



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0019637
(43) 공개일자 2019년02월27일

(51) 국제특허분류(Int. Cl.)
G06F 16/00 (2019.01) G06F 17/27 (2006.01)
(52) CPC특허분류
G06F 16/31 (2019.01)
G06F 16/93 (2019.01)
(21) 출원번호 10-2017-0104824
(22) 출원일자 2017년08월18일
심사청구일자 2017년08월18일

(71) 출원인
경희대학교 산학협력단
경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)
(72) 발명자
권오병
서울특별시 강남구 남부순환로 3032 미도아파트 107동 1001호(대치동, 미도아파트)
서한솔
경기도 용인시 기흥구 예현로 15, 101동 1203호(서천동, 서그내마을에스케이아파트)
(74) 대리인
서재승

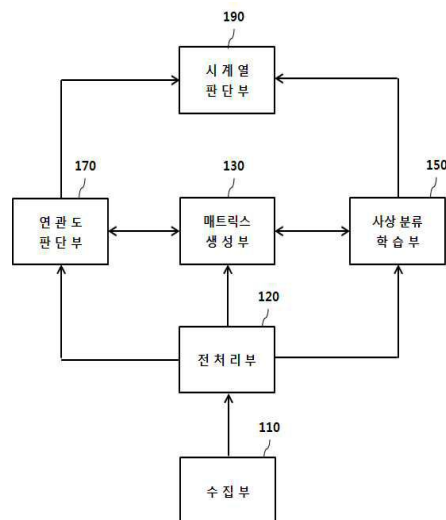
전체 청구항 수 : 총 16 항

(54) 발명의 명칭 인문 사상의 분류 방법

(57) 요약

본 발명은 인문 사상의 분류 방법에 관한 것으로, 보다 구체적으로 종래 철학 사상의 텍스트로부터 철학 사상을 분류하기 위한 사상 분류 알고리즘을 생성하며 신규 철학 사상의 텍스트가 수집되는 경우 사상 분류 알고리즘을 통해 신규 철학 사상의 텍스트가 가지는 철학 사상을 자동 분류할 수 있으며, 신규 철학 사상 텍스트와 종래 철학 사상 사이의 관계를 시계열적 연관도로 판단함으로써 신규 철학 사상 텍스트가 영향을 받은 과거 철학 사상과 신규 철학 사상 텍스트가 영향을 준 미래 철학 사상을 판단할 수 있는 인문 사상의 분류 방법에 관한 것이다.

대표도 - 도2



(52) CPC특허분류

G06F 17/2755 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 2017-0-01122

부처명 미래창조과학부

연구관리전문기관 정보통신기술진흥센터

연구사업명 ICT유망기술개발지원(ICT융합핵심기술개발지원)

연구과제명 지능형 잇츠마이스토리 서비스 제공을 위한 언어 사용패턴 분석 기반 개인 프로파일링 및
개인 맞춤형 챗봇 기술 개발

기 여 율 1/1

주관기관 경희대학교 산학협력단

연구기간 2017.05.01 ~ 2018.12.31

명세서

청구범위

청구항 1

사상 분류 장치에서 인문 사상을 분류하는 방법에 있어서,

철학 사상 텍스트에서 명사를 추출하고, 추출한 명사별 가중치로 이루어진 텍스트-명사 매트릭스(Document-Term Matrix)를 생성하는 단계;

상기 텍스트-명사 매트릭스로부터 학습 데이터와 테스트 데이터를 구분하여 생성하고, 상기 학습 데이터와 상기 테스트 데이터로부터 철학 사상을 분류하는데 이용되는 사상 분류 알고리즘을 생성하는 단계;

신규 철학 사상 텍스트를 전처리하여 추출한 명사의 가중치를 상기 사상 분류 알고리즘에 적용하여 상기 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계; 및

상기 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 상기 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 2

제 1 항에 있어서, 상기 텍스트-명사 매트릭스를 생성하는 단계는

철학 사상 텍스트를 형태소 분석기에 입력하여 상기 철학 사상 텍스트를 구성하는 명사를 추출하는 단계;

추출한 명사에 대한 가중치를 계산하는 단계; 및

상기 철학 사상 텍스트의 철학자와 추출한 명사에 대한 가중치로 이루어진 텍스트-명사 매트릭스를 생성하는 단계를 포함하는 특징으로 하는 인문 사상의 분류 방법.

청구항 3

제 2 항에 있어서, 상기 텍스트-명사 매트릭스를 생성하는 단계는

상기 철학 사상 텍스트를 구성하는 명사 중 상기 철학 사상 텍스트를 수집하는데 사용된 키워드를 제외하고 명사를 추출하는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 4

제 3 항에 있어서, 상기 추출한 명사에 대한 가중치는

단위 수의 철학 사상 텍스트 중 상기 명사가 존재하는 철학 사상 텍스트의 수의 비율과 상기 철학 사상 텍스트에 상기 추출한 명사가 나타난 빈도로부터 계산되는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 5

제 4 항에 있어서, 상기 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계

상기 신규 철학 사상 텍스트를 형태소 분석기에 입력하여 상기 신규 철학 사상 텍스트를 구성하는 명사를 추출하는 단계;

상기 신규 철학 사상 텍스트로부터 추출한 명사에 대한 가중치를 계산하는 단계; 및

추출한 명사에 대한 가중치를 상기 사상 분류 알고리즘에 적용하여 상기 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계를 포함하는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 6

제 5 항에 있어서, 상기 시계열적 연관도를 생성하는 단계는

상기 텍스트-명사 매트릭스(Document-Term Matrix)를 구성하는 명사별 가중치를 철학자 식별자로 분류하는

단계;

상기 철학자 식별자에 따라 분류한 상기 텍스트-명사 매트릭스의 명사 가중치와 신규 철학 사상 텍스트로부터 추출한 명사 가중치 사이의 거리를 계산하는 단계;

상기 신규 철학 사상 텍스트의 작성 시기를 판단하는 단계; 및

상기 계산한 거리와 상기 신규 철학 사상 텍스트의 작성 시기로부터 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 상기 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 7

제 6 항에 있어서, 상기 거리를 계산하는데 이용되는 상기 텍스트-명사 매트릭스의 명사 가중치는 상기 철학자 식별자에 따라 분류한 철학 사상 텍스트들의 명사 가중치의 평균값인 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 8

제 7 항에 있어서, 상기 신규 철학 사상 텍스트의 작성 시기는

상기 신규 철학 사상 텍스트에서 설정 간격으로 추출한 년도 데이터 중 가장 많은 빈도수가 높은 년도 데이터들의 평균값으로 계산되는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 9

