



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2016년06월13일
(11) 등록번호 10-1629210
(24) 등록일자 2016년06월03일

(51) 국제특허분류(Int. Cl.)
G06F 17/24 (2006.01) G06F 17/30 (2006.01)
(52) CPC특허분류
G06F 17/24 (2013.01)
G06F 17/30011 (2013.01)
(21) 출원번호 10-2015-0015487
(22) 출원일자 2015년01월30일
심사청구일자 2015년01월30일
(56) 선행기술조사문헌
KR1020030039576 A
KR101033611 B1*
KR101413444 B1*
KR1020130064757 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
인하대학교 산학협력단
인천광역시 남구 인하로 100, 인하대학교 (용현동)
(주)다울소프트
서울특별시 서초구 서초중앙로20길 29, 2층 (서초동, 원성빌딩)
(72) 발명자
김유성
인천 중구 신도시북로43번길 27, 112동 101호 (운서동, 동원베네스트영종타운하우스)
송광호
서울 구로구 디지털로31길 90, 115동 1502호 (구로동, 삼성래미안아파트)
민지홍
경기 부천시 원미구 평천로 616, 2110동 1504호 (상동, 다정한마을삼성래미안아파트)
(74) 대리인
양성보

전체 청구항 수 : 총 9 항

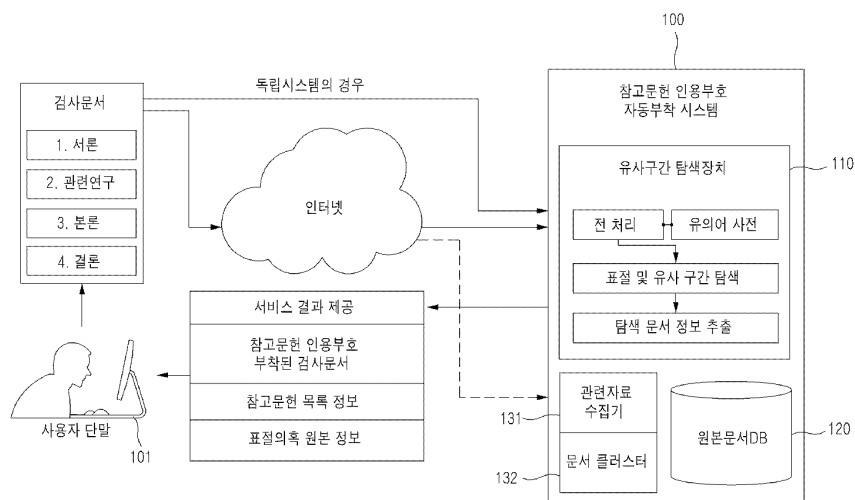
심사관 : 경연정

(54) 발명의 명칭 온라인 서비스가 가능한 참고문헌 인용부호 자동부착 지원 시스템 및 서비스 제공 방법

(57) 요약

온라인 서비스가 가능한 참고문헌 인용부호 자동부착 지원 시스템 및 서비스 제공 방법이 개시된다. 참고문헌 인용부호 자동부착 시스템은, 적어도 하나의 프로그램이 로딩된 메모리; 및 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 상기 프로그램의 제어에 따라, 검사문서를 문장 단위로 분리한 검사문장을 대상으로 원본문서에 포함된 원본문장과 유사도 검사를 수행하는 과정; 및 상기 검사문장과 상기 원본문장의 유사도가 일정기준을 초과하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 정보로 제공하는 과정을 처리한다.

대표도



공지예외적용 : 있음

명세서

청구범위

청구항 1

적어도 하나의 프로그램이 로딩된 메모리; 및

적어도 하나의 프로세서

를 포함하고,

상기 적어도 하나의 프로세서는, 상기 프로그램의 제어에 따라,

검사문서를 문장 단위로 분리한 검사문장을 대상으로 원본문서에 포함된 원본문장과 유사도 검사를 수행하는 과정; 및

상기 검사문장과 상기 원본문장의 유사도가 일정기준을 초과하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 정보로 제공하는 과정

을 처리하고,

상기 제공하는 과정은,

상기 검사문서를 구분하는 목차를 기준으로 상기 검사문장의 위치를 판단하는 과정;

상기 검사문장이 상기 목차 중 서론 또는 관련연구에 위치하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 상기 검사문장에 부착하는 과정; 및

상기 검사문장이 상기 목차 중 본론 또는 결론에 위치하는 경우 상기 검사문장의 수정을 지원하는 관련 정보로서 상기 원본문장 및 상기 서지정보를 상기 검사문장과 함께 표시하는 과정

을 포함하는 참고문헌 인용부호 자동부착 시스템.

청구항 2

제1항에 있어서,

상기 제공하는 과정은,

상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 상기 검사문장에 부착하는 것

을 특징으로 하는 참고문헌 인용부호 자동부착 시스템.

청구항 3

제1항에 있어서,

상기 제공하는 과정은,

상기 원본문서의 서지정보와 함께 상기 검사문장에 참고문헌에 대한 인용부호를 부착하기 위한 문서편집기용 API(application program interface)를 제공하는 것

을 특징으로 하는 참고문헌 인용부호 자동부착 시스템.

청구항 4

삭제

청구항 5

적어도 하나의 프로그램이 로딩된 메모리; 및

적어도 하나의 프로세서

를 포함하고,

상기 적어도 하나의 프로세서는, 상기 프로그램의 제어에 따라,

검사문서를 문장 단위로 분리한 검사문장을 대상으로 원본문서에 포함된 원본문장과 유사도 검사를 수행하는 과정; 및

상기 검사문장과 상기 원본문장의 유사도가 일정기준을 초과하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 정보로 제공하는 과정

을 처리하고,

상기 수행하는 과정은,

상기 원본문장과 상기 검사문장을 각각 단어 단위로 분할하여 유의어 사전에서 검색된 대표 유의어와 함께 데이터베이스에 저장하는 전처리 과정;

상기 원본문장 중에서 자카드 계수(Jaccard Coefficient) 기반의 유사도를 기준으로 상기 검사문장과 유사한 제1 문장을 선별하는 과정; 및

상기 제1 문장 중에서 코사인(cosine) 거리 기반의 유사도를 기준으로 상기 검사문장과 유사한 제2 문장을 선별하는 과정

을 포함하는 참고문헌 인용부호 자동부착 시스템.

