



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2019년05월07일  
(11) 등록번호 10-1975907  
(24) 등록일자 2019년04월30일

(51) 국제특허분류(Int. Cl.)  
G06F 16/00 (2019.01) G06Q 10/06 (2012.01)  
G06Q 10/10 (2012.01) G06Q 50/10 (2012.01)  
(52) CPC특허분류  
G06F 16/2465 (2019.01)  
G06Q 10/0637 (2013.01)  
(21) 출원번호 10-2018-0031342  
(22) 출원일자 2018년03월19일  
심사청구일자 2018년03월19일  
(56) 선행기술조사문헌  
KR101327546 B1\*  
KR1020170079648 A\*  
\*는 심사관에 의하여 인용된 문헌

(73) 특허권자  
인천대학교 산학협력단  
인천광역시 연수구 아카데미로 119 (송도동)  
(72) 발명자  
김관호  
인천광역시 연수구 컨벤시아대로42번길 77 (송도동, 더샵 엑스포)  
이레네 쇼스  
인천광역시 연수구 아카데미로 119 (송도동)  
(74) 대리인  
김효성

전체 청구항 수 : 총 10 항

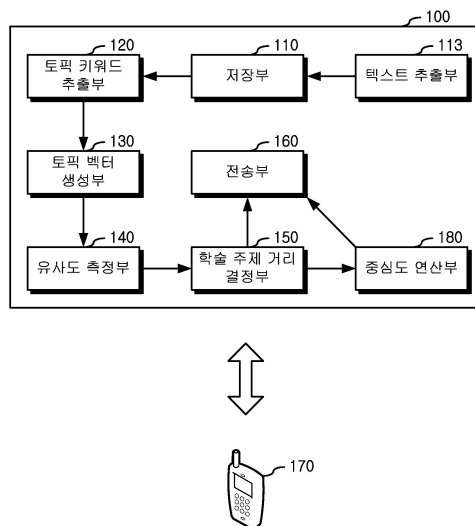
심사관 : 고재용

(54) 발명의 명칭 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치 및 이의 동작 방법

(57) 요약

본 발명에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치 및 이의 동작 방법은 제1 도시에 대한 제1 도시 토픽 벡터와 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 제1 도시와 다수의 도시들 각각 간의 유사도를 측정하고, 측정된 유사도에 기초하여, 제1 도시와 다수의 도시들 각각 간의 학술 주제 거리를 결정한 후 결정된 학술 주제 거리를 사용자 단말에 전송함으로써, 각 학회들이 포함된 지역들 간의 협업 가능성을 평가할 수 있는 정보를 자동으로 가공해서 사용자에게 제공하고자 한다.

대표도 - 도1



(52) CPC특허분류

G06Q 10/103 (2013.01)

G06Q 50/10 (2013.01)

G06F 2216/03 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1345263358

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 개인기초연구(교육부)

연구과제명 인간-공간 상호작용의 전체최적화를 위한 딥러닝 기반의 지능공간 의사결정 모델연구

기 여 율 1/2

주관기관 인천대학교

연구기간 2017.06.01 ~ 2018.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 C06370400100486835

부처명 중소벤처기업부

연구관리전문기관 중소기업기술정보진흥원

연구사업명 맞춤형 기술파트너 지원사업

연구과제명 딥러닝 기반 기업 연관성 분석 시스템 설계

기 여 율 1/2

주관기관 인천대학교 산학협력단

연구기간 2018.08.01 ~ 2019.04.30

## 명세서

### 청구범위

#### 청구항 1

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치에 있어서,

사전 설정된 다수의 토픽 키워드들 및 다수의 학회들 각각의 웹 페이지에 업로드된 다수의 학술 대회 정보들을 저장하는 저장부;

상기 저장부에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출하는 토픽 키워드 추출부;

상기 추출된 토픽 키워드의 중복 횟수에 기초하여 상기 제1 도시에 대한 제1 도시 토픽 벡터를 생성하는 토픽 벡터 생성부;

상기 제1 도시에 대한 제1 도시 토픽 벡터와 상기 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 유사도를 측정하는 유사도 측정부;

상기 측정된 유사도에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정하는 학술 주제 거리 결정부; 및

상기 결정된 학술 주제 거리를 사용자 단말에 전송하는 전송부

를 포함하고,

상기 학술 주제 거리 결정부는 상기 제1 도시와 상기 다수의 도시들 중 제2 도시 간의 유사도가 사전 설정된 기준 유사도 이상이면 상기 제1 도시와 상기 제2 도시 간의 학술 주제 거리를 1로 결정하며,

상기 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치는

상기 학술 주제 거리 결정부가 결정한 상기 제1 도시와 상기 다수의 도시들 각각 간의 상기 학술 주제 거리에 기초하여, 상기 제1 도시의 중심도를 연산하는 중심도 연산부

를 더 포함하고,

상기 전송부는 상기 결정된 학술 주제 거리 및 중심도를 상기 사용자 단말에 전송하는 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치.