신규 철학 사상 텍스트를 형태소 분석기에 입력하여 상기 신규 철학 사상 텍스트를 구성하는 명사를 추출하는 단계;

상기 신규 철학 사상 텍스트로부터 추출한 명사에 대한 가중치를 계산하는 단계;

추출한 명사에 대한 가중치를 사상 분류 알고리즘에 적용하여 상기 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계; 및

상기 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 상기 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 10

제 9 항에 있어서, 상기 사상 분류 알고리즘은

종래 철학 사상 텍스트에서 명사를 추출하고, 추출한 명사별 가중치로 이루어진 텍스트-명사 매트릭스(Document-Term Matrix)를 생성하는 단계;

상기 텍스트-명사 매트릭스로부터 학습 데이터와 테스트 데이터를 구분하여 생성하고, 상기 학습 데이터와 상기 테스트 데이터로부터 철학 사상을 분류하는데 이용되는 사상 분류 알고리즘을 생성하는 단계를 통해 생성되는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 11

제 10 항에 있어서, 상기 텍스트-명사 매트릭스는

상기 종래 철학 사상 텍스트를 형태소 분석기에 입력하여 상기 종래 철학 사상 텍스트를 구성하는 명사를 추출하는 단계;

추출한 명사에 대한 가중치를 계산하는 단계; 및

상기 종래 철학 사상 텍스트의 철학자와 추출한 명사에 대한 가중치로 이루어진 텍스트-명사 매트릭스를 생성하는 단계를 통해 생성되는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 12

제 11 항에 있어서, 상기 텍스트-명사 매트릭스는

상기 종래 철학 사상 텍스트를 구성하는 명사 중 상기 종래 철학 사상 텍스트를 수집하는데 사용된 키워드를 제외한 명사를 추출하여 생성되는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 13

제 12 항에 있어서, 상기 추출한 명사에 대한 가중치는

단위 수의 종래 철학 사상 텍스트 중 상기 명사가 존재하는 종래 철학 사상 텍스트의 수의 비율과 상기 종래 철학 사상 텍스트에 상기 추출한 명사가 나타난 빈도로부터 계산되는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 14

제 11 항에 있어서, 상기 시계열적 연관도를 생성하는 단계는

상기 텍스트-명사 매트릭스(Document-Term Matrix)를 구성하는 명사별 가중치를 철학자 식별자로 분류하는 단계;

상기 철학자 식별자에 따라 분류한 상기 텍스트-명사 매트릭스의 명사 가중치와 신규 철학 사상 텍스트로부터 추출한 명사 가중치 사이의 거리를 계산하는 단계;

상기 신규 철학 사상 텍스트의 작성 시기를 판단하는 단계; 및

상기 계산한 거리와 상기 신규 철학 사상 텍스트의 작성 시기로부터 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 상기 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 15

제 14 항에 있어서, 상기 거리를 계산하는데 이용되는 상기 텍스트-명사 매트릭스의 명사 가중치는 상기 철학자 식별자에 따라 분류한 종래 철학 사상 텍스트들의 명사 가중치의 평균값인 것을 특징으로 하는 인문 사상의 분류 방법.

청구항 16

제 15 항에 있어서, 상기 신규 철학 사상 텍스트의 작성 시기는

상기 신규 철학 사상 텍스트에서 설정 간격으로 추출한 년도 데이터 중 가장 많은 빈도수가 높은 년도 데이터들의 평균값으로 계산되는 것을 특징으로 하는 인문 사상의 분류 방법.

발명의 설명

기술 분야

[0001] 본 발명은 인문 사상의 분류 방법에 관한 것으로, 보다 구체적으로 종래 철학 사상의 텍스트로부터 철학 사상을 분류하기 위한 사상 분류 알고리즘을 생성하며 신규 철학 문언이 수집되는 경우 사상 분류 알고리즘을 통해 신규 철학 문언이 가지는 철학 사상을 자동 분류할 수 있고, 신규 철학 문언과 종래 철학 사상 사이의 관계를 시계열적 연관도로 판단함으로써 신규 철학 문언이 영향을 받은 과거 철학 사상과 신규 철학 사상 문언이 영향을 준 미래 철학 사상을 판단할 수 있는 인문 사상의 분류 방법에 관한 것이다.

배경 기술

[0002] 최근 빅데이터 분석 기법의 확산으로 디지털 인문학(Digital Humanities)이 주목 받고 있다. 디지털 인문학은 전통적인 인문학의 연구 주제를 계승하면서도 정보기술의 도움을 받아 새로운 방식으로 인문학 연구를 수행하고 인문 지식을 사회 전반으로 효과적으로 확산시키려는 목적을 가진다. 이에 따라 국가뿐만 아니라 Oxford, Stanford 등의 대학 연구기관과 Google 등의 기업들도 디지털 인문학 프로젝트를 진행 중이다. 인문학 분야에서 빅데이터를 활용하는 것은 인문학적 사유와 고찰을 대체하려는 것이 아니라 인문학을 더욱 풍성하게 하려는 목적을 가진다.

[0003] 이에 따라 문학, 지리학 등 다양한 학문 분야에서 디지털 인문학 연구가 진행되고 있으나 철학 분야에서의 디지털 인문학적 접근은 그 유용성에 대한 기대에도 불구하고 아직은 부족한 실정이다.

[0004] 최근 디지털 인문학 연구는 빅데이터 기술을 활용하기 시작했다. 먼저 가시화 도구는 인문학 문헌들의 구조와 관계성을 직관적으로 파악하기에 적합하다. 예를 들어 Data Sprint는 서적의 간략한 서지 정보 즉 저자나 책 제목 등을 활용하여 social-network analysis를 가시화하는 Amazon API를 활용하고 있다. 둘째로, 자료 검색도 빅데이터의 도움을 받을 수 있다. 방대한 문헌에서 정해진 짧은 시간 안에 검색 기능을 수행하는 연구를 진행 중이다.

[0005] 철학 연구에서 철학적 사상 간의 연결성 분석은 중요한 연구이며 또한 특정인의 사상이 기존 철학적 사상과 어떤 연관성이 있음을 분석하는 것도 중요한 연구이다. 이러한 연구를 위해서는 한 문헌이 지니는 철학적 성향과 특징을 파악하고 그와 관련된 철학 문헌들을 추출, 선별하고 목록화 하는 작업이 선행되어야 한다. 그러나 이러한 작업은 매우 광범위하고 많은 시간을 필요로 하는 과정이므로 연구자가 직접 수행하기보다 기존 연구 결과에 의존하거나 연구의 범위 자체가 축소되기도 하였다. 더구나 분석할 문헌의 양이 방대할수록 사람이 문헌을 읽고 철학 사상을 분류하는 작업은 많은 노력이 소요된다. 그러나 이러한 작업을 지원할 수 있는 정보기술 {information technology}에 대해서는 제안된 바가 드물다.

