조하면, 카테고리 문서들에 대하여 이러한 과정이 수행된 다음에는, 카테고리 문서 전체에 대하여 각 키워드가 차지하는 백분율을 합산하고, 합산된 키워드의 백분율이 높은 순서로 소정 개수의 특징 키워드를 대표 키워드로 선출할 수 있다.
- [0112] 예를 들어, 도 9에 도시된 각각의 키워드들 중에서 10개의 카테고리 문서 전체에서 각 키워드가 차지하는 백분율을 합산한 값이 키워드 B, 키워드 A, 키워드 E, 키워드 D, 키워드 O, 키워드 C, 키워드 K 순서로 높은 값을 갖는 경우에는, 선택된 문서들을 군집화하기 위한 대표 키워드로 키워드 B, 키워드 A, 키워드 E 및 키워드 D를 선택할 수 있다. 그리고, 선택된 대표 키워드를 대표 벡터의 성분으로 하여, 각각의 문서들에 대한 특징 벡터를 산정한다. 즉, 선택된 대표 키워드를 높은 확률분포 순서로 배열하여 이들을 대표 벡터의 성분으로 선정한다. 선택된 키워드 B, 키워드 A, 키워드 E 및 키워드 D를 기준으로 각 문서의 특징 벡터를 작성하는 과정이 수행되는데, 각 문서의 인덱스 파일중에서 상위 4개의 키워드들에 대하여 이러한 과정이 수행된다. 다만, 대표 벡터의 성분을 구성하는 대표 키워드로 4개가 선택되고, 각 문서에서 빈도수가 높은 4개의 키워드를 비교하여 각 문서의 특징벡터를 작성하는 것으로 설명하고 있으나, 이는 예시적인 사항일뿐 시스템의 관리자에 의하여 얼마든지 변경될 수 있다.
- [0113] 선택된 대표 키워드가 각 문서에 포함되어 있을 경우에는, 벡터 성분을 '1'로 설정하고, 포함되어 있지 않을 경우에는 '0'으로 설정할 수 있다. 다만, 이 역시 벡터 성분으로 1과 0 대신에 각 키워드에 대한 가중치를 부여한 값으로 벡터 성분을 작성할 수도 있다.
- [0114] 이렇게 작성된 각 문서의 특징 벡터는, 도 9에 도시된 바와 같이, 대표 키워드가 포함된 경우에는 '1', 포함되어 있지 않을 경우에는 '0'으로 하여 각 문서의 특징 벡터가 완성된다.
- [0115] 이러한 과정에 의하여, Document 1의 특징 벡터는 (1,1,1,1)이 되고, Document 2의 특징 벡터는 (1,1,0,1)이 된다. 각 특징 벡터의 성분이 1 또는 0으로 작성하였으나, 각 대표 키워드가 차지하는 발생빈도수에 따라 각 벡터 성분을 다른 값으로도 부여할 수 있다.
- [0116] 복수의 카테고리 문서를 이용하는 경우에, 이러한 각 문서의 특징벡터들을 이용하여 대표 벡터(또는 중심 벡터)를 선정하는 과정이 수행되는데, 여기서는 각 특징 벡터들 중에서 크기가 가장 큰 벡터를 군집화하기 위한 대표 벡터로 선정될 수 있다.
- [0117] 이러한 경우에, 도 9에 도시된 각각의 특징 벡터 중에서 Document 1의 특징 벡터(1,1,1,1)가 대표 벡터로 될 수 있으며, 선정된 대표 벡터를 이용함으로써 미분류된 제 2 그룹의 특허문서들을 군집화시킬 수 있다.
- [0118] 카테고리 문서로부터 도출되는 대표 벡터를 이용함으로써, 특정 카테고리나 소정의 유사도를 갖는 특허문서가 제 2 그룹에 포함되어 있는지 여부를 확인할 수 있으며, 이러한 유사도는 전술한 바와 같은 특징 벡터 또는 대표 벡터를 제 2 그룹의 특허문서들에 대해서도 수행함으로써 판단될 수 있다.
- [0119] 즉, 제 1 그룹의 소정 카테고리에 속하는 카테고리 문서와 제 2 그룹에 미분류된 문서와의 유사도는 각각의 특징 벡터 또는 대표 벡터에 의한 내적을 이용하여 산출될 수 있으며, 예컨대 카테고리 문서의 대표 벡터와 제 2 그룹의 특허문서에 대한 특징 벡터와의 내적을 통해서 내적된 연산의 값이 기 설정된 범위내에 속하는 경우에는 상기 대표 벡터와 함께 군집화될 수 있다. 즉, 상기 대표 벡터가 속하는 카테고리 내로 분류 및 군집화될 수 있다.
- [0120] 그리고, 상기 문서군집 수단(152)은 대표 벡터를 A라 하고, 유사도의 비교대상이 되는 문서의 특징 벡터를 B라고 하였을 때, 벡터A와 벡터B의 내적값을  $|A| \cdot |B| \cdot \cos \theta$ 로 나눈 값이 '1'로부터 얼마나 떨어져있는지에 따라 벡터 A에 해당하는 문서와 벡터 B에 해당하는 문서간의 유사도를 판단한다.