#### 청구항 2

삭제

#### 청구항 3

삭제

#### 청구항 4

제1항에 있어서,

상기 중심도 연산부는 하기 수학식 1을 이용하여 상기 제1 도시의 제1 중심도를 연산하는

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치.

[수학식 1]

$$C_{deg}(x) = \frac{n_{edge}(x)}{T_C - 1}$$

$C_{deg}(x)$ 는 제1 도시  $x$ 의 제1 중심도이고,  $n_{edge}(x)$ 는 상기 다수의 도시들 중 제1 도시  $x$ 와 학술 주제 거리가 1로 결정된 도시들의 개수이고,  $T_c$ 는 상기 다수의 도시들의 개수임.

#### 청구항 5

제4항에 있어서,

상기 중심도 연산부는 하기 수학식 2를 이용하여 상기 제1 도시의 제2 중심도를 연산하는

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치.

[수학식 2]

$$C_{do}(x) = \frac{1}{T_C - 1} \sum_{x \neq y \in G_C} len_{\min}(x, y)$$

$C_{do}(x)$ 는 제1 도시  $x$ 의 제2 중심도이고,  $y$ 는 도시 그룹  $G_C$ 에 포함되어 있는 도시이고,  $len_{\min}(x, y)$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 제1 도시  $x$  및 도시  $y$  간의 최단 거리임.

#### 청구항 6

제5항에 있어서,

상기 중심도 연산부는 하기 수학식 3을 이용하여 상기 제1 도시의 제3 중심도를 연산하는

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치.

[수학식 3]

$$C_{bet}(x) = \frac{1}{T_C - 1} \sum_{x \neq y \in G_C} \sum_{x \neq z \in G_C} \frac{\sigma_{y,z}(x)}{\sigma_{y,z}}$$

$C_{bet}(x)$ 는 제1 도시  $x$ 의 제3 중심도이고,  $z$ 는 도시 그룹  $G_C$ 에 포함되어 있는 도시이고,  $\sigma_{y,z}$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 도시  $y$ 와 도시  $z$  간의 다수의 최단 경로들의 개수이고,  $\sigma_{y,z}(x)$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 도시  $y$ 와 도시  $z$  간의 다수의 최단 경로들 중 제1 도시  $x$ 를 경유하는 최단 경로들의 개수임.

#### 청구항 7

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법에 있어서,

사전 설정된 다수의 토픽 키워드들 및 다수의 학회들 각각의 웹 페이지에 업로드된 다수의 학술 대회 정보들을 저장하는 저장부를 유지하는 단계;

상기 저장부에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출하는 단계;

상기 추출된 토픽 키워드의 중복 횟수에 기초하여 상기 제1 도시에 대한 제1 도시 토픽 벡터를 생성하는 단계;

상기 제1 도시에 대한 제1 도시 토픽 벡터와 상기 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 유사도를 측정하는 단계;

상기 측정된 유사도에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정하는 단계; 및

상기 결정된 학술 주제 거리를 사용자 단말에 전송하는 단계

를 포함하고,

상기 학술 주제 거리를 결정하는 단계는 상기 제1 도시와 상기 다수의 도시들 중 제2 도시 간의 유사도가 사전

설정된 기준 유사도 이상이면 상기 제1 도시와 상기 제2 도시 간의 학술 주제 거리를 1로 결정하며,

상기 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법은

상기 학술 주제 거리를 결정하는 단계에서 결정된 상기 제1 도시와 상기 다수의 도시들 각각 간의 상기 학술 주제 거리에 기초하여, 상기 제1 도시의 중심도를 연산하는 중심도 연산 단계

를 더 포함하고,

상기 전송하는 단계는 상기 결정된 학술 주제 거리 및 중심도를 상기 사용자 단말에 전송하는 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법.

#### 청구항 8

삭제

#### 청구항 9

삭제

#### 청구항 10

제7항에 있어서,

상기 중심도 연산 단계는 하기 수학적 식 1을 이용하여 상기 제1 도시의 제1 중심도를 연산하는

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법.

[수학적 식 1]

$$C_{deg}(x) = \frac{n_{edge}(x)}{T_C - 1}$$

$C_{deg}(x)$ 는 제1 도시 x의 제1 중심도이고,  $n_{edge}(x)$ 는 상기 다수의 도시들 중 제1 도시 x와 학술 주제 거리가 1로 결정된 도시들의 개수이고,  $T_C$ 는 상기 다수의 도시들의 개수임.

#### 청구항 11

제10항에 있어서,

상기 중심도 연산 단계는 하기 수학적 식 2를 이용하여 상기 제1 도시의 제2 중심도를 연산하는

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법.

