(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2021 년 6 월 10 일 (10.06.2021)

WIPO | PCT

(10) 국제공개번호
WO 2021/112384 A1

(51) 국제특허분류:
*G06F 40/279* (2020.01) *G06F 40/268* (2020.01)
*G06F 40/205* (2020.01)

(21) 국제출원번호: PCT/KR2020/013525

(22) 국제출원일: 2020 년 10 월 6 일 (06.10.2020)

(25) 출원언어: 한국어

(26) 공개언어: 한국어

(30) 우선권정보:
10-2019-0160742 2019 년 12 월 5 일 (05.12.2019) KR

(71) 출원인: 울산대학교 산학협력단 (**UNIVERSITY OF ULSAN FOUNDATION FOR INDUSTRY COOPERATION**) [KR/KR]; 44610 울산시 남구 대학로 93 (무거동), Ulsan (KR).

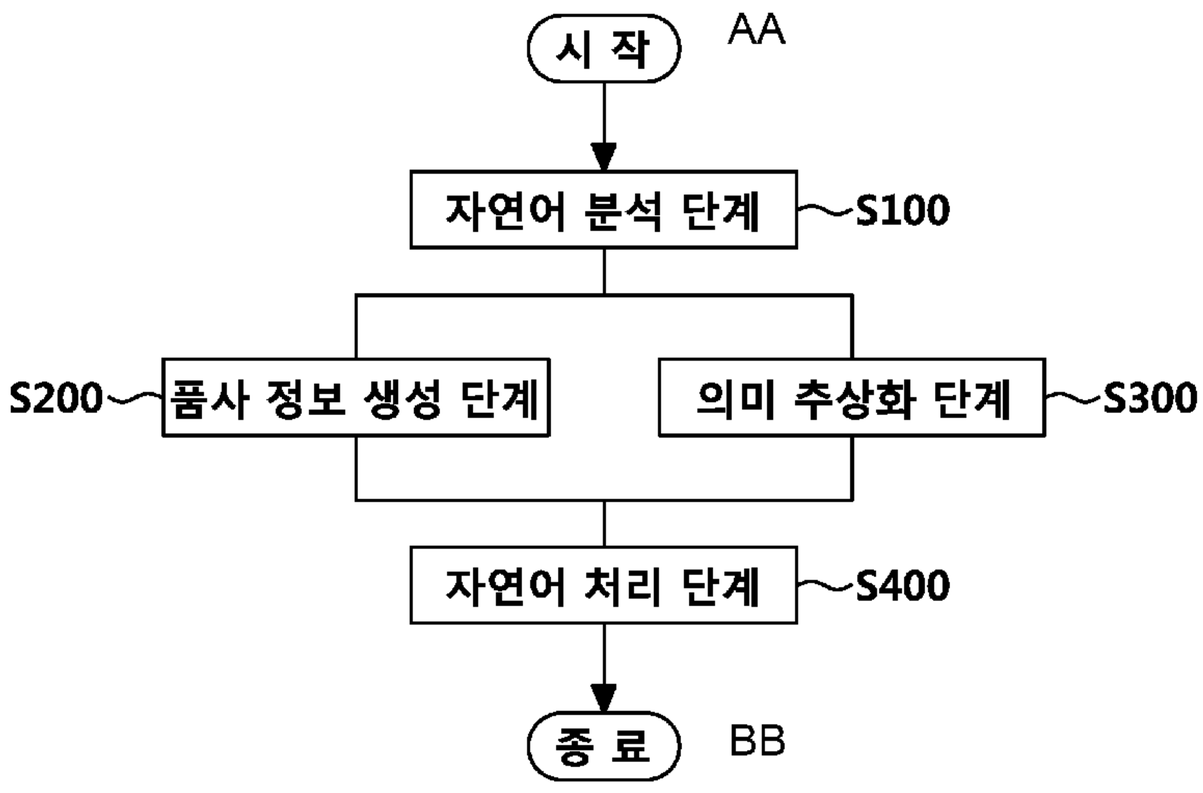
(72) 발명자: 옥철영 (**OCK, Cheol Young**); 44608 울산시 남구 신복로35번길 6, 101동 602호 (무거동), Ulsan (KR). 이주상 (**LEE, Ju Sang**); 44518 울산시 중구 학성로 179, 1007호 (학성동), Ulsan (KR).