- [0121] 그러나, 상기 대표 벡터와의 제 2 그룹 문서의 특징 벡터간의 내적된 연산의 값이 기준값을 벗어나는 경우에는, 상기 대표 벡터와 함께 군집화되지 아니하며, 다른 군집을 위한 문서로 사용된다.
- [0122] 카테고리를 대표하는 대표 벡터와 제 2 그룹 문서의 특징 벡터간의 이러한 유사도 산출 및 판정에 따라, 도 7과 같이, 제 2 그룹에 속하는 제 20문서(P20)는 제 1 그룹의 A분류로 군집화될 수 있으며, 제 2 그룹의 제 21문서(P21)는 제 1 그룹의 B분류로 군집화될 수 있다.
- [0123] 다만, 전술한 실시예 외에 문서분류 모듈(150)에 의하여 문서 분류가 수행되면, 그 결과로서 상기 문서분류 모듈(150)은 카테고리를 대표하는 기술분류 코드(IPC 또는 F-term)를 선정할 수 있다. 이 경우, 상기 문서군집 수단(152)에 의한 제 2 그룹 문서들의 분류 및 군집은 전술한 유사도 판단 이외에 기술분류 코드를 이용한다.
- [0124] 예를 들면, 상기 문서군집 수단(152)은 간접 인용관계를 이용해서 문서를 분류한 결과인 각각의 카테고리들에 대하여 높은 빈도수를 기록하는 F-term들을 이용해서, 제 2 그룹 문서들이 갖는 F-term과의 유사도를 판단할 수 있다.
- [0125] F-term의 경우, 해결과제 또는 해결수단에 따라 분류된 것이기에, 문서의 벡터화를 이용한 유사도 판단과 함께 사용된다면, 보다 효과적인 문서 군집화를 수행할 수 있을 것이다.
- [0126] 그 다음, 본 실시예에 따라 특허문서의 분류 및 그 분류 결과를 이용한 군집화가 수행된 다음에는, 상기 문서분류 모듈(150) 및 UI 출력수단(112)에 의하여 도 11 내지 도 15와 같은 다양한 정보를 갖는 UI들이 사용자에게 제공될 수 있다.
- [0127] 도 11은 문서의 분류 및 군집으로부터 획득될 수 있는 정보에 대한 UI의 제 1 실시예이다.
- [0128] 본 실시예의 문서 분석 시스템에 의하여 특허문서의 분류가 수행되고, 분류된 결과를 이용하여 다른 특허문서들을 군집화시킨 다음에는, 사용자의 기간 설정 또는 출원인(또는 등록권자) 설정에 따라 도 11과 같은 특허문서 분석 UI가 사용자에게 제공될 수 있다.
- [0129] 예를 들면, 사용자가 자사의 설정을 "LGE(대표명화를 포함)"로 하고, 경쟁사를 "A사"로 설정하여 둔 경우에, 군집화된 결과 내에서 국가별 출원 건수와 해당 건들의 평가치등이 도표로 보여질 수 있다. 특히, 상기 문서평가 모듈(140)에 의하여 부여된 평가치가 포함될 수 있으며, 해당 항목에 포함되는 건들의 평가치 합을 보여주거나 해당 항목에 포함된 건들의 평균 평가치를 보여줄 수 있다.