청구항 6

제5항에 있어서,

상기 제2 문장을 선별하는 과정은,

상기 제1 문장에 포함된 단어의 대표 유의어와 해당 단어의 출현 빈도를 저장하는 제1 벡터와 상기 검사문장에 포함된 단어의 대표 유의어와 해당 단어의 출현 빈도를 저장하는 제2 벡터를 생성하는 과정;

상기 제1 벡터와 상기 제2 벡터의 차원을 동기화 하여 코사인 유사도를 계산하는 과정; 및

상기 코사인 유사도가 표절 판정 기준 이상인 제1 문장을 상기 제2 문장으로 선별하는 과정

을 포함하는 참고문헌 인용부호 자동부착 시스템.

청구항 7

제6항에 있어서,

상기 코사인 유사도를 계산하는 과정은,

상기 제1 벡터와 상기 제2 벡터를 비교하여 서로에게 없는 단어의 빈도를 0으로 하여 상기 제1 벡터와 상기 제2 벡터의 차원을 동기화 하는 과정;

상기 제1 벡터와 상기 제2 벡터를 각각 정규화 하여 크기를 1로 생성하는 과정; 및

정규화 된 상기 제1 벡터와 상기 제2 벡터를 이용하여 상기 코사인 유사도를 계산하는 과정

을 포함하는 참고문헌 인용부호 자동부착 시스템.

청구항 8

컴퓨터로 구현되는 참고문헌 인용부호 자동부착 방법에 있어서,

검사문서를 문장 단위로 분리한 검사문장을 대상으로 원본문서에 포함된 원본문장과 유사도 검사를 수행하는 단계; 및

상기 검사문장과 상기 원본문장의 유사도가 일정기준을 초과하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 정보로 제공하는 단계

를 포함하고,

상기 제공하는 단계는,

상기 검사문서를 구분하는 목차를 기준으로 상기 검사문장의 위치를 판단하는 단계;

상기 검사문장이 상기 목차 중 서론 또는 관련연구에 위치하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 상기 검사문장에 부착하는 단계; 및

상기 검사문장이 상기 목차 중 본론 또는 결론에 위치하는 경우 상기 검사문장의 수정을 지원하는 관련 정보로서 상기 원본문장 및 상기 서지정보를 상기 검사문장과 함께 표시하는 단계

를 포함하는 참고문헌 인용부호 자동부착 방법.

청구항 9

제8항에 있어서,

상기 제공하는 단계는,

상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 상기 검사문장에 부착하는 것

을 특징으로 하는 참고문헌 인용부호 자동부착 방법.

청구항 10

제8항에 있어서,

상기 제공하는 단계는,

상기 원본문서의 서지정보와 함께 상기 검사문장에 참고문헌에 대한 인용부호를 부착하기 위한 문서편집기용 API(application program interface)를 제공하는 것

을 특징으로 하는 참고문헌 인용부호 자동부착 방법.

청구항 11

삭제

발명의 설명

기술 분야

[0001] 본 발명의 실시예들은 참고문헌에 대한 인용부호를 부착하는 기술에 관한 것이다.

배경 기술

[0002] 참고문헌은 논문이나 보고서 등의 문헌에 첨부된 관련 문헌으로, 매우 중요한 부분이다. 흔히 논문의 경우 목차와 서론, 그리고 참고문헌만 보면 그 논문을 평가할 수 있는데 이는 논문을 작성하는데 그 주제와 서론에서 밝힌 내용들에 적합한 문헌을 참고하였는가가 중요하기 때문이다.

[0003] 이러한 참고문헌에 대한 인용 색인은 미국의 ISI사에서 개발한 SCI(Science Citation Index)를 포함한 여러 지표들이 연구되어 왔다.

[0004] 그러나, 이러한 지표들은 참고 문헌을 수작업에 의존해 분석한 결과이며, 문헌 간의 인용 정보 부여로 그 역할이 한정되어 있다.

[0005] 한편, 국내외에서 기술문서에 대한 표절 문제가 대두되고 있으며 이러한 표절 문제가 큰 논란이 되고 있다. 이러한 문제를 해결하기 위하여 독립적으로 작동하거나 또는 인터넷을 통해 온라인으로 기술문서의 표절 여부를 판단해주는 기술들이 연구되어 있다.

[0006] 그러나, 작성중인 기술문서에 대해 향후 표절의혹을 해소할 수 있도록 지원하는 서비스는 현재까지 존재하지 않

는다.

발명의 내용

해결하려는 과제

- [0007] 작성 중인 기술문서를 표절의혹으로부터 자유로울 수 있도록 도움을 지원하는 시스템 및 방법을 제공한다.
- [0008] 기술문서에 참고문헌에 대한 인용부호를 자동으로 부착하는 시스템 및 방법을 제공한다.
- [0009] 표절 의심문장의 기술문서 내 위치에 따라 해당 문장의 수정을 제안하는 시스템 및 방법을 제공한다.

