(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0067507
(43) 공개일자 2024년05월17일

(51) 국제특허분류(Int. Cl.)
G06F 40/205 (2020.01) *G06F 40/268* (2020.01)
G06F 40/279 (2020.01) *G06F 40/30* (2020.01)
G06N 3/08 (2023.01)
(52) CPC특허분류
G06F 40/205 (2020.01)
G06F 40/268 (2020.01)
(21) 출원번호 10-2022-0148588
(22) 출원일자 2022년11월09일
심사청구일자 2022년11월09일

(71) 출원인
국립창원대학교 산학협력단
경상남도 창원시 의창구 창원대학로 20(퇴촌동)
(72) 발명자
차정원
경상남도 창원시 성산구 동산로 115, 105동 1407호 (상남동, 대동아파트)
홍성태
경상남도 창원시 의창구 북면 감계로 233, 403동 1801호 (창원 감계 힐스테이트 4차)
(74) 대리인
김종선, 이형석

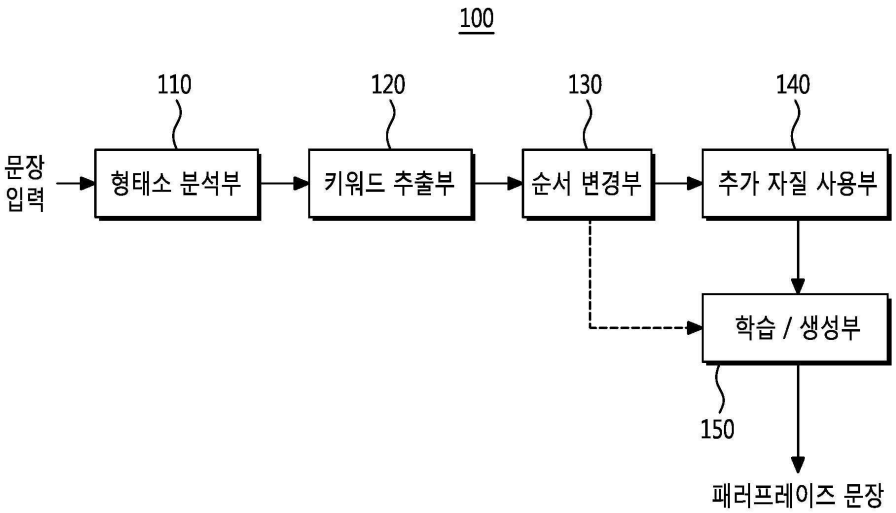
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 **키워드를 이용한 패러프레이즈 문장 생성장치 및 방법**

(57) 요 약

본 발명은 하나의 문장만으로 표현이 다양한 패러프레이즈 문장들을 생성하는 장치 및 방법을 제안한다. 본 발명의 패러프레이즈 문장 생성장치는, 입력 문장에 대해 형태소를 분석하는 형태소 분석부, 상기 분석 결과에서 실질 형태소만 키워드로 추출하는 키워드 추출부, 상기 추출된 키워드의 순서를 변경하는 순서 변경부, 및 상기 변경된 키워드 순서를 학습하여 패러프레이즈 문장을 생성하는 학습/생성부를 포함하여 구성된다. 상기 순서 변경부의 변경 횟수만큼 패러프레이즈 문장의 개수는 상이하게 제공된다.

대 표 도 - 도1



(52) CPC특허분류
G06F 40/279 (2020.01)
G06F 40/30 (2020.01)
G06N 3/08 (2023.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711152953
과제번호 2021-0-00354-002
부처명 과학기술정보통신부
과제관리(전문)기관명 정보통신기획평가원
연구사업명 차세대인공지능핵심원천기술개발
연구과제명 비정형 텍스트를 학습하여 쟁점별 사실과 논리적 근거 추론이 가능한 인공지능 원천기술
기 여 율 1/1
과제수행기관명 창원대학교 산학협력단
연구기간 2021.01.01 ~ 2022.12.31

명 세 서

청구범위

청구항 1

입력 문장에 대해 형태소를 분석하는 형태소 분석부;

상기 분석 결과에서 실질 형태소만 키워드로 추출하는 키워드 추출부;

상기 추출된 키워드의 순서를 변경하는 순서 변경부; 및

상기 변경된 키워드 순서를 학습하여 패러프레이즈 문장을 생성하는 학습/생성부를 포함하여 구성되는, 패러프레이즈 문장 생성장치.