[0006] 본원발명의 배경기술로 한국공개특허 제10-2012-0109943호가 있다.

발명의 내용

해결하려는 과제

[0007] 본 발명이 이루고자 하는 목적은 철학 연구 분야에서 디지털 인문학을 적용하여 특정 문언이 가지는 사상을 자동으로 분류하여 판단할 수 있는 인문 사상의 분류 방법을 제공하는 것이다.

[0008] 본 발명이 이루고자 하는 다른 목적은 종래 철학 사상의 텍스트로부터 철학 사상을 분류하기 위한 사상 분류 알고리즘을 생성하며, 신규 철학 문언이 수집되는 경우 사상 분류 알고리즘을 통해 신규 철학 문언이 가지는 철학 사상을 자동 분류할 수 있는 인문 사상 분류 방법을 제공하는 것이다.

[0009] 본 발명이 이루고자 하는 또 다른 목적은 신규 철학 문언과 종래 철학 사상 사이의 연관도를 계산하여 신규 철학 문언과 관련이 있는 종래 철학 사상 사이의 관련성을 판단할 수 있는 인문 사상 분류 방법을 제공하는 것이다.

[0010] 본 발명이 이루고자 하는 또 다른 목적은 신규 철학 문언과 종래 철학 사상 사이의 관계를 시계열적 연관도로 판단하여 신규 철학 문언이 영향을 받은 과거 철학 사상과 신규 철학 문언이 영향을 준 미래 철학 사상을 판단할 수 있는 인문 사상 분류 방법을 제공하는 것이다.

과제의 해결 수단

[0011] 본 발명의 목적을 달성하기 위하여 본 발명의 일 실시예에 따른 인문 사상을 분류하는 방법은 철학 사상 텍스트에서 명사를 추출하고 추출한 명사별 가중치로 이루어진 텍스트-명사 매트릭스(Document-Term Matrix)를 생성하는 단계와, 텍스트-명사 매트릭스로부터 학습 데이터와 테스트 데이터를 구분하여 생성하고 학습 데이터와 테스트 데이터로부터 철학 사상을 분류하는데 이용되는 사상 분류 알고리즘을 생성하는 단계와, 신규 철학 사상 텍스트를 전처리하여 추출한 명사의 가중치를 사상 분류 알고리즘에 적용하여 상기 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계와, 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 한다.

[0012] 바람직하게, 텍스트-명사 매트릭스를 생성하는 단계는 철학 사상 텍스트를 형태소 분석기에 입력하여 철학 사상 텍스트를 구성하는 명사를 추출하는 단계와, 추출한 명사에 대한 가중치를 계산하는 단계와, 철학 사상 텍스트의 철학자와 추출한 명사에 대한 가중치로 이루어진 텍스트-명사 매트릭스를 생성하는 단계를 포함하는 특징으로 한다.

[0013] 바람직하게, 텍스트-명사 매트릭스를 생성하는 단계는 철학 사상 텍스트를 구성하는 명사 중 철학 사상 텍스트를 수집하는데 사용된 키워드를 제외하고 명사를 추출하는 것을 특징으로 한다.

[0014] 여기서 추출한 명사에 대한 가중치는 단위 수의 철학 사상 텍스트 중 명사가 존재하는 철학 사상 텍스트의 수의 비율과 철학 사상 텍스트에 추출한 명사가 나타난 빈도로부터 계산되는 것을 특징으로 한다.

[0015] 바람직하게 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계는 신규 철학 사상 텍스트를 형태소 분석기에 입력하여 신규 철학 사상 텍스트를 구성하는 명사를 추출하는 단계와, 신규 철학 사상 텍스트로부터 추출

한 명사에 대한 가중치를 계산하는 단계와, 추출한 명사에 대한 가중치를 사상 분류 알고리즘에 적용하여 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계를 포함하는 것을 특징으로 한다.

[0016] 바람직하게, 시계열적 연관도를 생성하는 단계는 텍스트-명사 매트릭스(Document-Term Matrix)를 구성하는 명사 별 가중치를 철학자 식별자로 분류하는 단계와, 철학자 식별자에 따라 분류한 텍스트-명사 매트릭스의 명사 가중치와 신규 철학 사상 텍스트로부터 추출한 명사 가중치 사이의 거리를 계산하는 단계와, 신규 철학 사상 텍스트의 작성 시기를 판단하는 단계와, 계산한 거리와 신규 철학 사상 텍스트의 작성 시기로부터 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 한다.

[0017] 여기서 거리를 계산하는데 이용되는 텍스트-명사 매트릭스의 명사 가중치는 철학자 식별자에 따라 분류한 철학 사상 텍스트들의 명사 가중치의 평균값인 것을 특징으로 한다.

[0018] 여기서 신규 철학 사상 텍스트의 작성 시기는 신규 철학 사상 텍스트에서 설정 간격으로 추출한 년도 데이터 중 가장 많은 빈도수가 높은 년도 데이터들의 평균값으로 계산되는 것을 특징으로 한다.

[0019] 한편, 본 발명의 목적을 달성하기 위하여 본 발명의 다른 실시예에 따른 인문 사상을 분류하는 방법은 신규 철학 사상 텍스트를 형태소 분석기에 입력하여 신규 철학 사상 텍스트를 구성하는 명사를 추출하는 단계와, 신규 철학 사상 텍스트로부터 추출한 명사에 대한 가중치를 계산하는 단계와, 추출한 명사에 대한 가중치를 사상 분류 알고리즘에 적용하여 신규 철학 사상 텍스트가 속한 철학 사상을 판단하는 단계와, 신규 철학 사상 텍스트에 영향을 미친 과거 철학 사상과 신규 철학 텍스트에 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 생성하는 단계를 포함하는 것을 특징으로 한다.

[0020] 여기서 사상 분류 알고리즘은 종래 철학 사상 텍스트에서 명사를 추출하고, 추출한 명사별 가중치로 이루어진 텍스트-명사 매트릭스(Document-Term Matrix)를 생성하는 단계와, 텍스트-명사 매트릭스로부터 학습 데이터와 테스트 데이터를 구분하여 생성하고 학습 데이터와 테스트 데이터로부터 철학 사상을 분류하는데 이용되는 사상 분류 알고리즘을 생성하는 단계를 통해 생성되는 것을 특징으로 한다.