과제의 해결 수단

- [0010] 적어도 하나의 프로그램이 로딩된 메모리; 및 적어도 하나의 프로세서를 포함하고, 상기 적어도 하나의 프로세서는, 상기 프로그램의 제어에 따라, 검사문서를 문장 단위로 분리한 검사문장을 대상으로 원본문서에 포함된 원본문장과 유사도 검사를 수행하는 과정; 및 상기 검사문장과 상기 원본문장의 유사도가 일정기준을 초과하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 정보로 제공하는 과정을 처리하는 참고문헌 인용부호 자동부착 시스템을 제공한다.
- [0011] 일 측면에 따르면, 상기 제공하는 과정은, 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 상기 검사문장에 부착할 수 있다.
- [0012] 다른 측면에 따르면, 상기 제공하는 과정은, 상기 원본문서의 서지정보와 함께 상기 검사문장에 참고문헌에 대한 인용부호를 부착하기 위한 문서편집기용 API(application program interface)를 제공할 수 있다.
- [0013] 또 다른 측면에 따르면, 상기 제공하는 과정은, 상기 검사문서를 구분하는 목차를 기준으로 상기 검사문장의 위치를 판단하는 과정; 상기 검사문장이 상기 목차 중 서론 또는 관련연구에 위치하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 상기 검사문장에 부착하는 과정; 및 상기 검사문장이 상기 목차 중 본론 또는 결론에 위치하는 경우 상기 검사문장의 수정을 지원하는 관련 정보로서 상기 원본문장 및 상기 서지정보를 상기 검사문장과 함께 표시하는 과정을 포함할 수 있다.
- [0014] 또 다른 측면에 따르면, 상기 수행하는 과정은, 상기 원본문장과 상기 검사문장을 각각 단어 단위로 분할하여 유의어 사전에서 검색된 대표 유의어와 함께 데이터베이스에 저장하는 전처리 과정; 상기 원본문장 중에서 자카드 계수(Jaccard Coefficient) 기반의 유사도를 기준으로 상기 검사문장과 유사한 제1 문장을 선별하는 과정; 및 상기 제1 문장 중에서 코사인(cosine) 거리 기반의 유사도를 기준으로 상기 검사문장과 유사한 제2 문장을 선별하는 과정을 포함할 수 있다.
- [0015] 또 다른 측면에 따르면, 상기 제2 문장을 선별하는 과정은, 상기 제1 문장에 포함된 단어의 대표 유의어와 해당 단어의 출현 빈도를 저장하는 제1 벡터와 상기 검사문장에 포함된 단어의 대표 유의어와 해당 단어의 출현 빈도를 저장하는 제2 벡터를 생성하는 과정; 상기 제1 벡터와 상기 제2 벡터의 차원을 동기화 하여 코사인 유사도를 계산하는 과정; 및 상기 코사인 유사도가 표절 판정 기준 이상인 제1 문장을 상기 제2 문장으로 선별하는 과정을 포함할 수 있다.
- [0016] 또 다른 측면에 따르면, 상기 코사인 유사도를 계산하는 과정은, 상기 제1 벡터와 상기 제2 벡터를 비교하여 서로에게 없는 단어의 빈도를 0으로 하여 상기 제1 벡터와 상기 제2 벡터의 차원을 동기화 하는 과정; 상기 제1 벡터와 상기 제2 벡터를 각각 정규화 하여 크기를 1로 생성하는 과정; 및 정규화 된 상기 제1 벡터와 상기 제2 벡터를 이용하여 상기 코사인 유사도를 계산하는 과정을 포함할 수 있다.
- [0017] 컴퓨터로 구현되는 참고문헌 인용부호 자동부착 방법에 있어서, 검사문서를 문장 단위로 분리한 검사문장을 대상으로 원본문서에 포함된 원본문장과 유사도 검사를 수행하는 단계; 및 상기 검사문장과 상기 원본문장의 유사도가 일정기준을 초과하는 경우 상기 원본문서의 서지정보를 상기 검사문장에 대한 참고문헌 정보로 제공하는 단계를 포함하는 참고문헌 인용부호 자동부착 방법을 제공한다.

발명의 효과

- [0018] 본 발명의 실시예에 따르면, 기술문서에 참고문헌에 대한 인용부호를 자동으로 부착함으로써 기술문서의 표절의

혹 해소를 통한 표절 문제의 사회적 이슈화를 사전에 방지할 수 있다.

[0019] 본 발명의 실시예에 따르면, 작성 중인 기술문서 초안에 대하여 유사 문장 검색을 통해 선제적으로 표절의혹 해소를 위한 서비스를 제공함으로써 표절의혹으로부터 자유로운 기술문서 작성 환경을 서비스 할 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 일 실시예에 있어서, 사용자 단말과 참고문헌 인용부호 자동부착 시스템 간의 개괄적인 모습을 도시한 것이다.

도 2는 본 발명의 일 실시예에 있어서, 참고문헌 인용부호 자동부착 시스템의 구조도를 도시한 것이다.

도 3 내지 도 5는 본 발명의 일 실시예에 있어서, 참고문헌 정보를 제공하는 서비스 화면을 예시적으로 도시한 것이다.

도 6은 본 발명의 일 실시예에 있어서, 참고문헌 인용부호 자동부착 방법을 도시한 순서도이다.

도 7은 본 발명의 일 실시예에 있어서, 유의어 사전을 이용한 유사구간 탐색 과정을 도시한 순서도이다.

도 8은 본 발명의 일 실시예에 있어서, 유사구간 탐색을 위한 전처리 과정을 설명하기 위한 예시 도면이다.

도 9는 본 발명의 일 실시예에 있어서, 자카드 계수 기반 필터링 단계의 구체적인 과정을 설명하기 위한 도면이다.

도 10은 본 발명의 일 실시예에 있어서, 벡터 공간 모델의 필터링 단계의 구체적인 과정을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하, 본 발명의 실시예를 첨부된 도면을 참조하여 상세하게 설명한다.

[0022] 본 실시예들은 참고문헌에 대한 인용부호를 부착하는 기술에 관한 것으로, 대내외적으로 공개하는 기술문서에 대해서 사전에 표절의혹 해소를 지원하는 분야로서 학교의 리포트, 학위논문, 기술연구 보고서 등 기술문서를 작성하는 모든 분야에 적용 가능하다.