청구항 2

제1항에 있어서,

상기 순서 변경부는, 적어도 1회 이상 수행될 수 있고,

상기 순서 변경부의 변경 횟수에 따라 상기 패러프레이즈 문장의 개수가 결정되는, 패러프레이즈 문장 생성장치.

청구항 3

제2항에 있어서,

상기 순서 변경부는, 키워드 순서 변경을 더 이상 수행할 수 없을 때까지 반복 수행되는, 패러프레이즈 문장 생성장치.

청구항 4

제2항에 있어서,

상기 순서 변경부는, 사용자에 의해 기 설정된 횟수로만 키워드 순서의 변경을 수행하는, 패러프레이즈 문장 생성장치.

청구항 5

제1항에 있어서,

상기 순서 변경부에 의해 순서가 변경된 키워드마다 구문론적 또는 의미론적 정보를 적용하는 추가 자질 사용부를 더 포함하여 구성되는, 패러프레이즈 문장 생성장치.

청구항 6

제1항에 있어서,

상기 학습/생성부가 생성한 적어도 둘 이상의 패러프레이즈 문장을 재 학습하여, 상기 입력 문장과 의미가 다른 패러프레이즈 문장을 추출하는 추출부를 더 포함하여 구성되는, 패러프레이즈 문장 생성장치.

청구항 7

하나의 입력된 문장에 대해 패러프레이즈 문장을 생성하는 장치가,

입력 문장의 형태소를 분석하는 분석 단계;

상기 형태소의 분석결과에서 실질 형태소만 키워드로 추출하는 추출 단계;

상기 추출된 키워드의 순서를 변경하는 변경 단계; 및

상기 순서가 변경된 키워드를 학습하여 상기 입력된 문장에 대해 적어도 하나의 패러프레이즈 문장을 생성하는

생성 단계를 포함하는, 패러프레이즈 문장 생성방법.

청구항 8

제7항에 있어서,

상기 변경 단계의 키워드 순서 변경 회수와 대응하여 상기 생성 단계의 패러프레이즈 문장 개수가 정해지는, 패러프레이즈 문장 생성방법.

청구항 9

제7항에 있어서,

상기 키워드의 순서 변경은, 기 설정된 횟수 또는 더 이상 변경할 수 없을 때까지 반복하여 실시되는, 패러프레이즈 문장 생성방법.

청구항 10

제7항에 있어서,

상기 변경 단계는, 유의어 사전을 이용하여 상기 키워드를 사용 가능한 다른 단어로 변경하는 단어 변경 단계를 더 포함하여 수행되는, 패러프레이즈 문장 생성방법.

청구항 11

제7항에 있어서,

상기 생성 단계 이후, 상기 입력된 문장과 상기 생성된 적어도 둘 이상의 패러프레이즈 문장을 비교하고, 의미가 다른 패러프레이즈 문장을 추출하는 추출 단계를 더 수행하는, 패러프레이즈 문장 생성방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 하나의 문장만으로 표현이 다양한 패러프레이즈 문장들을 생성하는 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 자연어 처리분야는 인공지능과 딥러닝(deep learning)을 활용하면서 더 우수한 성능을 보이고 있다. 자연어 처리분야에서 자연어의 이해를 위해서는 다양하게 표현된 문장들로부터 동일한 의미를 파악하는 것이 요구될 수 있다. 동일한 의미에도 불구하고 언어는 다양하게 표현될 수 있기 때문이다.

[0003] 따라서 동일한 의미에 대응하는 다수의 문장들, 즉 패러프레이즈(paraphrase) 문장들을 준비하는 것은 자연어 이해의 중요한 기반이 될 수 있다. 패러프레이즈 문장은 서로 다른 표현을 사용하여 동일한 의미를 전달하는 두 개 이상의 문장을 의미한다. 패러프레이즈 문장 생성은 입력으로 사용되는 하나의 문장에 대해 동일한 의미의 또 다른 문장을 생성하는 방법이 있다.

[0004] 그러나 종래의 패러프레이즈 문장 생성은 유의어 및 동의어 등의 어휘로 대체하여 코퍼스를 구축하는 방법으로 단순하게 문장 내의 단어를 치환하는 경우가 대부분이다. 이처럼 단어가 구로 치환되거나 구조가 치환되어 패러프레이즈 문장을 생성하는 경우, 해당하는 표현들이 다양하게 생성할 수 없는 문제가 있다.