[0021] 여기서 텍스트-명사 매트릭스는 종래 철학 사상 텍스트를 형태소 분석기에 입력하여 종래 철학 사상 텍스트를 구성하는 명사를 추출하는 단계와, 추출한 명사에 대한 가중치를 계산하는 단계와, 종래 철학 사상 텍스트의 철학자와 추출한 명사에 대한 가중치로 이루어진 텍스트-명사 매트릭스를 생성하는 단계를 통해 생성되는 것을 특징으로 한다.

발명의 효과

[0022] 본 발명에 따른 인문 사상의 분류 방법은 다음과 같은 효과를 가진다.

[0023] 첫째, 본 발명에 따른 인문 사상의 분류 방법은 종래 철학 사상의 텍스트로부터 철학 사상을 분류하기 위한 사상 분류 알고리즘을 생성하며 신규 철학 사상의 텍스트가 수집되는 경우 사상 분류 알고리즘을 통해 신규 철학 사상의 텍스트가 가지는 철학 사상을 자동 분류함으로써, 철학 사상을 분류하는데 소요되는 시간과 노력을 줄일 수 있다.

[0024] 둘째, 본 발명에 따른 인문 사상의 분류 방법은 신규 철학 사상 텍스트와 종래 철학 사상 사이의 연관도를 계산함으로써, 신규 철학 사상 텍스트와 관련이 있는 종래 철학 사상 사이의 관련성을 판단할 수 있다.

[0025] 셋째, 본 발명에 따른 인문 사상의 분류 방법은 신규 철학 사상 텍스트와 종래 철학 사상 사이의 관계를 시계열적 연관도로 판단함으로써, 신규 철학 사상 텍스트가 영향을 받은 과거 철학 사상과 신규 철학 사상 텍스트가 영향을 준 미래 철학 사상을 판단할 수 있다.

도면의 간단한 설명

[0026] 도 1은 본 발명에 따른 인문 사상의 분류 시스템을 설명하기 위한 기능 블록도이다.

도 2는 본 발명의 일 실시예에 따른 사상 분류 장치를 설명하기 위한 기능 블록도이다.

도 3은 본 발명에 따른 전처리부의 일 예를 설명하기 위한 기능 블록도이다.

도 4는 본 발명에 따른 연관도 판단부의 일 예를 설명하기 위한 기능 블록도이다.

도 5는 본 발명의 일 실시예에 따른 인문 사상의 분류 방법을 설명하기 위한 흐름도이다.

도 6은 텍스트-명사 매트릭스를 생성하는 단계의 일 예를 설명하기 위한 흐름도이다.

도 7은 데이터베이스부에서 수집한 철학 문언을 형태소 분석하여 추출한 명사의 일 예를 도시하고 있다.

도 8은 텍스트-명사 매트릭스의 일 예를 도시하고 있다.

도 9는 코사인 유사성 방식으로 계산한, 신규 철학 문언이 가지는 사상과 텍스트-명사 매트릭스를 구성하는 철학 사상과의 연관도의 일 예를 도시하고 있다.

도 10은 신규 철학 사상의 시계열적 연관성의 일 예를 설명하기 위한 도면이다.

도 11은 신규 철학 문언의 작성 시기를 판단하는 방식을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0027] 본 발명에서 사용되는 기술적 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아님을 유의해야 한다. 또한, 본 발명에서 사용되는 기술적 용어는 본 발명에서 특별히 다른 의미로 정의되지 않는 한, 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 의미로 해석되어야 하며, 과도하게 포괄적인 의미로 해석되거나, 과도하게 축소된 의미로 해석되지 않아야 한다. 또한, 본 발명에서 사용되는 기술적인 용어가 본 발명의 사상을 정확하게 표현하지 못하는 잘못된 기술적 용어일 때에는, 당업자가 올바르게 이해할 수 있는 기술적 용어로 대체되어 이해되어야 할 것이다.
- [0028] 또한, 본 발명에서 사용되는 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한 복수의 표현을 포함한다. 본 발명에서, "구성된다" 또는 "포함한다" 등의 용어는 발명에 기재된 여러 구성 요소들, 또는 여러 단계를 반드시 모두 포함하는 것으로 해석되지 않아야 하며, 그 중 일부 구성 요소들 또는 일부 단계들은 포함되지 않을 수도 있고, 또는 추가적인 구성 요소 또는 단계들을 더 포함할 수 있는 것으로 해석되어야 한다.
- [0029] 또한, 첨부된 도면은 본 발명의 사상을 쉽게 이해할 수 있도록 하기 위한 것일 뿐, 첨부된 도면에 의해 본 발명의 사상이 제한되는 것으로 해석되어서는 아니 됨을 유의해야 한다.
- [0030] 이하 첨부한 도면을 참고로 본 발명에 따른 인문 사상의 분류 방법에 대해 보다 구체적으로 살펴본다.
- [0031] 도 1은 본 발명에 따른 인문 사상의 분류 시스템을 설명하기 위한 기능 블록도이다.
- [0032] 도 1을 참고로 보다 구체적으로 살펴보면, 네트워크(30)에는 데이터베이스부(10)와 인문 사상 분류 장치(50)가 접속되어 있다.
- [0033] 데이터베이스부(30)에는 다양한 종류의 인문 문언이 저장되어 있는데, 예를 들어 역사, 철학, 문학 등과 관련된 인문 문언이 저장되어 있다.
- [0034] 인문 사상 분류 장치(50)는 네트워크(30)를 통해 데이터베이스부(10)로부터 인문 문언을 수신하고 수신한 인문 문언으로부터 인문 문언을 분류하기 위한 사상 분류 알고리즘을 학습 생성하는데, 데이터베이스부(10)로부터 신규 인문 문언을 수집하거나 또는 네트워크(30)를 통해 다른 소스로 신규 인문 문언을 수집하는 경우 사상 분류 알고리즘에 기초하여 신규 인문 문언이 가지는 사상을 분류하거나 신규 인문 문언이 가지는 사상과 다른 인문 사상과의 연관도를 판단하거나, 신규 인문 문언이 영향을 받은 과거 인문 사상 또는 신규 인문 문언이 영향을 준 미래 인문 사상과의 연관도를 판단한다.
- [0035] 이하에서는 인문 문언의 일 예로 철학 문언을 사용하여 본 발명을 설명하도록 하나, 철학 문언 이외에 다양한 종류의 인문 문언이 사용될 수 있으며 이는 본 발명의 범위에 속한다.
- [0036] 도 2는 본 발명의 일 실시예에 따른 사상 분류 장치를 설명하기 위한 기능 블록도이다.
- [0037] 도 2를 참고로 보다 구체적으로 살펴보면, 수집부(110)는 입력된 키워드에 기초하여 데이터베이스부 또는 다른 소스를 통해 철학 문언을 수집한다. 바람직하게, 수집부(110)는 네트워크를 통해 데이터를 송수신하기 위한 통신 모듈을 구비하며, 수집부(110)는 네트워크를 통해 입력된 키워드에 해당하는 철학 문언을 수집한다.
- [0038] 전처리부(120)는 철학 문언의 텍스트를 형태소 분석기에 입력하여 철학 문언의 텍스트에서 명사를 추출하고 추출한 명사별 가중치를 계산하며, 매트릭스 생성부(130)는 이미 알고 있는 철학 문언의 철학 사상, 철학자 등과 각 철학 문언 텍스트에서 추출한 명사의 가중치로 이루어진 텍스트-명사 매트릭스(Document-term matrix)를 생성한다.