[수학적 식 2]

$$C_{do}(x) = \frac{1}{T_C - 1} \sum_{x \neq y \in G_C} len_{min}(x, y)$$

$C_{do}(x)$ 는 제1 도시 x의 제2 중심도이고, y는 도시 그룹  $G_C$ 에 포함되어 있는 도시이고,  $len_{min}(x, y)$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 제1 도시 x 및 도시 y 간의 최단 거리임.

#### 청구항 12

제11항에 있어서,

상기 중심도 연산 단계는 하기 수학적 식 3을 이용하여 상기 제1 도시의 제3 중심도를 연산하는

텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법.

[수학식 3]

$$C_{bet}(x) = \frac{1}{T_C - 1} \sum_{x \neq y \in G_C, x \neq z \in G_C} \frac{\sigma_{y,z}(x)}{\sigma_{y,z}}$$

$C_{bet}(x)$ 는 제1 도시  $x$ 의 제3 중심도이고,  $z$ 는 도시 그룹  $G_C$ 에 포함되어 있는 도시이고,  $\sigma_{y,z}$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 도시  $y$ 와 도시  $z$  간의 다수의 최단 경로들의 개수이고,  $\sigma_{y,z}(x)$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 도시  $y$ 와 도시  $z$  간의 다수의 최단 경로들 중 제1 도시  $x$ 를 경유하는 최단 경로들의 개수임.

### 청구항 13

제7항, 제10항, 제11항 또는 제12항 중 어느 한 항의 방법을 컴퓨터로 하여금 수행하도록 하는 프로그램을 기록한 컴퓨터 판독 가능 기록 매체.

### 청구항 14

제7항, 제10항, 제11항 또는 제12항 중 어느 한 항의 방법을 컴퓨터와의 결합을 통해 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램.

## 발명의 설명

### 기술 분야

[0001] 본 발명은 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치 및 이의 동작 방법에 대한 것이다.

### 배경 기술

[0002] 인터넷의 보급이 확대됨에 따라 다양한 정보들을 포함하고 있는 웹 페이지들이 등장하고 있다.

[0003] 이렇게 다양한 정보들을 포함하고 있는 웹 페이지가 생산되고 배포됨에 따라, 이러한 웹 페이지들에 포함되어 있는 정보들의 분석을 통해서, 경제 흐름이나 여론 흐름 등을 분석하는 빅데이터 기반의 분석 기술들도 등장하고 있다.

[0004] 웹 페이지를 통한 정보 분석은 비정형 텍스트 데이터에서 새롭고 유용한 정보를 찾아내는 기술인 텍스트 마이닝을 통해 웹 페이지에 포함되어 있는 각종 텍스트들을 분석하고, 이로부터 소정의 의미를 찾아내는 형태로 이루어지고 있다.

[0005] 예컨대, 연구에 대한 정보를 주고받는 학회와 관련된 웹 페이지에서 학회 관리자가 웹 페이지 상에 남겨놓은 학술 정보들을 분석해서 현재 어떠한 주제가 인기를 끌고 있는지 등을 예측하는 시스템들이 존재한다.

[0006] 최근에는 연구 협업 등의 이유로 학회들의 정보를 데이터베이스로 구축하려고 하는 시도가 증가하고 있다. 다양한 학회들의 정보를 수집하여 데이터베이스로 구축할 때에는 각 학회들이 어떠한 주제들에 대해 논의하고 있는지를 확인해서 지역별로 각 학회들이 논의하고 있는 주제 정보를 저장해야 할 필요가 있다.

[0007] 이와 관련해서, 대부분의 학회들은 홍보의 목적으로 다양한 정보가 포함된 웹 페이지를 구축해서 운영하고 있다는 점에서, 이러한 각 학회들의 웹 페이지에 대한 정보 분석을 통해 각 학회들이 포함된 지역들 간의 협업 가능성을 평가할 수 있는 정보를 자동으로 가공해서 관리자에게 제공할 수 있는 기술에 대한 연구가 필요하다.

## 발명의 내용

### 해결하려는 과제

[0008] 본 발명에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치 및 이의 동작 방법은 제1 도시에 대한 제1 도시 토픽 벡터와 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 제1 도시와 다수의 도시들 각각 간의 유사도를 측정하고, 측정된 유사도에 기초하여, 제1 도시와 다수의 도시들 각각 간의 학술 주제 거리를 결정한 후 결정된 학술 주제 거리를 사용자 단말에 전송함으로써, 각 학회들이 포함된 지역들 간의 협업 가능성을 평가할 수 있는 정보를 자동으로 가공해서 사용자에게 제공하고자 한다.

## 과제의 해결 수단

[0009] 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치는 사전 설정된 다수의 토픽 키워드들 및 다수의 학회들 각각의 웹 페이지에 업로드된 다수의 학술 대회 정보들을 저장하는 저장부, 상기 저장부에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출하는 토픽 키워드 추출부, 상기 추출된 토픽 키워드의 중복 횟수에 기초하여 상기 제1 도시에 대한 제1 도시 토픽 벡터를 생성하는 토픽 벡터 생성부, 상기 제1 도시에 대한 제1 도시 토픽 벡터와 상기 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 유사도를 측정하는 유사도 측정부, 상기 측정된 유사도에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정하는 학술 주제 거리 결정부 및 상기 결정된 학술 주제 거리를 사용자 단말에 전송하는 전송부를 포함한다.