[0005] 또 패러프레이즈 연구를 위한 공개된 학습 코퍼스는 두 문장 이상으로 이루어진 쌍으로 구성되는데, 이러한 학습 코퍼스를 구축하기 위해서는 상당한 시간과 비용이 소요되는 문제가 있다. 특히 구축된 학습 코퍼스에 대한 검토도 쉽지 않다.

선행기술문헌

특허문헌

[0006] (특허문헌 0001) 한국등록특허 10-2143157호

(특허문헌 0002) 한국공개특허 10-2017-0025424호

발명의 내용

해결하려는 과제

[0007] 본 발명의 목적은 상기와 같은 문제점을 해결하기 위하여 안출된 것으로서, 하나의 문장만으로 표현이 다양한 패러프레이즈 문장들을 생성할 수 있는 장치 및 방법을 제공하는 것이다.

[0008] 본 발명의 기술적 과제들은 이상에서 언급한 기술적 과제로 제한되지 않으며, 언급되지 않은 또 다른 기술적 과제들은 아래의 기재로부터 당업자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0009] 이와 같은 목적을 달성하기 위한 본 발명의 일 실시 예에 따른 패러프레이즈 문장 생성장치는, 입력 문장에 대해 형태소를 분석하는 형태소 분석부; 상기 분석 결과에서 실질 형태소만 키워드로 추출하는 키워드 추출부; 상기 추출된 키워드의 순서를 변경하는 순서 변경부; 및 상기 변경된 키워드 순서를 학습하여 패러프레이즈 문장을 생성하는 학습/생성부를 포함하여 구성되는 것을 특징으로 한다.

[0010] 상기 순서 변경부는, 적어도 1회 이상 수행될 수 있고, 상기 순서 변경부의 변경 횟수에 따라 상기 패러프레이즈 문장의 개수가 결정된다.

[0011] 상기 순서 변경부는, 키워드 순서 변경을 더 이상 수행할 수 없을 때까지 반복 수행될 수 있다.

[0012] 상기 순서 변경부는, 사용자에 의해 기 설정된 횟수로만 키워드 순서의 변경을 수행할 수 있다.

[0013] 상기 순서 변경부에 의해 순서가 변경된 키워드마다 구문론적 또는 의미론적 정보를 적용하는 추가 자질 사용부를 더 포함할 수 있다.

[0014] 상기 학습/생성부가 생성한 적어도 둘 이상의 패러프레이즈 문장을 재 학습하여, 상기 입력 문장과 의미가 다른 패러프레이즈 문장을 추출하는 추출부를 더 포함할 수 있다.

[0015] 본 발명의 다른 목적을 달성하기 위한 패러프레이즈 문장 생성방법은, 하나의 입력된 문장에 대해 패러프레이즈 문장을 생성하는 장치가, 입력 문장의 형태소를 분석하는 분석 단계; 상기 형태소의 분석결과에서 실질 형태소만 키워드로 추출하는 추출 단계; 상기 추출된 키워드의 순서를 변경하는 변경 단계; 및 상기 순서가 변경된 키워드를 학습하여 상기 입력된 문장에 대해 적어도 하나의 패러프레이즈 문장을 생성하는 생성 단계를 포함하는 것을 특징으로 한다.

[0016] 상기 변경 단계의 키워드 순서 변경 회수와 대응하여 상기 생성 단계의 패러프레이즈 문장 개수가 정해진다.

[0017] 상기 키워드 순서의 변경은, 기 설정된 횟수 또는 더 이상 변경할 수 없을 때까지 반복하여 실시된다.

[0018] 상기 변경 단계는, 유의어 사전을 이용하여 상기 키워드를 사용 가능한 다른 단어로 변경하는 단어 변경 단계를 더 포함한다.

[0019] 상기 생성 단계 이후, 상기 입력된 문장과 상기 생성된 적어도 둘 이상의 패러프레이즈 문장을 비교하고, 의미가 다른 패러프레이즈 문장을 추출하는 추출 단계를 더 수행할 수 있다.

발명의 효과

[0020] 이와 같은 본 발명에 따르면, 하나의 입력 문장에 대해 다양한 표현을 가지는 패러프레이즈 문장을 생성할 수 있는 효과가 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 바람직한 실시 예에 따른 패러프레이즈 문장 생성장치의 구성도이다.