(74) 대리인: 김종선 (**KIM, Jong Sun**); 06242 서울시 강남구 역삼로3길 11, 12층 (역삼동, 광성빌딩), Seoul (KR).

(81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD,

(54) **Title:** NATURAL LANGUAGE REPRESENTATION METHOD AND COMPUTER PROGRAM STORED IN MEDIUM TO EXECUTE SAME

(54) **발명의 명칭:** 자연어 표상 방법 및 이를 수행하기 위해 매체에 저장된 컴퓨터 프로그램




S100 ... Natural language analysis step
S200 ... Part-of-speech information generation step
S300 ... Meaning abstraction step
S400 ... Natural language processing step
AA ... Start
BB ... End

(57) **Abstract:** The present invention can rapidly and highly accurately represent and learn natural language by comprising: a natural language analysis step for dividing natural language into basic units; a part-of-speech information generation step for generating part-of-speech information corresponding to basic units; a meaning abstraction step for generating higher-level concept information comprising the basic units; and natural language representation step for processing the natural language using the part-of-speech information and higher-level concept information.

(57) **요약서:** 본 발명은 자연어를 기본 단위로 분리하는 자연어 분석 단계, 기본 단위에 품사 정보를 생성하는 품사 정보 생성 단계, 기본 단위를 포함하는 상위 개념 정보를 생성하는 의미 추상화 단계, 및 품사 정보 및 상위 개념 정보를 이용하여 자연어를 처리하는 자연어 표상 단계를 포함하여, 자연어 표상 및 학습을 빠르고 정확도 높게 수행할 수 있다.

[다음 쪽 계속]

WO 2021/112384 A1

SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

**규칙 4.17에 의한 선언서:**

— 신규성을 해치지 아니하는 개시 또는 신규성 상실의 예외에 관한 선언 (규칙 4.17(v))

**공개:**

— 국제조사보고서와 함께 (조약 제21조(3))

# 명세서

# 발명의 명칭: 자연어 표상 방법 및 이를 수행하기 위해 매체에 저장된 컴퓨터 프로그램

## 기술분야

[1] 본 발명은 자연어 표상 방법 및 이를 수행하기 위해 매체에 저장된 컴퓨터 프로그램에 관한 것으로서, 더욱 상세하게는 빠르고 정확성 있게 자연어를 표상하는 자연어 표상 방법 및 이를 수행하기 위해 매체에 저장된 컴퓨터 프로그램에 관한 것이다.

## 배경기술

[2] 자연어 표상은 자연어를 컴퓨터가 이해할 수 있는 형태로 표현하는 방법이다. 인간은 텍스트로 표현된 자연어에 대해 사전에 습득한 지식을 통해 정보를 획득할 수 있다. 그러나, 컴퓨터의 경우 단순 텍스트로 표현된 자연어에서 단순히 문자라는 정보만을 획득이 가능하다.

[3] 인간과 비슷한 수준의 정보를 컴퓨터에게 제공하기 위해 자연어 표상을 사용한다. 최근 자연어 처리 분야에서 심층 학습(딥러닝)이 주목받기 시작하면서 자연어 표상 방법의 중요도가 증가했다.

[4] 종래 기술에 따르면, 자연어는 고차원의 벡터로 표상된다. 벡터로 표상한 자연어는 벡터 차원수에 의해 정보가 한정적인 문제점이 있다. 자연어의 경우 복합적인 관계를 가지며 문법적 요소와 의미적 요소로 인해 단순히 고차원의 벡터로 표현하기 부족하기 때문이다.

[5] 다른 종래 기술의 경우에, 문맥을 분석하여 복수의 의미를 가지는 경우, 문맥에 따라 해석하여 정확한 해석이 가능했다. 즉, 복수의 의미 중 문맥에 부합하는 의미를 해석할 수 있다. 그러나, 많은 양의 학습 데이터를 사용해야 하고, 많은 연산이 필요해 학습에 많은 시간이 소요되는 문제점이 있다.

## 발명의 상세한 설명

## 기술적 과제

[6] 본 발명의 일 실시 예는, 정확하고 빠른 자연어 표상 및 학습을 가능하게 하는 자연어 표상 방법 및 이를 수행할 수 있게 매체에 저장된 컴퓨터 프로그램을 제공하는 것을 목적으로 한다.