[0010] 또한, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법은 사전 설정된 다수의 토픽 키워드들 및 다수의 학회들 각각의 웹 페이지에 업로드된 다수의 학술 대회 정보들을 저장하는 저장부를 유지하는 단계, 상기 저장부에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출하는 단계, 상기 추출된 토픽 키워드의 중복 횟수에 기초하여 상기 제1 도시에 대한 제1 도시 토픽 벡터를 생성하는 단계, 상기 제1 도시에 대한 제1 도시 토픽 벡터와 상기 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 유사도를 측정하는 단계, 상기 측정된 유사도에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정하는 단계 및 상기 결정된 학술 주제 거리를 사용자 단말에 전송하는 단계를 포함한다.

## 발명의 효과

[0011] 본 발명에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치 및 이의 동작 방법은 제1 도시에 대한 제1 도시 토픽 벡터와 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 제1 도시와 다수의 도시들 각각 간의 유사도를 측정하고, 측정된 유사도에 기초하여, 제1 도시와 다수의 도시들 각각 간의 학술 주제 거리를 결정한 후 결정된 학술 주제 거리를 사용자 단말에 전송함으로써, 각 학회들이 포함된 지역들 간의 협업 가능성을 평가할 수 있는 정보를 자동으로 가공해서 사용자에게 제공할 수 있다.

## 도면의 간단한 설명

[0012] 도 1은 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 구성을 도시한 도면이다.

도 2는 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법을 도시한 흐름도이다.

## 발명을 실시하기 위한 구체적인 내용

[0013] 이하에서는 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명하기로 한다. 이러한 설명은 본 발명을 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 유사한 구성요소에 대해 사용하였으며, 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 본 명세서 상에서 사용되는 모든 용어들은 본 발명이 속하는 기술분야에서 통상의 지식을 가진 사람에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다.

[0014] 도 1은 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)의 구조를 도시한 도면이다.

[0015] 도 1을 참조하면, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)는 저장부(110), 토픽 키워드 추출부(120), 토픽 벡터 생성부(130), 유사도 측정부(140), 학술 주제 거리 결정부(150) 및 전송부(160)를 포함할 수 있다.

[0016] 먼저, 저장부(110)는 사전 설정된 다수의 토픽 키워드들을 하기 표 1과 같이 유사한 키워드별로 그룹화하여 저장할 수 있다.

표 1

[0017]

| 토픽 키워드 그룹 | 토픽 키워드 |
|-----------|--------|
| 그룹 1      | 법      |
|           | 공공     |
|           | 정치     |
|           | ...    |
| 그룹 2      | 인포메이션  |
|           | 콘텐츠    |
|           | 미디어    |
|           | ...    |
| ...       | ...    |

[0018] 여기에서, 그룹 1은 사회과학 분야에 대한 토픽 키워드들을 포함하는 토픽 키워드 그룹이고, 그룹 2는 전자상거래 분야에 대한 토픽 키워드들을 포함하는 토픽 키워드 그룹이다. 또한, 저장부(110)는 다수의 학회들 각각의 웹 페이지로부터 획득된 학회 정보를 하기 표 2와 같이 저장할 수 있다. 여기에서, 학회 정보는 학회 명칭, 주소 정보, 학술 대회 정보 및 발표 주제 정보를 포함할 수 있다.

표 2

[0019]

| 학회   | 주소   | 학술 대회   | 발표 주제   |
|------|------|---------|---------|
| 학회 1 | 주소 1 | 학술 대회 1 | 발표 주제 1 |
|      |      | ...     | ...     |
| 학회 2 | 주소 2 | 학술 대회 2 | 발표 주제 2 |
|      |      | ...     | ...     |
| ...  | ...  | ...     | ...     |

[0020] 여기에서, 웹 페이지 주소가 'http://www.calsec.or.kr/' 인 경우, 상기 웹 페이지로부터 획득되는 학회 정보에 포함되는 학회 1은 '한국전자거래학회' 이고, 주소 1은 '서울시 강남구 역삼로 120 정보역삼빌딩 4층' 이며, 학술 대회 1은 '2017년도 추계학술대회' 이고, 발표 주제 1은 '스마트 인포메이션', '인포메이션 시스템', '콘텐츠/서비스' 및 '스마트 미디어' 일 수 있다. 여기에서, 저장부(110)에 저장되는 학회 정보는 텍스트 추출부(113)에 입력된 웹 페이지에 대한 접속 주소에 기초하여 추출된 정보일 수 있다. 구체적으로, 텍스트 추출부(113)는 제1 학회의 웹 페이지에 대한 접속 주소가 입력되면, 상기 접속 주소를 기초로 상기 제1 학회의 웹 페이지에 접속하여 상기 제1 학회의 웹 페이지로부터 상기 제1 학회의 웹 페이지 상에 존재하는 복수의 제1 텍스트들을 추출할 수 있다.