도 2는 본 발명의 바람직한 실시 예에 따른 패러프레이즈 문장 생성방법의 흐름도이다.

도 3은 본 발명의 실시 예에 따라 패러프레이즈 문장을 생성하는 예시 구성도이다.

발명을 실시하기 위한 구체적인 내용

[0022] 본 발명은 다양한 변환을 가할 수 있고 여러 가지 실시 예를 가질 수 있는 바, 특정 실시 예들을 도면에 예시하고 상세하게 설명하고자 한다. 그러나, 이는 본 발명의 특정한 실시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변환, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 본 발명을 설명함에 있어서 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다.

[0023] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다.

[0024] 본 발명에서 사용한 용어는 단지 특정한 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.

[0025] 공간적으로 상대적인 용어인 아래(below, beneath, lower), 위(above, upper) 등은 도면에 도시되어 있는 바와 같이 하나의 소자 또는 구성 요소들과 다른 소자 또는 구성 요소들과의 상관 관계를 용이하게 기술하기 위해 사용될 수 있다. 공간적으로 상대적인 용어는 도면에 도시되어 있는 방향에 더하여 사용시 또는 동작시 소자의 서로 다른 방향을 포함하는 용어로 이해되어야 한다. 예를 들면, 도면에 도시되어 있는 소자를 뒤집을 경우, 다른 소자의 아래(below, beneath)로 기술된 소자는 다른 소자의 위(above, upper)에 놓여질 수 있다. 따라서, 예시적인 용어인 아래는 아래와 위의 방향을 모두 포함할 수 있다. 소자는 다른 방향으로도 배향될 수 있고, 이에 따라 공간적으로 상대적인 용어들은 배향에 따라 해석될 수 있다.

[0026] 본 발명에서 사용되는 “부” 또는 “부분” 등의 일부분을 나타내는 표현은 해당 구성요소가 특정 기능을 포함할 수 있는 장치, 특정 기능을 포함할 수 있는 소프트웨어, 또는 특정 기능을 포함할 수 있는 장치 및 소프트웨어의 결합을 나타낼 수 있음을 의미하나, 꼭 표현된 기능에 한정된다고 할 수는 없으며, 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

[0027] 따라서, 본 발명의 사상은 설명된 실시 예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

[0029] 이하에서는 도면에 도시한 실시 예에 기초하면서 본 발명에 대하여 더욱 상세하게 설명하기로 한다.

[0031] 도 1은 본 발명의 바람직한 실시 예에 따른 패러프레이즈 문장 생성장치의 구성도이다.

[0032] 도 1에 도시한 바와 같이 패러프레이즈 문장 생성장치(100)는, 형태소 분석부(110), 키워드 추출부(120), 순서 변경부(130), 추가 자질 사용부(140) 및 학습/생성부(150)를 포함한다.

[0033] 형태소 분석부(110)는 입력 문장에 대한 형태소를 분석하는 작업을 수행하는 유닛이다. 알려진 바와 같이 형태소는 단어 또는 어절을 구성하고 있는 최소한의 언어요소이다. 이러한 형태소는 실질적인 의미를 가지고 있는 실질(의미) 형태소(meaning morpheme)와 기능적 요소를 가지는 형식(기능) 형태소로 구분할 수 있다. 또 형태소 분석은 자연 언어 분석의 첫 단계로서 단어 또는 어절을 구성하는 각각의 형태소들을 인식하고 불규칙 활용이나, 축약, 탈락 현상이 일어난 경우 원형을 복원하는 과정을 말하기도 한다.

[0034] 본 실시 예에서 형태소 분석부(110)는, 예를 들어 기계학습을 통해 문장 분석을 수행할 수 있는 형태소 분석기를 말할 수 있다.

[0035] 키워드 추출부(120)는 형태소 분석부(110)의 분석 결과에서 키워드를 추출하는 작업을 수행하는 유닛이다. 키워드 추출부(120)가 추출하는 키워드는 형태소 분석에 따른 정보 중 실질 형태소만 추출한다. 앞서 언급했듯이 형태소는 2가지로 분류할 수 있고, 이때 형식 형태소는 실질 형태소에 붙어 주로 말과 말 사이의 관계를 표시하는

형태소로서, 주로 조사, 접사, 어미 등이 해당된다. 이처럼 형식 형태소는 문장의 의미를 구성하는데 중요한 정보가 아니기 때문에, 키워드 추출부(120)는 실제 어휘적 의미를 가지고 있는 실질 형태소만 사용하는 것이다.