[0023] 도 1은 본 발명의 일 실시예에 있어서, 사용자 단말과 참고문헌 인용부호 자동부착 시스템 간의 개괄적인 모습을 도시한 것이다. 도 1에서는 참고문헌 인용부호 자동부착 시스템(100) 및 사용자 단말(101)을 도시하고 있다. 도 1에서 화살표는 참고문헌 인용부호 자동부착 시스템(100)과 사용자 단말(101) 간에 유/무선 네트워크를 통해 데이터가 송수신될 수 있음을 의미할 수 있다.

[0024] 사용자 단말(101)은 PC, 노트북, 스마트폰(smart phone), 태블릿(tablet), 웨어러블 컴퓨터(wearable computer) 등으로, 참고문헌 인용부호 자동부착 시스템(100)과 관련된 웹/모바일 사이트의 접속 또는 서비스 전용 어플리케이션의 설치 및 실행이 가능한 모든 단말 장치를 의미할 수 있다. 이때, 사용자 단말(101)은 웹/모바일 사이트 또는 전용 어플리케이션의 제어 하에 서비스 화면 구성, 데이터 입력, 데이터 송수신, 데이터 저장 등 서비스 전반의 동작을 수행할 수 있다.

[0025] 참고문헌 인용부호 자동부착 시스템(100)은 유사문서 탐색을 통해 참고문헌에 대한 정보를 제공한다. 특히, 참고문헌 인용부호 자동부착 시스템(100)은 사용자가 작성 중인 기술문서 초안에 대해 향후 표절의혹을 해소하는데 도움을 주기 위해 표절의혹의 문장에 참고문헌 인용부호를 자동으로 부착하는데 필요한 정보 또는 직접 부착하는 기능을 제공한다.

[0026] 상기한 참고문헌 인용부호 자동부착을 위한 시스템 및 서비스는 단일 컴퓨터 서버에서 독립적으로 운영될 수 있으며, 혹은 인터넷을 통해 지정된 서버로 접속해서 온라인 서비스를 제공 받는 형식으로 구현될 수 있다. 그리고, 참고문헌 인용부호 자동부착 시스템(100)은 적어도 일부의 구성 요소가 사용자 단말(101) 상에 설치되는 어플리케이션 형태로 구현되거나, 혹은 클라이언트-서버 환경에서 서비스를 제공하는 플랫폼에 포함되는 형태로 구현되는 것 또한 가능하다.

[0027] 도 2는 본 발명의 일 실시예에 있어서, 참고문헌 인용부호 자동부착 시스템의 구조도를 도시한 것이다.

[0028] 도 2에 도시한 바와 같이, 일 실시예에 따른 참고문헌 인용부호 자동부착 시스템(100)은 유사구간 탐색장치

(110) 및 그에 부속된 원본문서 DB(120)와 인터넷과 연결된 관련자료 수집기(131) 및 문서 클러스터(132)를 포함한다. 이때, 유사구간 탐색장치(110)는 원본문서 DB(120)에서 기술문서와 유사한 원본문서를 찾는 것으로 유의어 사전을 이용하여 표절 및 유사구간을 탐색하는 역할을 한다. 그리고, 관련자료 수집기(131)는 인터넷으로부터 자동으로 기술문서를 수집하는 웹 크롤링 장치이고 이를 통하여 수집된 자료는 문서 클러스터(132)를 통하여 원본문서 DB(120)에 저장된다.

[0029] 도 2를 참조하면, 사용자(기술문서 작성자)는 사용자 단말(101)을 통해 작성한 기술문서 초안을 직접 또는 인터넷을 통해 참고문헌 인용부호 자동부착 시스템(100)에 검사문서로 입력한다. 이때, 검사문서는 사용자에게 의해 작성된 일반적인 기술문서를 의미하며 세부 목차로서 서론, 관련 연구, 본문, 결론을 갖는 문서이다.

[0030] 이에, 참고문헌 인용부호 자동부착 시스템(100)은 유사구간 탐색장치(110)를 통해 검사문서를 문장 단위로 분석하여 내부에 저장되거나 인터넷으로부터 검색한 원본문서의 문장과 비교하여 일정 기준 이상의 유사성을 가진 문장이 발견되면 이를 포함하는 원본문서의 서지정보를 참고문헌 목록에 포함시키고 해당 참고문헌에 대한 인용부호를 검사문서의 해당 문장에 자동으로 부착하는 기능 또는 부착을 지원하는 응용 프로그램 인터페이스(application program interface: 약칭 API)를 제공할 수 있다.

[0031] 기술문서에 대한 표절의혹을 해결하기 위해서는 제공된 검사문서를 문장 단위로 분할하고 해당 문장이 표절된 것으로 의심 받을 수 있는 유사도가 높은 원본문서의 문장을 참고문헌 인용부호 자동부착 시스템(100)에 저장되어 있는 원본문서 DB(120)로부터 검색한다.

[0032] 이때, 원본문서 DB(120)에 저장된 문서는 참고문헌 인용부호 자동부착 시스템(100)으로 직접 등록된 문서뿐만 아니라, 외부 문서 수집기인 관련자료 수집기(131)가 인터넷으로부터 수집하여 문서 분야 및 종류에 따라 문서 클러스터(132)를 이용하여 정리해 놓은 문서를 모두 포함할 수 있다.

[0033] 검사문서의 문장이 표절된 것으로 의심 받을 수 있는 원본문서의 문장을 탐색하기 위해서는 복제(cloning) 형식의 표절뿐만 아니라 유의어 대체 표절, 문장 구조를 변경한 형태의 표절, 축약 형태의 표절 등까지도 대상으로 검색할 수 있다. 또한, 표절로 의심을 받는 검사문서의 해당 문장이 검사문서 내의 어느 부분에 위치하였는지에 따라 다른 형태로 사용자를 지원할 수 있다.