- [0130] 그리고, 이러한 정보와 함께, 특허당 인용 인덱스(CPP), 기술영향력 인덱스(CII), 기술력 인덱스(TS), 영향력 측정 인덱스(TII), 기술진보 측정 인덱스(TCT) 및 기술자립도 인덱스(TI)등이 보여질 수 있다.
- [0131] 여기서, 특허당 인용 인덱스는 보유특허의 평균 피인용 횟수를 가리키고, 기업의 기술적 진보정도를 평가하기 위한 항목으로서 해당 특허문서의 인용된 횟수를 전체특허수로 나눈 값이 될 수 있다. 기술영향력 인덱스는, 예를 들어, 과거 5년동안 기업의 특허가 인용된 정보를 나타내어, 기업의 기술이 최근에 미치는 영향력 정보를 평가하기 위한 것으로서,  $CII = (\text{연도별피인용도} \times \text{연도별 특허수의 총합} / \text{전년도 특허수 총합})$ 으로 계산될 수 있다.
- [0132] 그리고, 기술력 인덱스는 기업의 기술영향력을 양적으로 평가하기 위한 항목으로서,  $(CII \times \text{특허건수})$ 로 계산될 수 있다. 영향력 측정 인덱스는 특정 기술분야에서 상위 10% 이상 인용되는 특허가 해당 기술분야의 전체인용횟수에서 차지하는 비율을 가리키는 항목으로서, 기업별로 특정 기술분야에 미치는 영향력을 평가하기 위하여,  $(\text{피인용상위 10\% 이상에 속하는 특허의 피인용횟수} / \text{전체피인용횟수})$ 로 계산될 수 있다.
- [0133] 그리고, 기술진보 측정 인덱스는 기업의 기술진보 속도에 대한 평가 항목으로서, 인용특허들과의 년도차 중 중간값에 해당하는 년도차의 평균을 나타내며,  $(\text{중간인용특허와의 연도차 총합} / \text{특허수})$ 로 계산될 수 있다. 기술자립도 인덱스는 자사 기술의 독립성을 평가하기 위한 항목으로서, 자사의 특허를 인용하는 정도를 획득하기 위하여  $(\text{자사특허 인용횟수} / \text{전체인용횟수})$ 로 계산될 수 있다.
- [0134] 이러한 다양한 종류의 인덱스는 문서의 분류 및 군집화가 수행된 다음에 상기 문서분류 모듈(150)에 의하여 연산될 수 있으며, 그러한 연산의 결과는 UI 출력수단(112)에 의하여 도 11 내지 도 15와 같은 도표 또는 그래프 등으로 나타낼 수 있다.
- [0135] 도 12는 문서의 분류 및 군집으로부터 획득될 수 있는 정보에 대한 UI의 제 2 실시예이다. 제 2 실시예의 경우는, 설정된 기간 내에서의 출원인 별 특허문서 건수가 도표로 도시되고, 해당 출원인은 사용자가 선택하여 둔