[0036] 순서 변경부(130)는 나중에 생성된 패러프레이즈에서 다양한 구조적 및 의미적 변경을 추출하기 위하여 상기 키워드 추출부(120)가 추출한 키워드를 섞는(shuffle) 작업을 수행하는 유닛이다. 본 실시 예에 따르면 키워드의 셔플 수행에 따라 생성되는 패러프레이즈 문장에는 차이가 있다. 셔플을 수행하지 않고 생성된 문장과 셔플을 수행한 후에 생성된 문장의 표현력이 다를 수 있는 것이다. 키워드를 셔플시킬 경우 최초에 입력된 원 문장의 의미는 유지하면서 다양한 표현이 가능한 문장을 효율적으로 생성하는 것이 가능하다.

[0037] 즉 순서 변경부(130)는 키워드의 순서를 변경하여 다양한 표현을 가지는 패러프레이즈 문장이 생성되도록 하는 것이다.

[0038] 본 실시 예에서 순서 변경부(130)는 적어도 1회 이상 수행될 수 있고, 상기 순서 변경부(130)의 변경 횟수에 따라 최종 생성되는 패러프레이즈 문장의 개수가 정해진다. 또한 순서 변경부(130)는, 키워드 순서 변경을 더 이상 수행할 수 없을 때까지 반복하여 수행되거나 사용자에 의해 기 설정된 횟수만큼 수행될 수 있다.

[0039] 본 실시 예에 따르면 순서 변경부(130)는 필요에 따라 구문론적 또는 의미론적 정보를 사용할 수도 있다. 그래서 추가 자질 사용부(140)가 필요할 수 있다. 예를 들어 구문 분석기를 이용하여 문장의 구의 정보를 보존하면서 셔플링하는 방법은 대체로 긴 문장에서도 연속될 필요가 있는 정보들을 쉽게 유지할 수 있다. 물론 본 발명은 단순하게 랜덤으로 섞어 문장 내의 어떠한 순서 정보도 없이 셔플링하는 방법, 또는 셔플링하지 않는 방법도 적용할 수 있다.

[0040] 본 실시 예에서 순서 변경부(130)는 키워드의 순서를 변경하는 것 이외에 유의어 사전을 이용하여 같은 문맥에서 사용 가능한 다른 단어로 치환하는 기능도 수행할 수 있다.

[0041] 학습/생성부(150)는 상기 순서 변경부(130)의 변경 결과에 따른 문장을 학습하고 추론하여 제공하는 유닛이다. 학습/생성부(140)의 결과가 패러프레이즈 문장이 된다.

[0042] 본 발명에 따르면 생성된 패러프레이즈 문장들 중에서 입력 문장과 비교하여 의미가 다른 패러프레이즈 문장을 추출하는 추출부(도면 미도시)가 더 포함될 수 있다. 즉 키워드의 순서를 변경하기 때문에 매우 다양한 표현을 가지는 패러프레이즈 문장을 생성할 수 있는데, 이렇게 생성된 패러프레이즈 문장들 중에서 입력 문장과 의미가 다른 문장이 있을 수 있기 때문이다.

[0044] 도 2는 본 발명의 바람직한 실시 예에 따라 패러프레이즈 문장을 생성하는 방법을 설명하는 흐름도이다.

[0045] 패러프레이즈 문장 생성장치(100)가 패러프레이즈 문장 생성을 위한 문장을 입력받는 것에 의해 개시된다.

[0046] 형태소 분석부(110)는 입력된 문장에 대한 형태소 분석 작업을 수행하게 된다. 형태소 분석은 실질 형태소와 형식 형태소를 구분하는 작업일 수 있고, 실질적으로 문장에서 어휘적 의미를 가지는 단어들을 추출하는 것을 말할 수 있다. 그래서 형태소 분석부(110)의 형태소 분석 작업이 완료되면, 문장은 그 문장을 구성하는 실질적인 의미를 가지는 실질 형태소와 조사/접사/어미 등과 같은 형식 형태소로 구분되어 제공될 수 있을 것이다. 이러한 형태소 분석을 수행하지 않게 되면 문장 의미를 파악하는데 불필요한 정보들이 포함될 수 있고, 이는 최초 입력된 문장과 의미가 그대로이고 표현이 다양한 최종 결과물인 패러프레이즈 문장을 생성하는데 어려움이 생길 수 있다.