[0021] 이때, 본 발명의 일실시예에 따르면, 텍스트 추출부(113)는 상기 제1 학회의 웹 페이지를 구성하는 HTML(Hypertext Markup Language) 코드를 파싱(parsing)하여 상기 HTML 코드 상에서 텍스트 입력과 연관된 태그(tag)를 통해 삽입되어 있는 텍스트들을 추출함으로써, 상기 제1 학회의 웹 페이지 상에 존재하는 상기 복수의 제1 텍스트들을 추출하되, 상기 HTML 코드 상에 하이퍼링크 태그가 존재하는 경우, 상기 하이퍼링크 태그를 통해 링크되어 있는 서브 페이지에 접속하여 상기 서브 페이지의 HTML 코드로부터 텍스트 입력과 연관된 태그를 통해 삽입되어 있는 텍스트들도 함께 추출함으로써, 상기 제1 학회의 웹 페이지 상에 존재하는 상기 복수의 제1 텍스트들에 대한 추출을 수행할 수 있다.

[0022] 관련해서, 텍스트 추출부(113)는 제1 학회의 웹 페이지를 구성하는 HTML 코드에서 텍스트 입력과 연관된 태그를 통해 삽입되어 있는 텍스트들을 추출하되, "<a href>"와 같은 하이퍼링크 태그가 존재하는 경우, 해당 하이퍼링크 태그를 통해 링크되어 있는 서브 페이지에 접속해서 상기 서브 페이지의 HTML 코드로부터 텍스트 입력과 연관된 태그를 통해 삽입되어 있는 텍스트들을 함께 추출함으로써, 상기 제1 학회의 웹 페이지 상에 존재하는 복수의 제1 텍스트들을 추출한 후 저장부(110)에 저장할 수 있다.

[0023] 토픽 키워드 추출부(120)는 상기 저장부(110)에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트

콘텐츠들로부터 상기 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출할 수 있다. 구체적으로, 토픽 키워드 추출부(120)는 학회의 주소 정보에 기초하여 학회가 제1 도시에 위치하는지 여부를 판별한 후 학회가 제1 도시에 위치하면, 상기 학회에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출할 수 있다.

[0024] 예를 들어, 제1 도시가 서울인 경우, 토픽 키워드 추출부(120)는 서울에 위치하는 한국전자거래학회에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들, 즉, 학술 대회 정보 및 발표 주제 정보에 포함된 텍스트 콘텐츠들로부터 상기 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 ‘인포메이션’, ‘콘텐츠’ 및 ‘미디어’를 토픽 키워드로 추출할 수 있다. 한편, 발표 주제 1에 대한 텍스트 콘텐츠들 중 ‘스마트’, ‘시스템’ 및 ‘서비스’가 저장부(110)에 토픽 키워드로 저장되어 있는 경우, 토픽 키워드 추출부(120)는 ‘스마트’, ‘시스템’ 및 ‘서비스’를 토픽 키워드로 추출할 수도 있다. 또한, 학회 2에 대한 주소 정보가 서울인 경우, 토픽 키워드 추출부(120)는 학회 2의 학회 정보로부터 토픽 키워드를 추출할 수 있다.

[0025] 토픽 벡터 생성부(130)는 상기 추출된 토픽 키워드의 중복 횟수에 기초하여 상기 제1 도시에 대한 제1 도시 토픽 벡터를 생성할 수 있다. 구체적으로, 토픽 벡터 생성부(130)는 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들의 저장 순번에 기초하여 제1 도시 토픽 벡터에 포함된 각 원소들의 순번을 결정하고, 제1 도시 토픽 벡터에 포함된 각 원소들 중 제1 원소의 값을 추출된 토픽 키워드들의 전체 개수 대비 상기 제1 원소의 순번과 상응하는 저장 순번을 가지는 토픽 키워드가 추출된 개수의 비율로 결정할 수 있다.

[0026] 예를 들어, 추출된 토픽 키워드들의 전체 개수가 10개이고, ‘인포메이션’, ‘콘텐츠’ 및 ‘미디어’가 저장부(110)에 저장되어 있는 순번이 5 내지 7인 경우, ‘인포메이션’이 추출된 개수가 2개이고, ‘콘텐츠’가 추출된 개수가 1개이고, ‘미디어’가 추출된 개수가 1개이므로, 토픽 벡터 생성부(130)는 서울의 도시 토픽 벡터의 5 내지 7번째 원소의 값을 0.2, 0.1 및 0.1로 결정할 수 있다.