- [0039] 사상 분류 학습부(150)는 매트릭스 생성부(130)에서 생성된 텍스트-명사 매트릭스로부터 트레이닝 데이터와 테스트 데이터를 추출하고 추출한 트레이닝 데이터와 테스트 데이터로부터 철학 사상을 분류하는데 이용되는 사상 분류 알고리즘을 생성한다. 여기서 사상 분류 학습부(150)에서 사용되는 텍스트-명사 매트릭스는 철학 사상을 알고 있는 텍스트와 각 테스트에서 추출한 명사의 가중치로 이루어지며, 트레이닝 데이터와 테스트 데이터는 텍스트-명사 매트릭스에서 임의적으로 구분하여 추출할 수 있다.
- [0040] 한편, 수집부(110)를 통해 신규 철학 문언이 수집되는 경우, 전처리부(120)는 신규 철학 문언의 텍스트를 형태소 분석기에 입력하여 신규 철학 문언의 텍스트에서 명사를 추출하고 추출한 명사의 가중치를 계산한다. 신규 철학 문언의 명사별 가중치는 사상 분류 학습부(150)로 제공되는데, 사상 분류 학습부(150)는 신규 철학 문언의 명사별 가중치를 생성한 사상 분류 알고리즘에 적용하여 신규 철학 문언이 속하는 철학 사상을 판단한다.
- [0041] 한편, 전처리부(120)는 신규 철학 문언의 명사별 가중치를 연관도 판단부(170)로 제공하는데, 연관도 판단부(170)는 매트릭스 생성부(130)에서 생성된 철학 사상별 명사의 가중치와 신규 철학 문언의 명사 가중치 사이의 거리를 계산하여 신규 철학 문언이 가지는 사상과 다른 철학 사상 사이의 연관도를 계산한다.
- [0042] 시계열 판단부(190)는 사상 분류 학습부(150)에서 분류한 신규 철학 문언이 속하는 철학 사상과 연관도 판단부(170)에서 판단한 다른 철학 사상과의 연관도에 기초하여 신규 철학 문언과 다른 철학 사상 사이의 시계열적 연관도를 판단한다. 즉, 시계열 판단부(190)는 신규 철학 문언이 작성된 시기에 기초하여 신규 철학 문언이 영향을 받은 과거 철학 사상 및 신규 철학 문언이 영향을 준 미래 철학 사상 사이의 시계열적 연관도를 판단한다.
- [0043] 도 3은 본 발명에 따른 전처리부의 일 예를 설명하기 위한 기능 블록도이다.
- [0044] 도 3을 참고로 보다 구체적으로 살펴보면, 형태소 분석부(121)는 수집한 철학 문언의 텍스트로부터 명사를 추출하는데, 여기서 형태소는 뜻을 가진 가장 작은 말의 단위로 형태소 분석 방식은 규칙기반 형태 분석 방식과 음절단위 형태 분석 방식 등 다양한 형태소 분석 알고리즘이 사용되고 있으며, 이에 대한 자세한 설명은 생략한다.
- [0045] 형태소 분석부(121)를 통해 철학 문언의 텍스트로부터 추출한 명사에서 철학 문언을 수집하는데 사용된 키워드를 제거한다. 추출한 명사에서 철학 문언을 수집하는데 사용된 키워드를 제거함으로써, 수집한 철학 문언으로부터 생성되는 사상 분류 알고리즘이 추출한 명사에 과적합(overfitting)되는 것을 방지할 수 있다.
- [0046] 가중치 계산부(125)는 추출한 명사가 철학 문언의 텍스트에서 가지는 가중치를 계산하는데, 추출한 명사의 가중치는 단위 수의 철학 문언 중 해당 명사가 존재하는 철학 문언의 수의 비율, 해당 철학 문언에서 추출한 명사가 나타난 빈도로부터 계산될 수 있다.
- [0047] 도 4는 본 발명에 따른 연관도 판단부의 일 예를 설명하기 위한 기능 블록도이다.
- [0048] 도 4를 참고로 보다 구체적으로 살펴보면, 텍스트-명사 매트릭스에서 명사별 가중치를 철학 사상 식별자별로 분류한다. 여기서 철학 사상 식별자로 철학자 식별자가 사용될 수 있다.
- [0049] 평균값 계산부(173)는 철학 사상 식별자별로 각 명사의 가중치 평균값을 계산하며, 연관도 계산부(175)는 각 명사의 가중치 평균값을 이용하여 철학 사상 식별자별로 각 명사의 거리를 계산하여 각 철학 사상이 가지는 연관도를 계산한다.
- [0050] 도 5는 본 발명의 일 실시예에 따른 인문 사상의 분류 방법을 설명하기 위한 흐름도이다.
- [0051] 도 5를 참고로 보다 구체적으로 살펴보면, 데이터베이스부에 저장되어 있는 철학 문언을 수집하고 수집한 철학 문언의 텍스트에서 명사를 추출하여 추출한 명사의 가중치로 이루어진 텍스트-명사 매트릭스를 생성한다(S110).
- [0052] 생성한 텍스트-명사 매트릭스로부터 트레이닝 데이터와 테스트 데이터를 분류하여 추출하고, 추출한 트레이닝 데이터와 테스트 데이터로부터 사상 분류 알고리즘을 생성한다(S130). 본 발명이 적용되는 분야에 따라 사상 분류 알고리즘을 생성하기 위한 다양한 종류의 학습 알고리즘이 사용될 수 있는데, 결정 트리, Deep Neural Network, k-NN classifier, Multinomial logistic regression, Naive Bayes, Random Forest, SVM 등의 학습 알고리즘을 사용할 수 있으며, 이에 대한 자세한 설명은 생략한다.
- [0053] 신규 철학 문언이 수집되는 경우(S150), 신규 철학 문언을 전처리하여 신규 철학 문언의 텍스트에서 명사를 추출하고 추출한 명사의 가중치를 계산한다(S160). 추출한 명사의 가중치를 생성한 사상 분류 알고리즘에 적용하여 신규 철학 문언이 속하는 철학 사상을 판단하고(S170), 신규 철학 문언의 명사 가중치와 텍스트-명사 매트릭스를 구성하는 철학 사상별 명사 가중치 사이의 거리를 측정하여 신규 철학 문언이 가지는 사상과 텍스트-명사

매트릭스를 구성하는 종래 철학 사상과의 연관도를 계산한다(S180).