경우가 될 수 있다.

- [0136] 그리고, 각 시기에 해당되는 특허문서들의 평균 평가치가 W/F로 도시될 수 있으며, 사용자는 이러한 UI에 함께 표시되는 W/F 항목으로부터 그 기술개발의 변곡점이 될 수 있는 위치를 확인할 수 있다. 또한, 사용자가 평균 평가치인 W/F가 높은 시점을 선택하는 경우에는, 실시예의 문서분류 모듈(150) 및 UI 출력수단(112)은 해당 시점의 특허문서들의 정보를 별도의 UI로 제공하거나 해당 시점에서 평가치가 가장 높은 문서 또는 대표화 문서를 별도의 UI로 제공할 수 있다.
- [0137] 도 13은 문서의 분류 및 군집으로부터 획득될 수 있는 정보에 대한 UI의 제 3 실시예이다. 도 13에는 사용자가 설정한 시기 및 출원인에 대한 특허당 인용지수(CPP), 기술영향력 지수(CII), 특허당 인용지수(CPP) 및 기술영향력 지수(CII)등의 정보가 포함된 UI가 도시되어 있으며, 이러한 UI내에 출원인별 특허당 인용지수를 시기별로 표시한 그래프가 더 포함될 수 있다.
- [0138] 즉, 도 13 아래에 도시된 UI에는, 삼성전자와 샤프등의 출원인이 높은 인용지수를 갖는 것으로 예시되어 있다.
- [0139] 이외에, 기술분야별 특허활동 평가, 특허활동지수, 특허 포트폴리오 분석(HHI) 및 특허 다각화 지표에 대한 정보가 더 제공될 수 있으며, 기술분야별 특허활동 평가는 선택된 기간 내에 분야별 특허활동을 정량적으로 비교하는 것으로서, 기술분야별 출원(또는 공개)건을 비교하는 것에 의하여 가능하다.
- [0140] 그리고, 특허활동 지수는 특정 기술분야에서 차지하는 비율을 가리키는 것으로서, {(특정분야의 특허수/회사전체 특허수)/(회사전체 특허수/모든 분야 전체 특허수)}로 계산될 수 있다.
- [0141] 그리고, 특허 포트폴리오 분석 지수는 기업들이 시장에서 경쟁하는 형태를 파악하기 위한 항목으로서, 각 기업의 상위 IPC 분야별로 산출하고, 기업별로 과점하고 있는 기술분야와 경쟁하고 있는 기술분야를 산출할 수 있다. 예를 들면, 발명자당 출원 건수는 발명자 1인당 출원 건수의 상대평가(총 출원 건수/회사발명자수) 지수를 가리키고, 발명자당 청구항 수는 발명자 1인당 획득한 청구항 수의 상대 평가(총 보유청구항 수 / 회사발명자수) 지수를 가리키고, 유효특허 평균 잔존기간은 보유 특허의 평균 잔존 기간(유효특허의 잔존기간 총합/유효특허 전체 건수)의 지수를 가리킬 수 있다.
- [0142] 그리고, 공동 출원 비율은 공동 연구 활동의 활발한 정도를 평가하는 항목으로서, (공동출원 건수/전체 특허건수)로 계산될 수 있다.
- [0143] 도 14 및 도 15는 문서의 분류 및 군집으로부터 획득될 수 있는 정보에 대한 UI의 제 4 및 제 5 실시예이다.
- [0144] 도 14 및 도 15에는, 특정 기간 내의 회사별 인용횟수에 대한 그래프와, 높은 인용횟수를 갖는 특허문서에 대한 도표를 갖는 UI가 도시되어 있으며, 높은 인용횟수를 갖는 특허문서의 표시에 있어서는 상기 문서평가 모듈(140)에 의해 부여된 평가치가 함께 표시되도록 할 수 있다.
- [0145] 또한, 사용자가 인용횟수가 높은 순서대로 배열된 도표를 보고서, 특정 특허문서의 번호(출원번호, 등록번호 등)를 선택할 경우에는, 해당 특허문서에 대한 추가 정보 또는 해당 명세서를 사용자에게 제공할 수 있다.
- [0146] 전술한 바와 같은 본 실시예의 문서 분석 시스템에 의하여 제공되는 문서 분류 결과 또는 문서 군집화의 결과는 시스템의 설정에 따라 내용 저장 및 다른 사용자와의 공유도 가능한 것이며, 이러한 경우에 특히 특허개발을 유도하는 기업 또는 팀에서는 매우 유용할 것이다.