[0027] 유사도 측정부(140)는 상기 제1 도시에 대한 제1 도시 토픽 벡터와 상기 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 유사도를 측정할 수 있다. 여기에서, 벡터 연산은 두 개의 벡터 간의 벡터 내적(또는, 두 개의 벡터가 이루는 각도의 코사인 값을 구하기 위한 연산)일 수 있다. 이에 따라, 유사도는 0 내지 1 사이의 값을 가질 수 있다.

[0028] 예를 들어, 상기 다수의 도시들이 서울, 대전, 대구 및 부산인 경우, 유사도 측정부(140)는 하기 표 3과 같이 상기 다수의 도시들 각각 간의 유사도를 측정할 수 있다.

표 3

|    | 서울  | 대전  | 대구  | 부산  |
|----|-----|-----|-----|-----|
| 서울 | 0   | 0.7 | 0.4 | 0.6 |
| 대전 | 0.7 | 0   | 0.8 | 0.3 |
| 대구 | 0.4 | 0.8 | 0   | 0.2 |
| 부산 | 0.6 | 0.3 | 0.2 | 0   |

[0030] 학술 주제 거리 결정부(150)는 상기 측정된 유사도에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정할 수 있다. 구체적으로, 상기 학술 주제 거리 결정부(150)는 상기 제1 도시와 상기 다수의 도시들 중 제2 도시 간의 유사도가 사전 설정된 기준 유사도 이상이면 상기 제1 도시와 상기 제2 도시 간의 학술 주제 거리를 1로 결정할 수 있다. 여기에서, 사전 설정된 기준 유사도는 0.5일 수 있다.

[0031] 예를 들어, 상기 다수의 도시들 각각 간의 유사도가 상기 표 3과 같은 경우, 학술 주제 거리 결정부(150)는 하기 표 4와 같이 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정할 수 있다.

표 4

|    | 서울 | 대전 | 대구 | 부산 |
|----|----|----|----|----|
| 서울 | 0  | 1  | 0  | 1  |
| 대전 | 1  | 0  | 1  | 0  |
| 대구 | 0  | 1  | 0  | 0  |
| 부산 | 1  | 0  | 0  | 0  |

- [0033] 전송부(160)는 상기 결정된 학술 주제 거리를 사용자 단말(170)에 전송할 수 있다.
- [0034] 상술한 바와 같은 본 발명에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)는 제1 도시에 대한 제1 도시 토픽 벡터와 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 제1 도시와 다수의 도시들 각각 간의 유사도를 측정하고, 측정된 유사도에 기초하여, 제1 도시와 다수의 도시들 각각 간의 학술 주제 거리를 결정한 후 결정된 학술 주제 거리를 사용자 단말에 전송함으로써, 각 학회들이 포함된 지역들 간의 협업 가능성을 평가할 수 있는 정보를 자동으로 가공해서 사용자에게 제공할 수 있다.
- [0035] 또한, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)는 중심도 연산부(180)를 더 포함할 수 있다.
- [0036] 중심도 연산부(180)는 상기 학술 주제 거리 결정부(150)가 결정한 상기 제1 도시와 상기 다수의 도시들 각각 간의 상기 학술 주제 거리에 기초하여, 상기 제1 도시의 중심도를 연산할 수 있다.
- [0037] 구체적으로, 상기 중심도 연산부는 하기 수학적 식 1 내지 3을 이용하여 상기 제1 도시의 제1 내지 3 중심도를 연산할 수 있다.

### 수학식 1

$$C_{deg}(x) = \frac{n_{edge}(x)}{T_C - 1}$$

- [0038]
- [0039] 여기에서,  $C_{deg}(x)$ 는 제1 도시 x의 제1 중심도이고,  $n_{edge}(x)$ 는 상기 다수의 도시들 중 제1 도시 x와 학술 주제 거리가 1로 결정된 도시들의 개수이고,  $T_C$ 는 상기 다수의 도시들의 개수이다.
- [0040] 예를 들어, 학술 주제 거리 결정부(150)가 상기 표 4와 같이 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정한 경우, 서울과 학술 주제 거리가 1로 결정된 도시들(즉, 대전 및 부산)의 개수는 2이다.

### 수학식 2

$$C_{do}(x) = \frac{1}{T_C - 1} \sum_{x \neq y \in G_C} len_{min}(x, y)$$

- [0041]
- [0042]  $C_{do}(x)$ 는 제1 도시 x의 제2 중심도이고, y는 도시 그룹  $G_C$ 에 포함되어 있는 도시이고,  $len_{min}(x, y)$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 제1 도시 x 및 도시 y 간의 최단 거리이다.
- [0043] 예를 들어, 학술 주제 거리 결정부(150)가 상기 표 4와 같이 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정한 경우, 서울과 대전, 부산 간의 학술 주제 거리에 기초하여 결정되는 최단 거리는 1이고, 서울과 대구 간의 학술 주제 거리에 기초하여 결정되는 최단 거리는 2이다.