[0054] 본 발명이 적용되는 분야에 따라 신규 철학 문언이 가지는 사상과 텍스트-명사 매트릭스를 구성하는 종래 철학 사상과의 연관도는 유클리드 거리(Euclidean distance) 또는 코사인 유사성(Cosine similarity)의 방식을 사용하여 계산될 수 있다. 도 9는 코사인 유사성 방식으로 계산한, 철학 사상 연관도의 일 예로 서로 유사한 철학 사상일수록 1에 가까운 값을 가지며, 서로 상이한 철학 사상일수록 0의 값을 가지게 된다. 코사인 유사성 방식은 2개의 텍스트를 구성하는 단어를 벡터로 표현하며 이들 사이의 내적을 통해 2개의 텍스트 사이의 유사도를 측정하는 방식으로 이에 대한 상세한 설명은 생략한다.

[0055] 신규 철학 문언이 작성된 시기를 신규 철학 문언에 나타나는 년도 데이터에 기초하여 판단하며, 신규 철학 문언이 작성된 시기, 신규 철학 문언이 속하는 철학 사상 및 신규 철학 문언과 텍스트-명사 매트릭스를 구성하는 철학 사상과의 연관도에 기초하여 신규 철학 문언에 영향을 준 과거 철학 사상과 신규 철학 문언이 영향을 준 미래 철학 사상과의 시계열적 연관도를 판단한다(S190).

[0056] 도 6은 텍스트-명사 매트릭스를 생성하는 단계의 일 예를 설명하기 위한 흐름도이다.

[0057] 도 6을 참고로 보다 구체적으로 살펴보면, 데이터베이스부에서 수집한 철학 문언을 형태소 분석하여 철학 문언의 텍스트를 구성하는 명사를 추출한다(S111). 도 7은 데이터베이스부에서 수집한 철학 문언을 형태소 분석하여 추출한 명사의 일 예를 도시하고 있다. 데이터베이스부에서 수집한 철학 문언의 경우 해당 철학 문언의 철학 사상이 또는 철학 사상을 이미 알고 있으며, 철학 문언의 텍스트를 구성하는 명사를 추출한다.

[0058] 추출한 명사에서 해당 철학 문언을 수집하는데 이용한 키워드를 제거한다(S113). 추출한 명사에서 해당 철학 문언을 수집하는데 이용한 키워드를 제거하는 것은 철학 문언에서 추출한 명사를 이용하여 사상 분류 알고리즘을 생성시 과적합의 문제가 발생하는 것을 미연에 방지하기 위한 것으로, 예를 들어 '데카르트'를 키워드로 철학 문언을 수집하는 경우 수집한 철학 문언의 텍스트에서 추출한 명사 중 키워드로 사용한 '데카르트'를 제거하여 사상 분류 알고리즘에서 과적합 문제가 발생하는 것을 방지한다.

[0059] 추출한 각 명사가 해당 철학 문언에서 가지는 가중치를 계산한다(S115). 여기서 가중치는 단위 수의 철학 문언 중 해당 명사가 존재하는 철학 문언의 수의 비율과 해당 철학 문언에 해당 명사가 나타난 빈도로부터 계산되는데, 바람직하게 각 명사(i)의 가중치(w_i)는 아래의 수학식 (1)과 같이 계산될 수 있다.

[0060] [수학식 1]

$$w_i = \left(\frac{n_i}{N} \right)^{-1} s f_i$$

[0061]

[0062] 여기서 w_i 는 명사 i의 가중치이며, N은 단위 수의 철학 문언, n_i 은 단위 수의 철학 문언에서 해당 명사(i)가 존재하는 문언의 수, f_i 는 해당 철학 문언에서 명사(i)가 나타난 빈도를 의미한다.

[0063] 이와 같이 단위 수의 철학 문언 중 해당 명사가 존재하는 철학 문언의 수의 비율에 대한 역수로부터 명사의 가중치를 계산함으로써, 추출한 명사가 해당 철학 문언의 철학 사상과 관련하여 중요한 명사인지를 판단할 수 있다.

[0064] 철학 문언을 구성하는 각 명사의 가중치와 해당 철학 문언의 철학 사상이 또는 철학 사상으로 부터 이루어진 텍스트-명사 매트릭스를 생성한다(S117).

[0065] 도 8은 텍스트-명사 매트릭스의 일 예로 텍스트-명사 매트릭스에는 수집한 철학 문언 텍스트의 각 명사에 대한 가중치와 각 철학 문언의 철학자가 매트릭스 형태로 생성된다.

[0066] 도 10은 신규 철학 사상의 시계열적 연관성의 일 예를 설명하기 위한 도면이다.

[0067] 도 10에 도시되어 있는 바와 같이, 각 철학 문언이 가지는 철학 사상은 사상 분류 알고리즘에 의해 +1과 -1로 서로 구분되어 분류되는데, 예를 들어 +1은 경험주의 철학 사상에 해당하며, -1은 이성주의 철학 사상에 해당하고, 경험주의 철학 사상과 이성주의 철학 사상 모두에 해당하지 않는 경우 0에 위치하게 된다.

[0068] 원 모양은 각 철학 사상을 의미하는데 철학 문언이 가지는 철학 사상과 다른 철학 사상 사이에 연관도가 있는 경우 원은 서로 연결선으로 연결되며, 연관도의 정도에 따라 연결선의 굵기가 상이하게 표시된다.