[0047] 키워드 추출부(120)는 형태소 분석부(110)의 분석 결과를 전달받는다. 상기 분석 결과는 문장에서 실질 형태소와 형식 형태소를 모두 포함하며, 키워드 추출부(120)는 실질적인 어휘적 의미를 포함하는 실질 형태소만 키워드로 추출한다. 실시 예에 따르면 실질 형태소만이 패러프레이즈 문장을 생성하는데 사용될 수 있다.

[0048] 순서 변경부(130)는 키워드 추출부(120)가 추출한 키워드를 전달받고 키워드의 배치 순서 등을 변경한다. 키워드를 섞는 방법은 단순한 shuffle, 구문정보를 이용한 shuffle, 의미정보를 이용한 shuffle 등이 가능하다.

[0049] 키워드의 순서를 변경하는 셔플링을 수행하는 이유는 생성될 패러프레이즈 문장에서 다양한 구조적 및 의미적 변경을 이끌어내기 위해서이다. 그래서 구조적 및 의미적 변경을 적용하기 위하여 추가자질 사용부(140)가 수행될 수 있다. 물론 이처럼 단순히 키워드를 셔플링한 후 구조적/의미적 정보를 적용할 수도 있지만, 키워드 셔플링 후 구조적/의미적 정보를 사용하지 않을 수도 있다.

[0050] 한편 본 실시 예는 키워드의 순서를 변경하는 것 이외에, 미리 제공되는 유의어 사전을 이용하여 상기 키워드를 사용 가능한 다른 단어로 변경하는 단어 변경 단계를 더 수행할 수 있다. 따라서 키워드의 순서를 변경하여 제공되는 학습 데이터보다 더 풍부한 학습 데이터를 제공할 수 있다.

[0051] 도 2에서 보듯이 본 실시 예는 학습 및 생성부(150)가 셔플링 미수행된 키워드들, 셔플링만 수행한 키워드들, 셔플링 수행 후 구조적/의미적 정보를 사용한 키워드 등의 학습 데이터를 대상으로 기계학습을 수행하고 패러프레이즈 문장을 생성하여 제공할 수 있다.

[0052] 각각의 경우에 따라 생성되는 패러프레이즈 문장은 차이가 있는데, 키워드의 순서를 변경하는 셔플링 과정을 수행한 결과가 더 다양한 패러프레이즈 문장을 생성할 수 있게 된다.

[0054] 본 실시 예에 따라 생성된 패러프레이즈 문장의 예는 도 3을 참조할 수 있다.

[0055] 도 3a를 참조하면, “사회심리학이라는 학문은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해서 연구한다.” 의 문장이 입력된다고 가정한다(S100).

[0056] 그러면 형태소 분석부(110)의 분석 결과(S110)에 따라, 키워드 추출부(120)는 '사회', '심리학', '학문', '문화', '제도', '규범', '인간', '심리', '미치', '영향', '대하', '연구'의 키워드를 추출할 수 있다(S120).

[0057] 이렇게 추출된 키워드를 대상으로 만약 키워드를 섞지 않고 기계학습을 수행하여 패러프레이즈 문장을 추론하면, “사회심리학이라는 학문은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해 연구한다.” 와 같이 최초 입력된 문장 대비 거의 차이가 없는 상태의 문장을 생성하게 된다. 입력 문장과 표현이 거의 동일하다고 할 수 있다.

[0058] 그러나 추출된 키워드를 섞는 작업을 하여(S130), 예를 들어 '대하', '미치', '사회', '심리학', '심리', '연구', '규범', '문화', '제도', '영향', '학문', '인간'과 같이 키워드의 순서를 변경하고 구조적/의미적 정보를 적용한 후 기계학습을 수행하면(S140, S150), “사회심리학은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해 연구하는 학문이다.” 와 같이 입력 문장과 의미는 그대로 유지하면서 다르게 표현된 패러프레이즈 문장을 생성할 수 있게 된다(S160).

[0059] 도면에는 도시하지 않았지만, 다른 예로 키워드 셔플링에 따라 키워드 순서가 '연구', '규범', '문화', '사회', '심리학', '제도', '영향', '대하', '미치', '심리', '학문', '인간'과 같이 변경될 수 있다. 이 경우 위에서 생성된 패러프레이즈 문장과는 다른 문장으로 예를 들어, “인간 심리에 문화, 제도, 규범 등이 미치는 영향에 대해 연구하는 학문이 사회심리학이다.” 와 같은 문장을 생성할 수 있다.