### 수학식 3

$$C_{bet}(x) = \frac{1}{T_C - 1} \sum_{x \neq y \in G_C, x \neq z \in G_C} \frac{\sigma_{y,z}(x)}{\sigma_{y,z}}$$

- [0044]
- [0045]  $C_{bet}(x)$ 는 제1 도시 x의 제3 중심도이고, z는 도시 그룹  $G_C$ 에 포함되어 있는 도시이고,  $\sigma_{y,z}$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 도시 y와 도시 z 간의 다수의 최단 경로들의 개수이고,  $\sigma_{y,z}(x)$ 는 상기 다수의 도시들 각각 간의 학술 주제 거리에 기초하여 결정되는 도시 y와 도시 z 간의 다수의 최

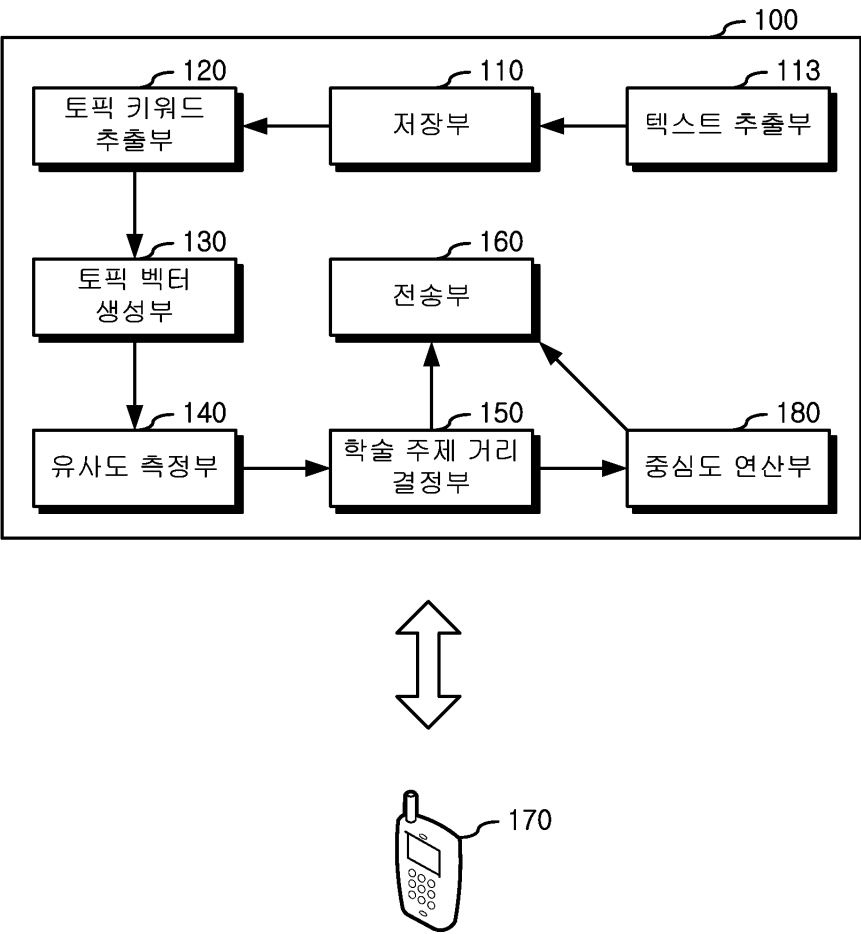
단 경로들 중 제1 도시  $x$ 를 경유하는 최단 경로들의 개수이다.