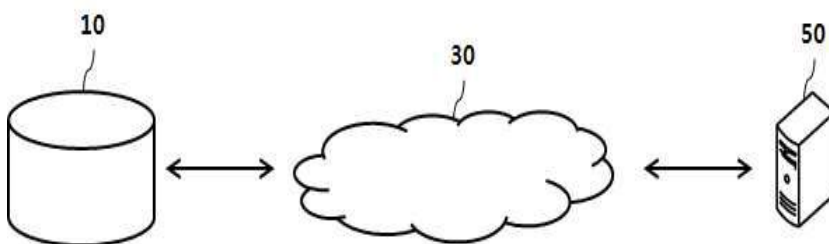
- [0069] 한편, 신규 철학 문언(N)이 수집된 경우, 신규 철학 문언 텍스트에서 추출한 명사의 가중치를 사상 분류 알고리즘에 적용하여 +1과 -1 사이에서 분류하며, 신규 철학 문언 텍스트에서 추출한 명사의 가중치로부터 다른 철학 사상과의 연관도를 계산한다. 신규 철학 문언이 작성된 시기, 신규 철학 문언과 다른 철학 사상 사이의 연관도에 기초하여 신규 철학 문언이 영향을 받은 과거 철학 사상 및 신규 철학 문언이 영향을 준 미래 철학 사상을 표시하여 판단한다. 즉, 신규 철학 문언과 연관도가 있는 다른 철학 사상 중 신규 철학 문언이 작성된 시기의 선후 관계로 서로 영향을 주거나 받은 시계열적 관계를 판단할 수 있다.
- [0070] 도 11은 신규 철학 문언의 작성 시기를 판단하는 방식을 설명하기 위한 도면이다.
- [0071] 도 11에 도시되어 있는 바와 같이, 신규 철학 문언에서 나타난 년도 데이터를 추출하고 추출한 년도 데이터의 빈도를 단위 기간(P)별로 나누어 카운트한다. 단위 기간 중 카운트한 년도 데이터의 빈도가 가장 많은 단위 기간을 제외한 나머지 단위 기간의 년도 데이터를 삭제하고, 빈도가 가장 많은 단위 시간에 나타난 년도 데이터의 평균값을 신규 철학 문언의 작성 시기로 판단한다.
- [0072] 한편, 상술한 본 발명의 실시 예들은 컴퓨터에서 실행될 수 있는 프로그램으로 작성 가능하고, 컴퓨터로 읽을 수 있는 기록 매체를 이용하여 상기 프로그램을 동작시키는 범용 디지털 컴퓨터에서 구현될 수 있다.
- [0073] 상기 컴퓨터로 읽을 수 있는 기록 매체는 마그네틱 저장 매체(예를 들어, 롬, 플로피 디스크, 하드디스크 등), 광학적 판독 매체(예를 들면, 시디롬, 디브이디 등) 및 캐리어 웨이브(예를 들면, 인터넷을 통한 전송)와 같은 저장 매체를 포함한다.
- [0074] 본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 등록청구범위의 기술적 사상에 의해 정해져야 할 것이다.

부호의 설명

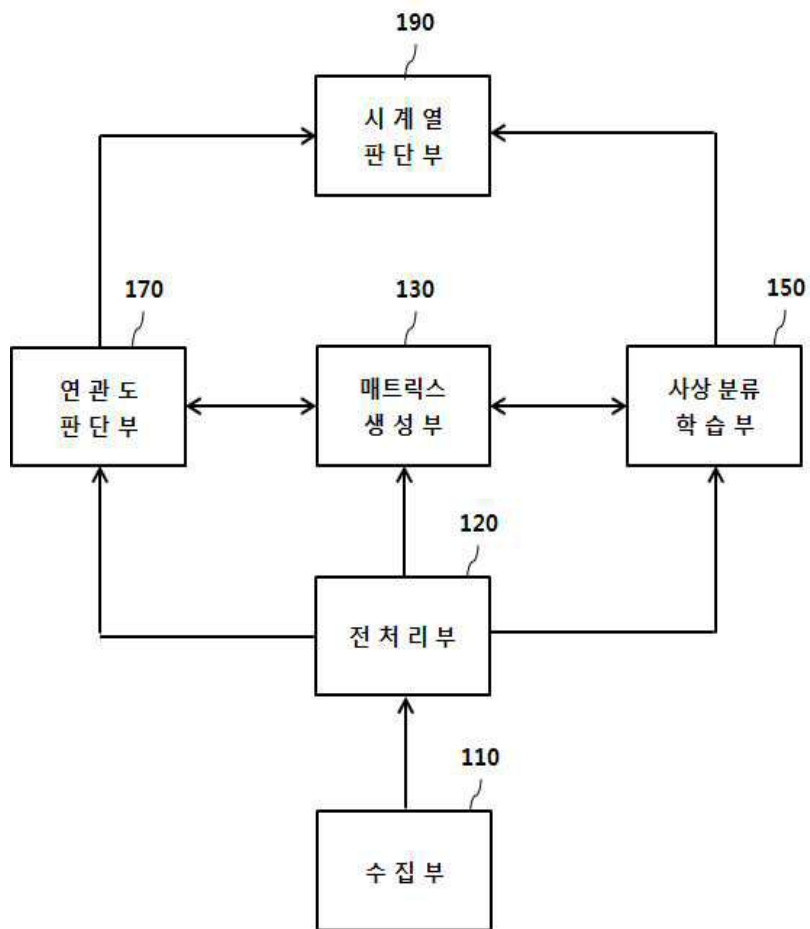
- [0075] 10: 데이터베이스부 30: 네트워크
50: 인문 사상 분류 장치 110: 수집부
120: 전처리부 130: 매트릭스 생성부
150: 사상 분류 학습부 170: 연관도 판단부
190: 시계열 판단부