[0060] 즉 추출된 키워드의 순서를 변경하여 학습을 할 경우, 입력된 문장의 의미는 그대로 유지하면서 다른 표현으로 사용이 가능한 패러프레이즈 문장을 생성하는 것이 가능함을 알 수 있다.

[0061] 본 발명에 따르면 상기한 키워드를 유의어 사전을 이용하여 문장 내에서 사용 가능한 다른 단어로 변경하는 것도 가능하다. 따라서 더 많은 키워드 조합을 제공할 수 있을 것이고, 결과적으로 의미는 동일하지만 더 다양하고 풍부한 패러프레이즈 문장을 생성할 수 있다.

[0063] 이상과 같이 본 발명의 도시된 실시 예를 참고하여 설명하고 있으나, 이는 예시적인 것들에 불과하며, 본 발명이 속하는 기술 분야의 통상의 지식을 가진 자라면 본 발명의 요지 및 범위에 벗어나지 않으면서도 다양한 변형, 변경 및 균등한 타 실시 예들이 가능하다는 것을 명백하게 알 수 있을 것이다. 따라서 본 발명의 진정한 기술적 보호 범위는 첨부된 청구범위의 기술적인 사상에 의해 정해져야 할 것이다.

부호의 설명

[0064] 100: 패러프레이즈 문장 생성장치

110: 형태소 분석부

120: 키워드 추출부

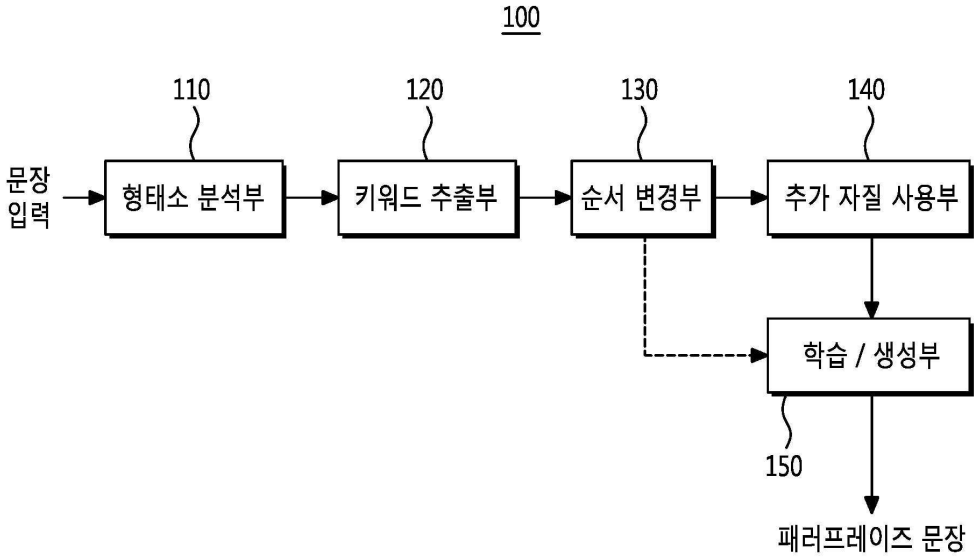