## 과제 해결 수단

[7] 이와 같은 목적을 달성하기 위한 본 발명은, 상기한 바와 같은 목적을 달성하기 위하여, 본 발명의 일 실시 예에 따른 자연어 표상 방법은, 자연어를 기본 단위로 분리하는 자연어 분석 단계, 기본 단위에 품사 정보를 생성하는 품사 정보 생성 단계, 기본 단위를 포함하는 상위 개념 정보를 생성하는 의미 추상화 단계, 및 품사 정보 및 상위 개념 정보를 이용하여 자연어를 처리하는 자연어 표상 단계를 포함하는 것을 특징으로 한다.

[8] 또한, 기본 단위는 형태소 단위인 것을 특징으로 한다.

[9] 또한, 형태소를 그 품사 정보에 따라 식별하는 형태소 식별 정보를 생성하는 식별 정보 생성 단계를 더 포함하는 것을 특징으로 한다.

[10] 또한, 품사 정보는 명사, 용언, 접두사, 접미사 및 어미를 포함하는 것을 특징으로 한다.

[11] 또한, 형태소 식별 정보는 형태소 및 특수 문자 중 적어도 하나를 포함하는 것을 특징으로 한다.

[12] 또한, 품사 정보가 명사인 경우, 형태소 식별 정보는 형태소만 포함하며, 품사 정보가 용언 또는 접두사인 경우, 형태소 식별 정보는 형태소와 형태소 뒤에 배치된 특수 문자를 포함하고, 품사 정보가 접미사 또는 어미인 경우, 형태소 식별 정보는 특수 문자와 특수 문자의 뒤에 배치된 형태소를 포함하는 것을 특징으로 한다.

[13] 또한, 의미 추상화 단계에서 상위 개념 정보는, 기 저장된 상위 개념 데이터베이스에서 기본 단위를 기초로 선택된 것을 특징으로 한다.

[14] 또한, 상위 개념 데이터베이스는 복수의 상위 개념 정보를 포함하며, 상위 개념 정보는 복수의 하위 개념 정보를 포함하는 것을 특징으로 한다.

[15] 또한, 상위 개념 정보와 하위 개념 정보는 명사인 것을 특징으로 한다.

[16] 또한, 본 발명의 일 실시 예에 따른 매체에 저장된 컴퓨터 프로그램은 본 발명의 일 실시 예에 따른 자연어 표상 방법을 실행한다.

## 발명의 효과

[17] 이상과 같은 본 발명의 자연어 표상 방법 및 이를 수행하기 위해 매체에 저장된 컴퓨터 프로그램에 따르면, 자연어를 정확도 높으면서도 빠르게 표상하여 학습할 수 있다.

## 도면의 간단한 설명

[18] 도 1은 본 발명의 일 실시 예에 따른 자연어 표상 방법을 도시한 플로우 차트이다.

[19] 도 2는 도 1에 도시된 자연어 표상 방법의 다른 실시 예를 도시한 플로우 차트이다.

[20] 도 3은 본 발명의 일 실시 예에 따른 형태소 식별 정보를 도시한 표이다.

[21] 도 4는 본 발명의 일 실시 예에 따른 상위 개념 데이터베이스를 도시한 도면이다.

## 발명의 실시를 위한 최선의 형태

[22] 본 발명의 목적 및 효과, 그리고 그것들을 달성하기 위한 기술적 구성들은 첨부되는 도면과 함께 상세하게 후술되어 있는 실시 예들을 참조하면 명확해질 것이다. 본 발명을 설명함에 있어서 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다.

[23] 그리고 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다.

[24] 그러나 본 발명은 이하에서 개시되는 실시 예들에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있다. 단지 본 실시 예들은 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.

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

[26] 도 1은 본 발명의 일 실시 예에 따른 자연어 표상 방법을 도시한 플로우 차트이다.

[27] 도 1을 참조하면, 본 발명의 일 실시 예에 따른 자연어 표상 방법은 자연어를 컴퓨터에서 빠르고 정확하게 표상하기 위해서, 자연어 분석 단계(S100), 품사 정보 생성 단계(S200), 의미 추상화 단계(S300), 및 자연어 표상 단계를 포함한다. 자연어 표상은 자연어를 컴퓨터가 이해할 수 있는 형태로 표현하는 방법이다. 자연어 처리는 자연어 표상된 정보를 통해 학습 등으로 활용하는 것이다.