도면

도면1



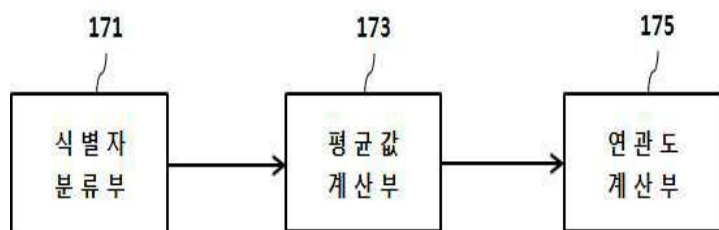
도면2



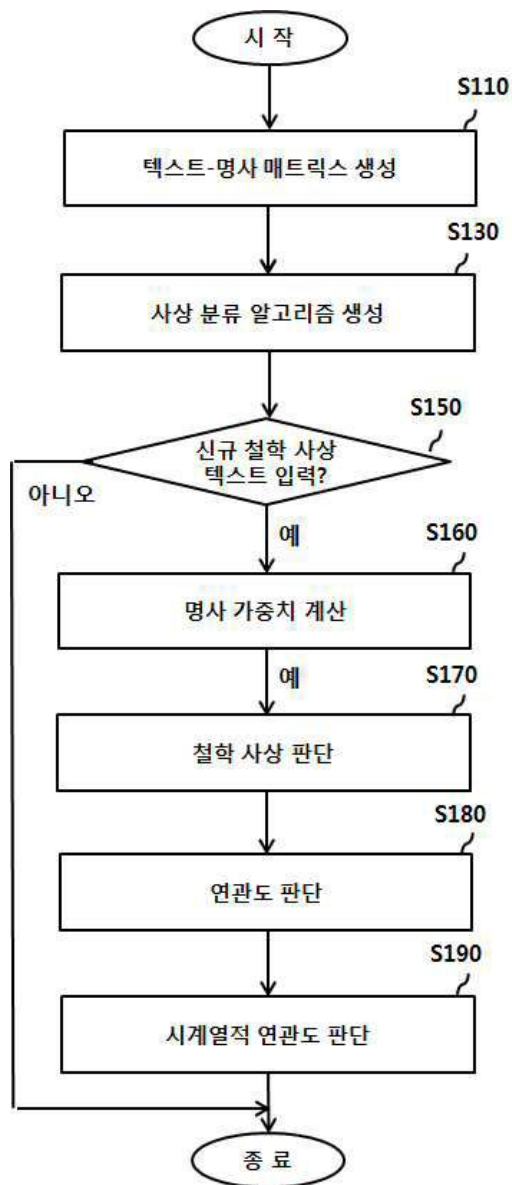
도면3



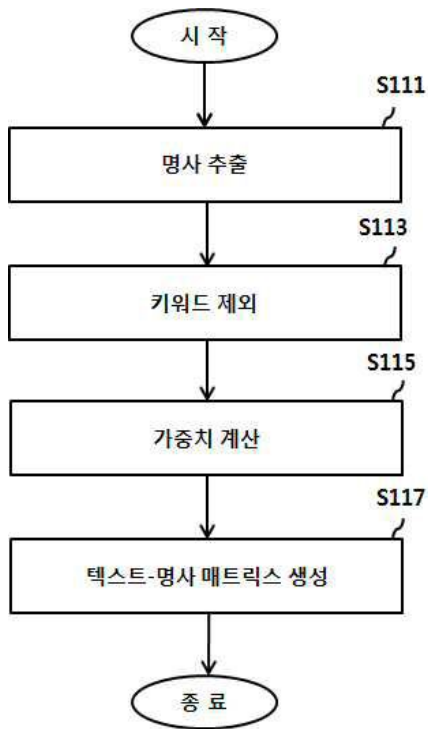
도면4



도면5



도면6



도면7

철학자	추출 명사
데카르트	프랑스 철학자 데카르트 나 나 말 데카르트 근대 인식 논쟁 대륙 합리 선두 근대 철학 포문 의미 근대 철학 아버지 평가 데카르트 세기 교회 권위 허무주의 회의 등
⋮	⋮
로크	책 인간 지식 직관 경험 합리 추론 인간 누구 신 지식 신 지식 가운데 누구 자신 생각 근거 까닭 지식 믿음 로크 경험 확인 해석 편견 혼란 객관적 지식 매력 등
⋮	⋮
흄	영국 경험 인식 비판 사람 흄 그 저서 이성 외적 사물 신의 존재 관념 원인 어디 우리 인식 마음 밖 자세 그때 외계 관념 길 경험 마음속 관념 움직임 등
⋮	⋮

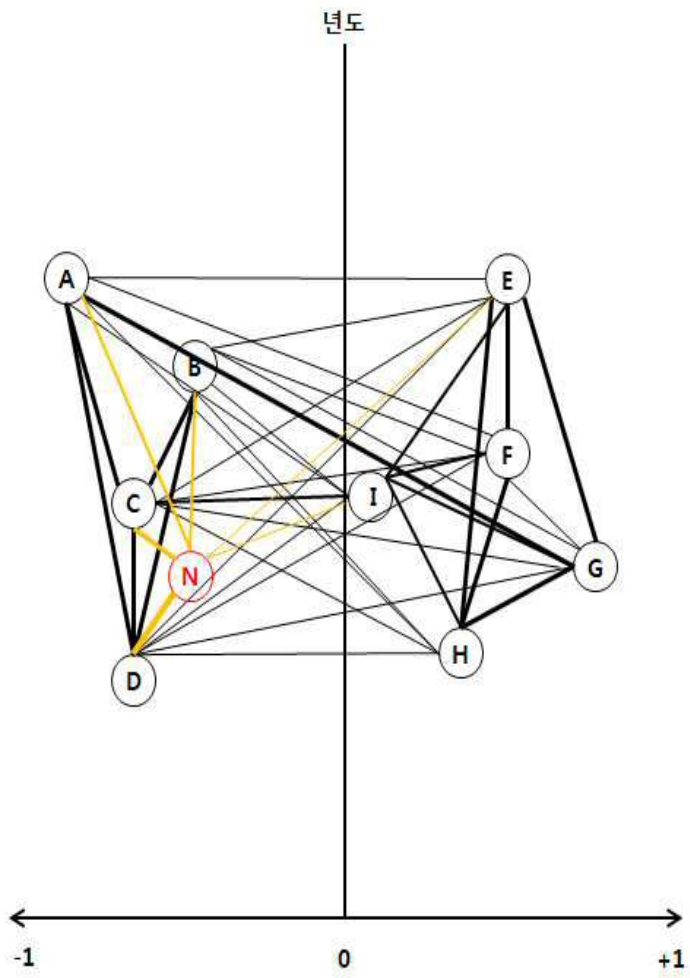
도면8

문서번호	감각	경험	인식	존재	필연	철학자
1	0.017359	0.003962	0.003626	0.001526	0.004109	데카르트
2	0.046697	0.003849	0.003131	0.000988	0.002661	데카르트
⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	0.021479	0.027086	0.029158	0.004908	0.003305	로크
42	0.050914	0.012589	0.019199	0	0	로크
⋮	⋮	⋮	⋮	⋮	⋮	⋮
105	0.010902	0.048524	0.019734	0.01661	0.005592	흄
106	0	0.031681	0.025769	0.004067	0	흄
⋮	⋮	⋮	⋮	⋮	⋮	⋮

도면9

	Berkeley	Decartes	Hume	Leibniz	Locke	Spinoza
Berkeley	1					
Decartes	0.447761	1				
Hume	0.336501	0.352719	1			
Leibniz	0.291569	0.359281	0.209868	1		
Locke	0.586265	0.396486	0.357334	0.253208	1	
Spinoza	0.35182	0.498597	0.280601	0.37575	0.311997	1

도면10



도면11