130: 순서 변경부

140: 추가 자질 사용부

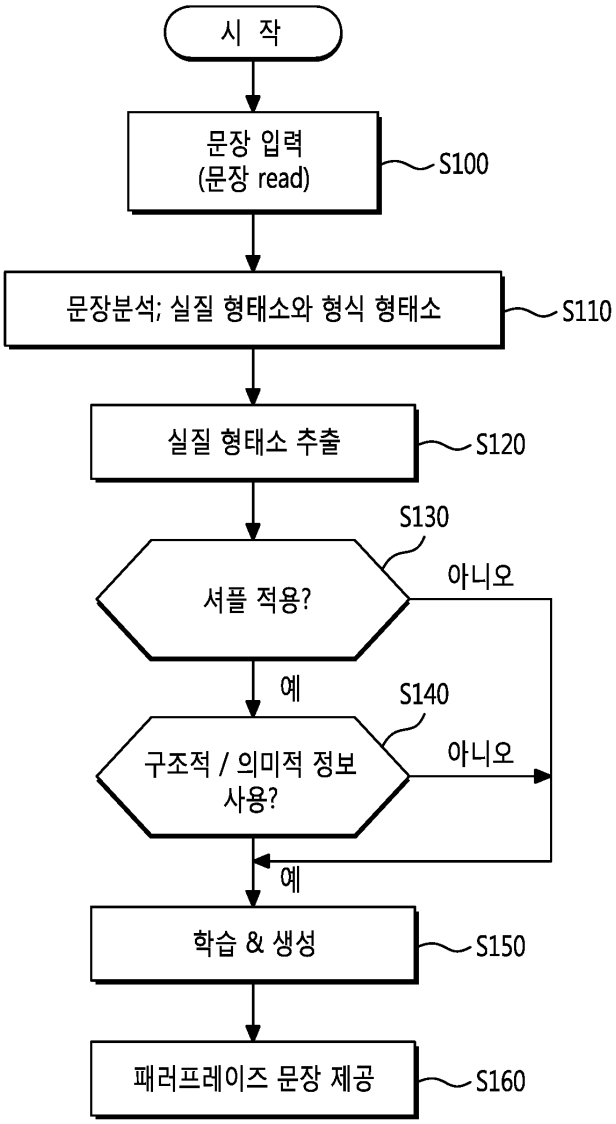
150: 학습/생성부

도면

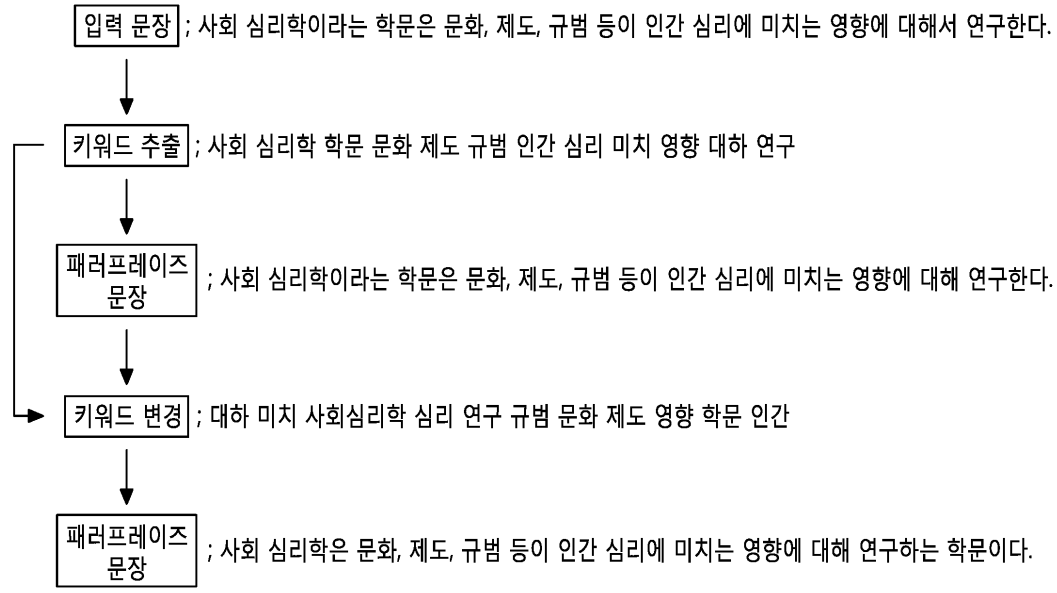
도면1



도면2


시 작
문장 입력
(문장 read)
S100
문장분석; 실질 형태소와 형식 형태소
S110
실질 형태소 추출
S120
S130
셔플 적용?
아니오
예
S140
구조적 / 의미적 정보
사용?
아니오
예
학습 & 생성
S150
패러프레이즈 문장 제공
S160

도면3


입력 문장
; 사회 심리학이라는 학문은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해서 연구한다.
키워드 추출
; 사회 심리학 학문 문화 제도 규범 인간 심리 미치 영향 대하 연구
패러프레이즈 문장
; 사회 심리학이라는 학문은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해 연구한다.
키워드 변경
; 대하 미치 사회심리학 심리 연구 규범 문화 제도 영향 학문 인간
패러프레이즈 문장
; 사회 심리학은 문화, 제도, 규범 등이 인간 심리에 미치는 영향에 대해 연구하는 학문이다.