**도면의 간단한 설명**

- [0147] 도 1은 본 실시예에 따른 문서 분석 시스템을 설명하기 위한 도면.
- [0148] 도 2는 본 실시예에 따른 문서 분석 시스템의 동작 흐름을 설명하기 위한 도면.
- [0149] 도 3은 본 실시예에 따른 문서평가 모듈의 평가팩터 테이블의 일례.
- [0150] 도 4는 본 실시예에 따른 문서의 검색 및 평가 결과를 도시한 일례.
- [0151] 도 5는 본 실시예에 따라 문서의 정보가 보여지는 UI의 일례.
- [0152] 도 6은 본 실시예에 따른 문서군집 수단의 구성을 보여주는 도면.
- [0153] 도 7은 본 실시예에 따른 간접인용 관계를 설명하기 위한 도면.
- [0154] 도 8은 본 실시예에 따라 제 2 그룹 문서가 제 1 그룹의 카테고리로 분류 및 군집화되는 것을 설명하기 위한 도

면.

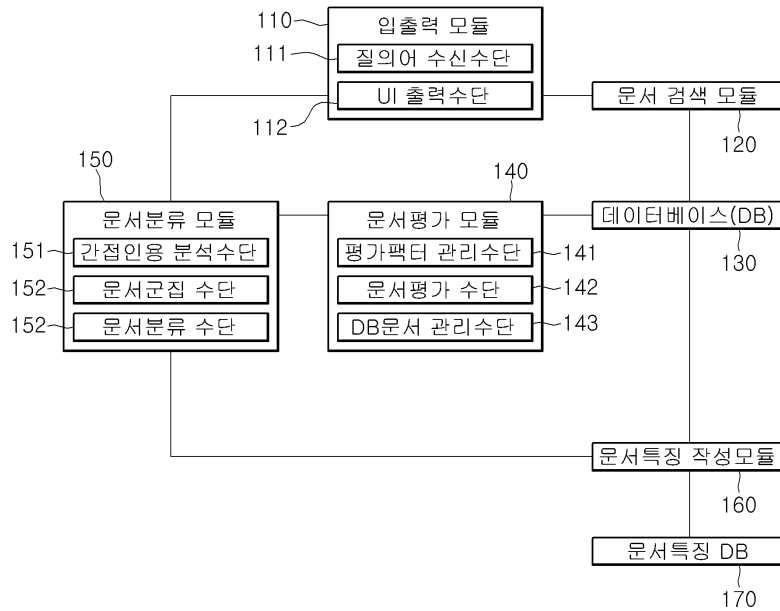
[0155] 도 9는 본 실시예에 따른 카테고리 문서 또는 제 2 그룹 문서의 속성 정보를 나타내는 일례.

[0156] 도 10은 본 실시예에 따른 카테고리 문서 또는 제 2 그룹 문서로부터 산출되는 특징 벡터를 나타내는 일례.

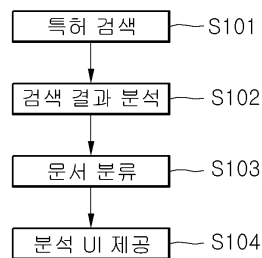
[0157] 도 11 내지 도 15는 본 실시예에 따른 문서 분류 및 군집의 결과로서 사용자에게 제공되는 다양한 종류의 UI.

도면

도면1



도면2







도면4

NO	원문	국가	특허번호 ▼	상태 ▲▼	출원일 ▲▼	특허일 ▲▼	발명의 명칭 ▲▼	IPC ▲▼	Weighting Value
1		KR	2008-0084358	A	2007/03/16	2008/09/19	자동차 에어컨 필터용 부직포의 제조 방법	B01D-039/02	77
2		KR	2008-0083766	A	2007/03/13	2008/09/19	사용자 체온 감지 기반 자동차용 온도 제어 장치	B60H-001/00	90
3		KR	2008-0083591	A	2007/03/11	2008/09/18	변환 회로	H03K-019/00	75
4		KR	2008-0082102	A	2007/03/07	2008/09/11	분리형 에어컨의 배관 연결부 개폐 장치	BF24F-013/20	80
5		KR	2008-0081713	A	2007/03/06	2008/09/10	점퍼 구조에서의 데이터 인식 방법	G06F-001/00	82
6		KR	2008-0079841	A	2007/03/28	2008/09/02	다기능 광고용 책자	B42D-001/00	76
7		KR	2008-0079395	A	2007/03/27	2008/09/01	소프트 키동수단을 구비한 모터	H02K-007/00	88