[28] 자연어는 일반적으로 인간이 사용하는 말을 말한다. 자연어 표상이란 딥러닝 학습을 위해 자연어를 컴퓨터가 인식하게 하기 위한 기술을 말한다. 자연어는 여러 단어, 어절, 문장 들로 이루어진다.

[29] 자연어 분석 단계(S100)의 경우, 복수의 단어, 문장을 포함하는 자연어를 기본 단위로 분석(분리)한다. 여기서, 기본 단위는 형태소 단위이다. 일 실시 예로, "학교에" 라는 어절을 먼저 "학교"와 "에"로 분리될 수 있다. 한국어는 교착어로 조사와 접사의 의미를 인식하기 위해 어절 단위보다 축소된 형태소 단위를 사용한다.

[30] 품사 정보 생성 단계(S200)의 경우, 기본 단위에 해당하는 품사를 적용한다. 즉, 분류된 형태소에 따라 품사 정보를 생성한다. 품사 정보는 명사, 용언, 접두사, 접미사 및 어미를 포함할 수 있다. 일 실시 예로, "학교가"라는 어절의 경우, "학교"와 "가"로 분석될 수 있다. 여기서, "학교"에 대한 품사 정보는 명사이다. 또한, "가"에 대한 품사 정보는 접미사이다.

## 발명의 실시를 위한 형태

[31] 도 2는 도 1에 도시된 자연어 표상 방법의 다른 실시 예를 도시한 플로우 차트이며, 도 3은 본 발명의 일 실시 예에 따른 형태소 식별 정보를 도시한 표이다.

[32] 도 2를 참조하면, 본 발명의 일 실시 예에 따른 자연어 표상 방법은 식별 정보 생성 단계(S210)를 더 포함할 수 있다. 식별 정보 생성 단계(S210)의 경우, 기본 단위 즉, 형태소를 그 품사 정보에 따라 식별(분류)하기 위한 형태소 식별 정보를

생성한다.

[33] 도 3을 참조하면, 형태소 식별 정보는 형태소 및 특수 문자 중 적어도 하나를 포함한다. 형태소 식별 정보는 형태소가 복수의 품사를 가지는 경우에 유용하다. 즉, 형태소 식별 정보는 복수의 품사를 가지는 형태소를 구분한다. 형태소 식별 정보는 빠른 자연어 표상을 가능하게 한다. 일 실시 예로, 형태소의 품사에 따라 특수 문자 "##"를 추가한다.

[34] 구체적으로, 품사 정보가 명사인 경우, 형태소 식별 정보는 형태소만을 포함한다. 즉, 명사의 경우 특수 문자가 추가되지 않는다. 예컨대,"사과"에 대한 형태소 식별 정보는 "사과"이다. 품사 정보가 용언 또는 접두사인 경우, 형태소 식별 정보는 형태소와 형태소 뒤에 배치된 특수 문자를 포함한다. 즉, 용언인 "갈"에 대한 형태소 식별 정보는 "갈##"이다. 또한, 접두사인 "수"에 대한 형태소 식별 정보는 "수##"이다.

[35] 품사 정보가 접미사 또는 어미인 경우, 형태소 식별 정보는 특수 문자와 특수 문자의 뒤에 배치된 형태소를 포함한다. 즉, 접미사인 "가"에 대한 형태소 식별 정보는 "##가"이다. 또한, 어미인 "다"에 대한 형태소 식별 정보는 "##다"이다.

[36] 도 4는 본 발명의 일 실시 예에 따른 상위 개념 데이터베이스를 도시한 도면이다.

[37] 도 4를 참조하면, 의미 추상화 단계(S300)의 경우, 기본 단위를 기초로 기본 단위(111)를 포함하는 상위 개념 정보(110)를 생성한다. 상위 개념 정보(110)는 상위 개념 데이터베이스(100)에서 기본 단위(111)를 기초로 선택된다. 상위 개념 데이터베이스(100)는 복수의 상위 개념 정보(110)를 포함한다. 상위 개념 정보(110)는 복수의 하위 개념 정보(111)를 포함한다. 여기서, 하위 개념 정보(111)는 기본 단위(111)이다. 상위 개념 데이터베이스(100)는 기 설정되어 저장된다. 상위 개념 정보(110)를 통해 기본 단위(111)의 개념 확장이 가능하여, 자연어 표상의 소요 시간이 단축될 수 있다. 즉, 자연어 학습 데이터를 확장하여 학습하는 효과를 가져온다.