- [0046] 예를 들어, 학술 주제 거리 결정부(150)가 상기 표 4와 같이 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정한 경우, 대전 및 부산 간의 최단 경로(즉, 대전, 서울 및 부산의 순서에 따른 경로)의 개수는 1개이고, 1개의 최단 경로 중 서울을 경유하는 최단 경로도 1개이다.
- [0047] 예를 들어, 학술 주제 거리 결정부(150)가 상기 표 4와 같이 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정한 경우, 서울에 대한 제1 내지 3 중심도는 0.67, 1.33 및 0.67로 결정될 수 있다.
- [0048] 상기 전송부(160)는 상기 결정된 학술 주제 거리 및 중심도를 상기 사용자 단말(170)에 전송할 수 있다.
- [0049] 또한, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)의 토픽 키워드 추출부(120)는 상기 저장부(110)에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출하되, 사전 설정된 기간에 상기 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들 중 사전 설정된 기간 이내에 웹 페이지에 업로드된 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출할 수 있다. 여기에서, 사전 설정된 기간은 사용자 단말(170)로부터 입력된 기간일 수 있다.
- [0050] 이에 따라, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)는 사용자에게 기간별 지역별 연구 트렌드 정보를 제공할 수 있다.
- [0051] 또한, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)의 토픽 키워드 추출부(120)는 상기 저장부(110)에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부(110)에 저장되어 있는 상기 다수의 토픽 키워드들 중 사전 설정된 토픽 키워드 그룹에 포함되어 있는 토픽 키워드들에 매칭되는 토픽 키워드를 추출할 수 있다. 여기에서, 사전 설정된 토픽 키워드 그룹은 사용자 단말(170)로부터 입력된 토픽 키워드 그룹일 수 있다.
- [0052] 이후, 전송부(160)는 상기 결정된 학술 주제 거리 및 중심도를 상기 사용자 단말(170)에 전송하면서, 동시에, 사전 설정된 토픽 키워드 그룹에 포함되어 있는 토픽 키워드들에 매칭되는 토픽 키워드에 기초하여 결정된 학술 주제 거리에 따른 사전 설정된 다수의 도시들 각각의 중심도에 기초하여, 사용자의 주소로부터 사전 결정된 범위 이내에 위치하는 도시들 중 가장 높은 중심도를 가지는 도시를 활동 추천 도시로 전송할 수 있다. 구체적으로, 전송부(160)는 제1 내지 3 중심도 각각에 따른 도시 순위를 결정한 후 도시 순위의 평균 값에 기초하여 가장 낮은 도시 순위 평균 값을 가지는 도시를 활동 추천 도시로 전송할 수 있다.
- [0053] 이에 따라, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)는 사용자에게 주제별 지역별 연구 트렌드 정보를 제공할 수 있다.
- [0054] 도 2는 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)의 동작 방법을 도시한 흐름도이다.
- [0055] 도 2를 참조하면, 단계(S210)에서는 사전 설정된 다수의 토픽 키워드들 및 다수의 학회들 각각의 웹 페이지에 업로드된 다수의 학술 대회 정보들을 저장하는 저장부를 유지할 수 있다.
- [0056] 단계(S220)에서는 상기 저장부에 저장되어 있는 상기 다수의 학술 대회 정보들을 중 사전 설정된 다수의 도시들 중 제1 도시에 위치하는 다수의 학회들 각각에 대한 다수의 학술 대회 정보들에 포함된 텍스트 콘텐츠들로부터 상기 저장부에 저장되어 있는 상기 다수의 토픽 키워드들에 매칭되는 토픽 키워드를 추출할 수 있다.
- [0057] 단계(S230)에서는 상기 추출된 토픽 키워드의 중복 횟수에 기초하여 상기 제1 도시에 대한 제1 도시 토픽 벡터를 생성할 수 있다.
- [0058] 단계(S240)에서는 상기 제1 도시에 대한 제1 도시 토픽 벡터와 상기 다수의 도시들 각각에 대한 제2 도시 토픽 벡터 간의 벡터 연산에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 유사도를 측정할 수 있다.
- [0059] 단계(S250)에서는 상기 측정된 유사도에 기초하여, 상기 제1 도시와 상기 다수의 도시들 각각 간의 학술 주제 거리를 결정할 수 있다.

- [0060] 상기 단계(S250)에서는 상기 제1 도시와 상기 다수의 도시들 중 제2 도시 간의 유사도가 사전 설정된 기준 유사도 이상이면 상기 제1 도시와 상기 제2 도시 간의 학술 주제 거리를 1로 결정할 수 있다.
- [0061] 단계(S260)에서는 상기 결정된 학술 주제 거리를 사용자 단말에 전송할 수 있다.
- [0062] 상술한 바와 같은 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치(100)의 동작 방법은 상기 단계(S250)에서 결정된 상기 제1 도시와 상기 다수의 도시들 각각 간의 상기 학술 주제 거리에 기초하여, 상기 제1 도시의 중심도를 연산하는 단계를 더 수행하고, 상기 단계(S260)에서는 상기 결정된 학술 주제 거리 및 중심도를 상기 사용자 단말에 전송할 수 있다.
- [0063] 한편, 상기 중심도 연산 단계는 상기 수학적 식 1을 이용하여 상기 제1 도시의 제1 중심도를 연산할 수 있다. 또한, 상기 중심도 연산 단계는 상기 수학적 식 2를 이용하여 상기 제1 도시의 제2 중심도를 연산할 수 있다. 마찬가지로, 상기 중심도 연산 단계는 상기 수학적 식 3을 이용하여 상기 제1 도시의 제3 중심도를 연산할 수 있다.
- [0064] 이상, 도 2를 참조하여 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법에 대해 설명하였다. 여기서, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법은 도 1을 이용하여 설명한 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작에 대한 구성과 대응될 수 있으므로, 이에 대한 보다 상세한 설명은 생략하기로 한다.
- [0065] 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법은 컴퓨터와의 결합을 통해 실행시키기 위한 저장매체에 저장된 컴퓨터 프로그램으로 구현될 수 있다.
- [0066] 또한, 본 발명의 일실시예에 따른 텍스트 마이닝 기반의 지역별 연구 트렌드 분석 장치의 동작 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.
- [0067] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.
- [0068] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

도면

도면1



도면2