34  
문서분류

31

32

33

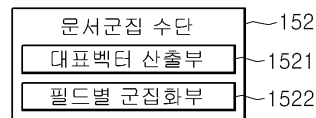
도면5

발명의 명칭	<input type="text" value="원문보기"/>	Factor	Value	평균
대표도면		유지기간		
		경과정보		
		⋮	⋮	⋮
요약				

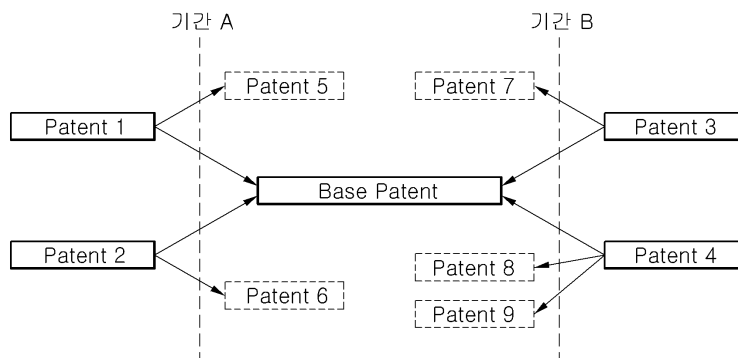
  

중요도	<input type="radio"/> S <input type="radio"/> A <input type="radio"/> B <input type="radio"/> C
제품 적용 가능성	<input type="radio"/> 상 <input type="radio"/> 중 <input type="radio"/> 하
기술분류	<input type="text" value="추가"/> <input type="text" value="삭제"/>

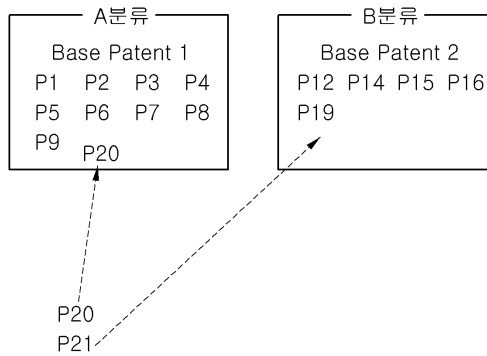
도면6



도면7



도면8



도면9

Document 1 (35,19,15,13,.....1) (A,B,C,D,.....,M)

Document 2 (24,22,20,15,.....,1) (B,C,A,D,.....,I)

Document 3 (27,15,10,8,.....,1) (B,A,C,K,.....,D)

⋮

Document 100 (33,21,15,15,.....,1) (O,P,Q,C,.....,Z)

도면10

Document 1 (1, 1, 1, 1)

Document 2 (1, 1, 0, 1)

Document 3 (1, 1, 1, 0)

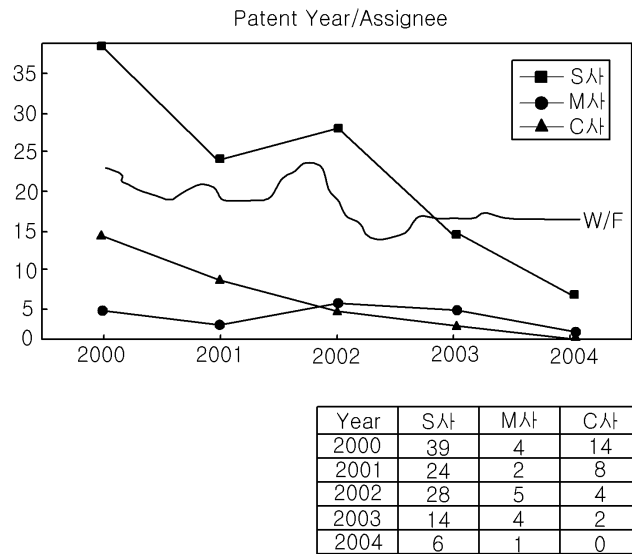
⋮

Document 10 (0, 1, 0, 0)