[0034] 예를 들면, 검사문서의 서론 및 관련연구 부분에서 표절 의심 문장이 발견된 경우에는 이와 유사한 문장을 갖고 있는 원본문서의 서지정보를 참고문헌 형태로 제공하고 검사문서의 해당 문장에는 원본문서에 대한 참고문헌 인용부호를 부착하기 위한 정보 또는 직접 인용부호를 부착하는데 사용할 수 있는 문서편집기용 API를 제공하여 참고문헌 인용부호의 자동부착을 지원한다.

[0035] 또한, 검사문서의 본문 및 결론 부분에서 원본문서의 문장과 유사한 표절의혹 문장이 검색된 경우에는 이 부분을 수정할 수 있도록 관련정보(원본문서의 서지정보 및 표절의심 원본 문장 등)를 지원할 수 있다. 이는 기술문서의 본문 및 결론 부분에서 다른 원본문서의 문장이 일정기준 이상의 높은 유사도로 인용되는 것이 타당하지 않기 때문에 인용부호 부착을 지원하지 않고 검사문서의 해당문장을 수정하도록 제안한다.

[0036] 참고문헌 인용부호 자동부착 시스템(100)은 기술문서를 입력한 사용자 단말(101)로 기술문서에 대한 서비스 결과를 제공한다. 참고문헌 인용부호 자동부착 시스템(100)은 도 3에 도시한 바와 같이 기술문서에 대하여 검사문장(301)과 유사한 원본 문장(302)을 서비스 결과로서 제공할 수 있으며, 일 예로 참고문헌 인용부호가 부착된 검사문서, 참고문헌 인용부호에 대한 부착 기능을 지원할 수 있는 API, 참고문헌 인용부호로 사용할 수 있는 정보 등을 제공할 수 있다. 다른 예로, 참고문헌 인용부호 자동부착 시스템(100)은 도 4에 도시한 바와 같이 참고문헌 목록의 형태로 구성된 원본문서 목록에 대한 정보(401)를 서비스 결과로 제공할 수 있다. 또 다른 예로, 참고문헌 인용부호 자동부착 시스템(100)은 도 5에 도시한 바와 같이 표절 의심 구간의 문장을 포함하는 원본문서의 간략한 정보 및 서지 정보(501)를 서비스 결과로 제공할 수 있다.

[0037] 도 6은 본 발명의 일 실시예에 있어서, 참고문헌 인용부호 자동부착 방법을 도시한 순서도이다.

[0038] 단계(610)에서 참고문헌 인용부호 자동부착 시스템은 사용자로부터 작성 중인 기술문서 초안을 검사문서로 입력 받는다. 이때, 검사문서는 서론, 관련 연구, 본문, 결론의 세부 목차로 작성된 문서이다.

[0039] 단계(620)에서 참고문헌 인용부호 자동부착 시스템은 사용자로부터 입력 받은 검사문서를 문장 단위로 분리하고 동시에 분리된 문장의 문서 내 위치정보를 파악하여 저장한다.

[0040] 단계(630)에서 참고문헌 인용부호 자동부착 시스템은 단계(620)에서 분리된 문장을 원본문서 DB 내의 문서에 속

해 있는 문장들과 비교하여 문장 간의 유사도 검사를 진행한다.

- [0041] 단계(640)에서 참고문헌 인용부호 자동부착 시스템은 단계(630)의 유사도 검사를 통해 검사문서의 표절 의심문장과 유사한 문장을 포함하는 원본문서의 서지정보를 파악하고 해당 문장 및 문서 내 위치를 추출한다.
- [0042] 단계(650)에서 참고문헌 인용부호 자동부착 시스템은 표절 의심문장으로 판단되는 문장의 검사문서 내 위치가 세부 목차(서론, 관련 연구, 본문, 결론) 중 어디에 속하는지 판단한다.
- [0043] 단계(660)에서 참고문헌 인용부호 자동부착 시스템은 표절 의심문장으로 판단되는 문장의 검사문서 내 위치가 서론 또는 관련 연구인 경우 서비스 결과로서 검사문서의 표절 의심문장에 대해 참고문헌 목록 및 인용부호 부착을 지원하는 정보를 제공한다.
- [0044] 단계(670)에서 참고문헌 인용부호 자동부착 시스템은 표절 의심문장으로 판단되는 문장의 검사문서 내 위치가 본문 또는 결론인 경우 서비스 결과로서 검사문서 내에 표절 의심문장을 표시하고 검사문서의 표절 의심문장과 유사한 문장을 포함하는 원본문서의 원본 문장 및 서지 정보를 제공한다.
- [0045] 단계(680)에서 참고문헌 인용부호 자동부착 시스템은 참고문헌 인용부호 부착을 지원하기 위해 단계(620)에서 분리된 문장이 검사문서 내의 마지막 문장인지 여부를 판단하여 마지막 문장인 경우 모든 프로세스를 종료하고 다음 문장이 존재하는 경우 상기한 단계(620~670)를 반복 수행한다.
- [0046] 따라서, 본 발명에서는 작성 중인 기술문서 초안에 대해 선제적으로 표절의혹 해소를 위한 다양한 형태의 서비스를 제공한다.
- [0047] 이하에서는 검사문서에 대한 유사문서를 탐색하는 과정을 구체적으로 설명하기로 한다.
- [0048] 본 발명에 따른 참고문헌 인용부호 자동부착 시스템의 유사구간 탐색장치에서는 전처리 과정에서 원본문서와 검사문서를 단어 단위로 분할한 후 색인어 집합과 이에 대한 대표 유의어를 유의어 사전에서 검색하여 색인어 자체와 색인어의 문장 내 위치 정보, 그리고 대표 유의어를 함께 저장하고 이를 기반으로 의역 및 문장 구조 변경 유형의 문서 표절을 검출할 수 있다.
- [0049] 도 7은 본 발명의 일 실시예에 있어서, 유의어 사전을 이용한 유사구간 탐색 과정을 도시한 순서도이다.
- [0050] 유의어 사전 기반의 유사구간 탐색장치는 문서 내 유사구간의 탐색을 위해 전처리 단계(710), 필터링 단계(720~730), 유사구간 탐색 단계(740)의 총 3단계로 이루어진 검사를 실시한다.
- [0051] 먼저, 전처리 단계(710)에서는 문서의 유사구간을 탐색하기 위해서 사전에 등록되는 원본문서를 대상으로 텍스트를 문장 단위와 단어 단위로 각각 분할한다. 분할된 단어들은 불용어(stop-word)를 제거하는 과정을 거치고 난 후 유의어 사전을 통하여 검색된 대표 유의어와 함께 문장 내 색인어의 위치, 그리고 문장 자체와 함께 데이터베이스에 저장된다. 예를 들어, 도 8을 참조하면 '사람은 누구나 자기를 알아주는 사람을 위해 헌신한다.'와 같은 문장으로 이루어진 문서가 원본문서로 입력되면 상기한 전처리 과정을 거쳐 도 8의 표와 같은 형태로 데이터베이스에 저장된다. 유사구간 탐색을 위해 사용자가 검사문서를 입력하는 경우에도 원본문서를 처리하는 전처리 과정과 동일한 전처리 과정을 거친다.
- [0052] 이렇게 구성된 검사문서 정보를 이용해 데이터베이스 내의 원본문서 정보를 대상으로 필터링 단계(720~730)를 진행한다. 먼저, 전처리 단계(710)가 끝나면 자카드 계수를 이용한 제1 필터링 단계(720)를 수행한다. 도 9를 참조하면, 제1 필터링 단계(720)에서는 원본문서와 검사문서에 대하여 각각의 색인어를 해당 단어의 대표 유의어로 대체하여 저장하는 벡터 A와 벡터 B를 생성한다. 그 후 생성된 벡터를 상호 비교하여 동일한 색인어의 개수를 계산한다. 이를 이용하여 자카드 계수를 계산하고 그 결과가 일정 기준(예컨대, 25%)을 초과하는 문서들을 다음 필터링 단계인 벡터 공간 모델 및 코사인 거리를 이용한 제2 필터링 단계(730)로 넘겨준다. 다음으로, 제2 필터링 단계(730)에서는 도 10에 도시한 바와 같이 검사문서 내의 색인어의 대표 유의어 정보 및 문서 내 출현빈도 정보를 이용해 벡터를 구성한다. 데이터베이스에 저장된 원본문서 중 제1 필터링 단계(720)에서 선별된 후보 원본문서에 대하여 동일한 방법으로 벡터를 구성한 뒤 두 벡터의 차원을 동기화 하여 상호 간의 코사인 유사도(cosine similarity)를 계산한다. 계산된 유사도를 이용해 후보 원본문서 중에서 일정 기준 이상 유사한 문서를 선별한다.
- [0053] 마지막으로, 제2 필터링 단계(730)에서 선별된 후보 원본문서의 문장들과 사용자가 입력한 검사문서의 문장들 간에 유클리디안 거리 알고리즘을 이용한 유사도 검사로 유사구간을 찾아내는 유사구간 탐색 단계(740)를 수행한다.

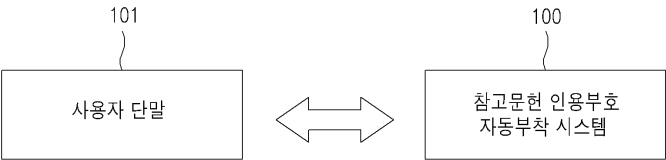
- [0054] 따라서, 본 발명에서는 원본문서와 검사문서로부터 형태소 분석 및 불용어 제거 등의 전처리 과정을 거쳐서 색인어 집합을 추출하고 대표 유의어와 함께 데이터베이스에 저장하여 원본문서에 대한 복제 및 축약 형태의 표절 유형은 물론, 의역 및 구조 변경의 표절 유형까지 검출할 수 있다.
- [0055] 상기한 참고문헌 인용부호 자동부착 방법은 도 1 내지 도 10을 통해 설명한 상세 내용을 바탕으로 보다 단축된 동작들 또는 추가의 동작들을 포함할 수 있다. 또한, 둘 이상의 동작이 조합될 수 있고, 동작들의 순서나 위치가 변경될 수 있다.
- [0056] 본 발명의 실시예에 따른 방법들은 다양한 컴퓨터 시스템을 통하여 수행될 수 있는 프로그램 명령(instruction) 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 또한, 본 실시예에 따른 프로그램은 PC 기반의 프로그램 또는 모바일 단말 전용의 어플리케이션으로 구성될 수 있다.
- [0057] 이와 같이, 본 발명의 실시예에 따르면, 기술문서에 참고문헌에 대한 인용부호를 자동으로 부착함으로써 기술문서의 표절의혹 해소를 통한 표절 문제의 사회적 이슈화를 사전에 방지할 수 있으며, 작성 중인 기술문서 초안에 대하여 유사 문장 검색을 통해 선제적으로 표절의혹 해소를 위한 서비스를 제공함으로써 표절의혹으로부터 자유로운 기술문서 작성 환경을 서비스 할 수 있다.
- [0058] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 컨트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPGA(field programmable gate array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다. 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다. 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 컨트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.
- [0059] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다. 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embodiment)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.
- [0060] 실시예에 따른 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 실시예를 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. 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0061] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될

수 있다.