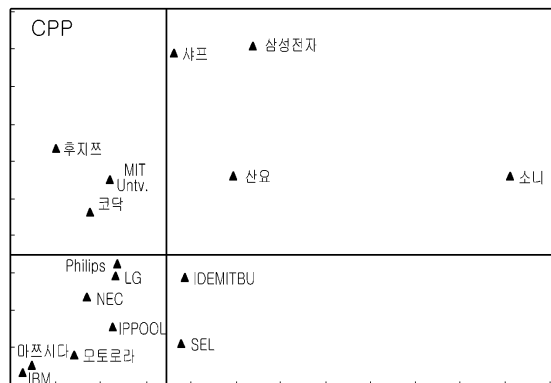
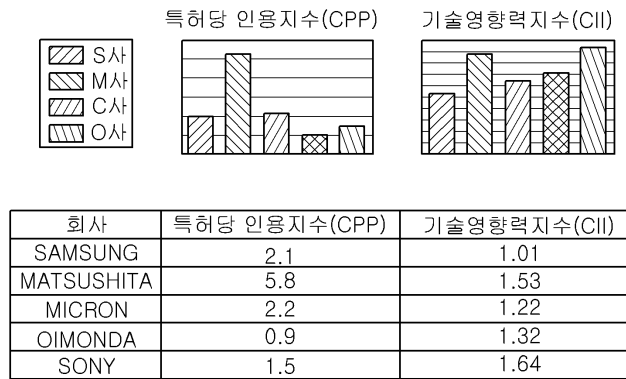
도면11

	LGE						경쟁사(A社)							
	Total		미국		.....	Index	Total		미국		.....	Index		
	件	WF	件	WF		CPP	TII	件	WF	件	WF		CPP	TII
기술분류1														
기술분류2														

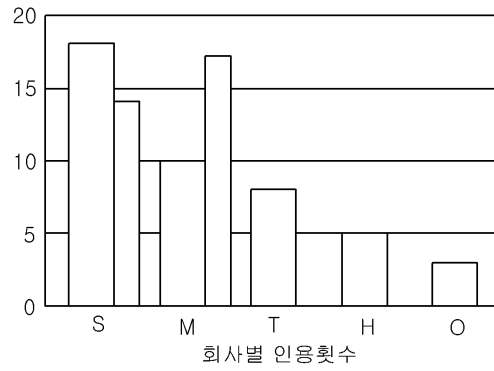
도면12



도면13

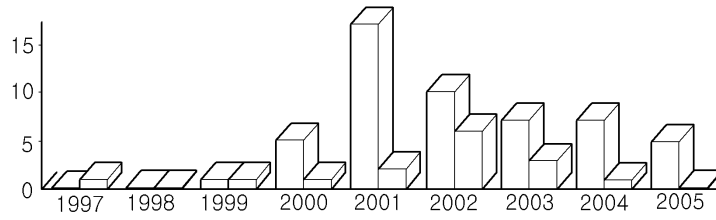


도면14



번호	피인용횟수	W/F	인용특허	회사
US 6487421	5	79	US 7451247	SONY
			US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
			US 7053215	HYNIX
US 6144794	5	85	US 7451247	SONY
			US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
			US 7053215	HYNIX
US 5478134	4	60	US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
			US 7053215	HYNIX
US 6614572	3	50	US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
US 6859742	2	40	US 6857845	TOSHIBA

도면15



	1997	1998	1999	2000	2001	2002	2003	2004	2005
A is referring B	0	0	5	5	17	10	7	7	5
B is referring A	1	0	1	1	2	6	3	1	0

번호	W/F	피인용횟수	인용특허	회사
US 6487421	79	5	US 7451247	SONY
			US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
			US 7053215	HYNIX
US 6144794	85	5	US 7451247	SONY
			US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
			US 7053215	HYNIX
US 5478134	60	4	US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
			US 7053215	HYNIX
US 6614572	50	3	US 6853215	SONY
			US 6933554	MATSUSHITA
			US 6857845	TOSHIBA
US 6859742	40	2	US 6857845	TOSHIBA