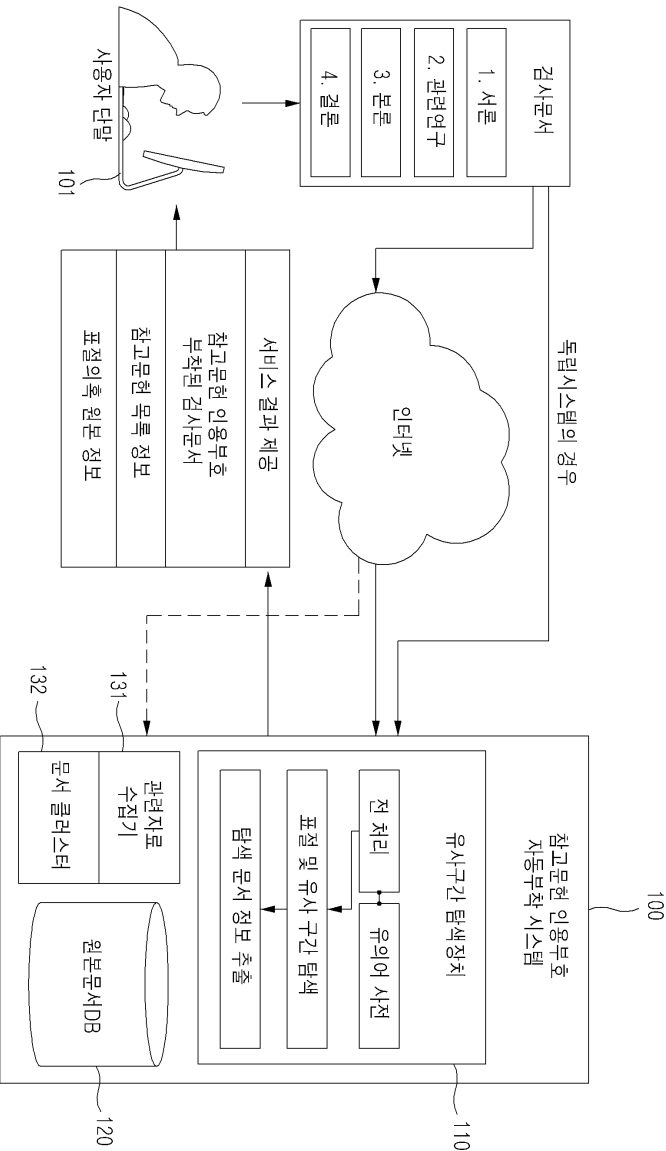
[0062] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

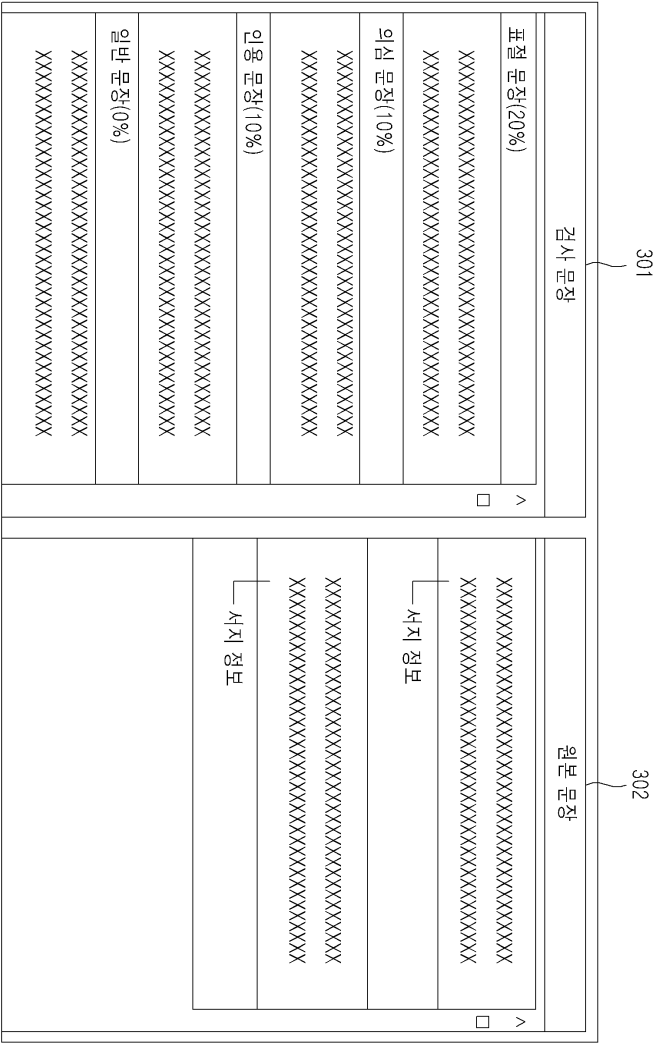
도면1



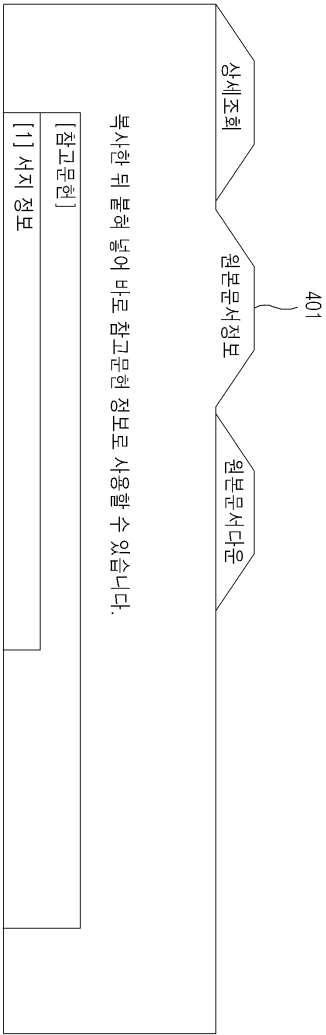
도면2



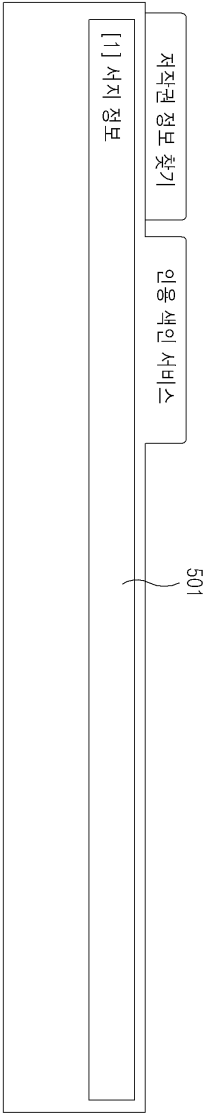
도면3



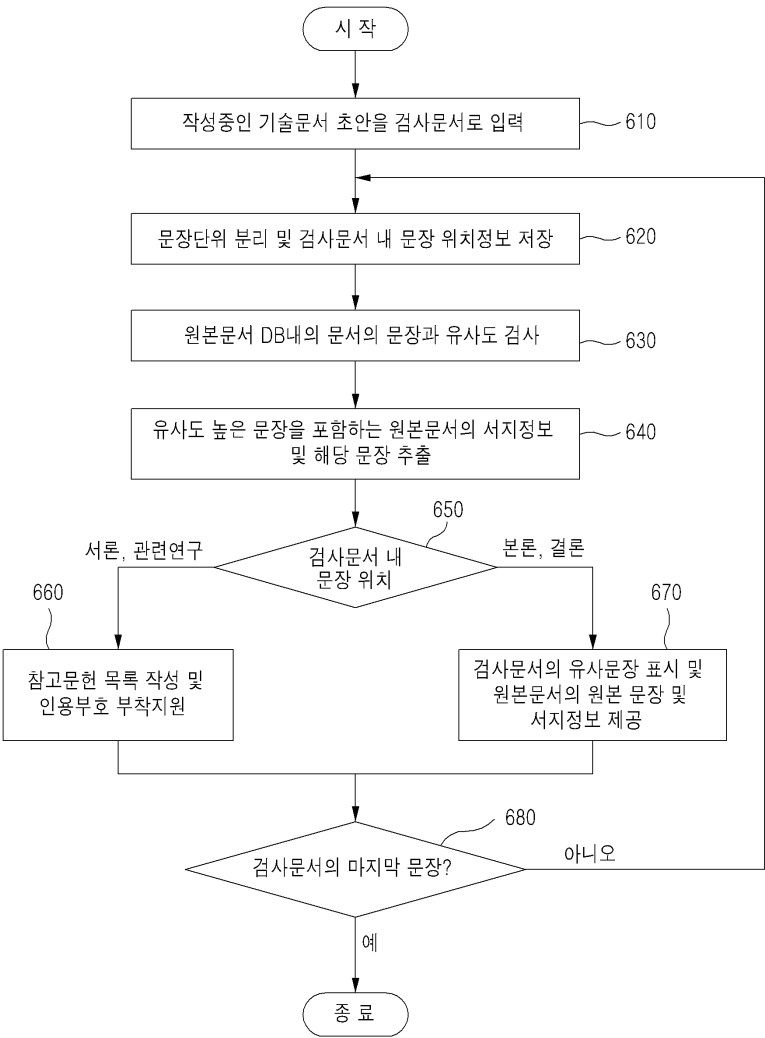
도면4



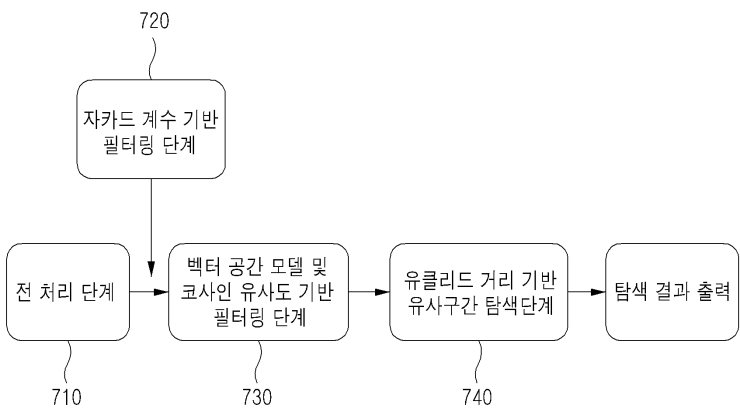
도면5



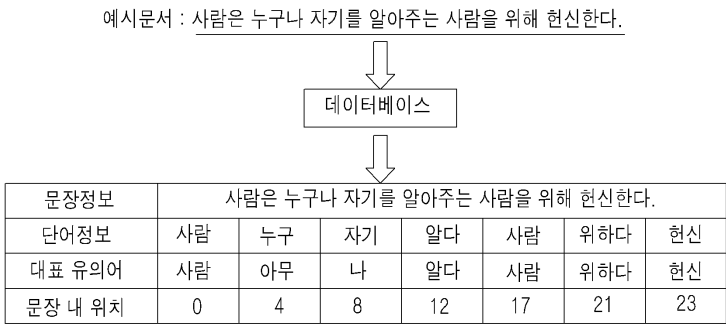
도면6



도면7



도면8



도면9

Algorithm	Jaccard Coefficient
1. 원본문서의 색인어를 저장하는 벡터 A를 만든다. 2. 검사문서의 색인어를 저장하는 벡터 B를 만든다. 3. 벡터 A와 B의 색인어들을 비교한다. 4. 자카드 계수를 계산하고 표절판정기준(현재 기준 25%) 이상의 유사문서를 선별한다. 5. 선별된 문서들을 벡터공간모델 기반의 다음 필터링 단계로 넘긴다.	

도면10

Algorithm	VectorSpaceModel
Procedure	
1. 원본문서와 검사문서 각각의 색인어와 그 빈도를 저장한 벡터 A(원본 문서)와 B(검사문서)를 생성한다. 2. 벡터 A와 B의 색인어를 비교한다. 3. 벡터 A와 B의 색인어 중 서로에게 없는 색인어의 빈도를 0으로 하여 벡터의 차원을 동기화 시킨다. 4. 벡터 A와 B를 각각 정규화 하여 그 크기를 1로 만든다. 5. 4. 에서 정규화 된 벡터 A와 B를 이용하여 코사인 유사도를 계산한다. 6. 5. 에서 계산한 코사인 유사도를 이용하여 문서 간 유사도를 평가한다.	