[38] 일 실시 예로, 상위 개념 정보(110)와 하위 개념 정보(111)는 명사이다. 기본 단위(111)인 하위 개념 정보(111)가 연필인 경우, 상위 개념 정보(110)는 필기 도구로 선택된다. 또한, 기본 단위인 하위 개념 정보(111)가 사과인 경우, 상위 개념 정보(110)는 과일로 선택된다.

[39] 상위 개념 정보(110)는 기본 단위(111)와 관계성 및 유사한 개념을 가진 하위 개념 정보(111)들을 군집하게 한다. 예컨대, "연필"이라는 기본 단위(111)를 "필기 도구"로 추상화하면, "연필"은 "필기 도구"의 하위 개념 정보(111)들과 관계성을 가지게 된다. 즉, "연필"의 주변 개념인 "필기 도구"의 하위 개념 정보(111)들을 학습하는 효과를 가져온다.

[40] 자연어 처리 단계(S400)의 경우, 기본 단위의 품사 정보와 상위 개념 정보를 활용하여 자연어를 처리한다. 본 발명의 경우, 품사 정보와 상위 개념 정보를 BERT(Bidirectional Encoder Representations from Transformers) 모델에 적용한다.

BERT는 Transformer[7]라는 encoder-decoder 모델의 encoder를 사용하여 자연어를 표상하는 모델이다.

[41] 한편, 본 발명의 일 실시 예에 따른 매체에 저장된 컴퓨터 프로그램은 본 발명의 일 실시 예에 따른 자연어 표상 방법을 실행한다.

[42] 본 발명에 따르면, 분석된 단어들 중 이전 단어 정보에 대한 종속성을 가지지 않아 빠른 자연어 표상 및 학습이 가능하다.

[43] 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.

## 산업상 이용가능성

[44] 정확하고 빠른 자연어 표상 및 학습을 가능하게 하는 컴퓨터 프로그램 및 이러한 컴퓨터 프로그램이 구비된 다양한 형태의 자연어 학습 디바이스 등에 이용할 수 있다.

# 청구범위

[청구항 1] 자연어를 기본 단위로 분리하는 자연어 분석 단계;
기본 단위에 품사 정보를 생성하는 품사 정보 생성 단계;
상기 기본 단위를 포함하는 상위 개념 정보를 생성하는 의미 추상화 단계; 및
상기 품사 정보 및 상기 상위 개념 정보를 이용하여 자연어를 처리하는 자연어 처리 단계를 포함하는 자연어 표상 방법.

[청구항 2] 제 1 항에 있어서,
상기 기본 단위는 형태소 단위인 자연어 표상 방법.

[청구항 3] 제 2 항에 있어서,
상기 형태소를 그 품사 정보에 따라 식별하는 형태소 식별 정보를 생성하는 식별 정보 생성 단계를 더 포함하는 자연어 표상 방법.

[청구항 4] 제 3 항에 있어서,
상기 품사 정보는 명사, 용언, 접두사, 접미사 및 어미를 포함하는 자연어 표상 방법.

[청구항 5] 제 4 항에 있어서,
상기 형태소 식별 정보는 형태소 및 특수 문자 중 적어도 하나를 포함하는 자연어 표상 방법.

[청구항 6] 제 5 항에 있어서,
상기 품사 정보가 명사인 경우, 상기 형태소 식별 정보는 형태소만 포함하며,
상기 품사 정보가 용언 또는 접두사인 경우, 상기 형태소 식별 정보는 상기 형태소와 형태소 뒤에 배치된 특수 문자를 포함하고,
상기 품사 정보가 접미사 또는 어미인 경우, 상기 형태소 식별 정보는 상기 특수 문자와 특수 문자의 뒤에 배치된 형태소를 포함하는 자연어 표상 방법.

[청구항 7] 제 1 항에 있어서,
상기 의미 추상화 단계에서 상위 개념 정보는,
기 저장된 상위 개념 데이터베이스에서 기본 단위를 기초로 선택된 자연어 표상 방법.

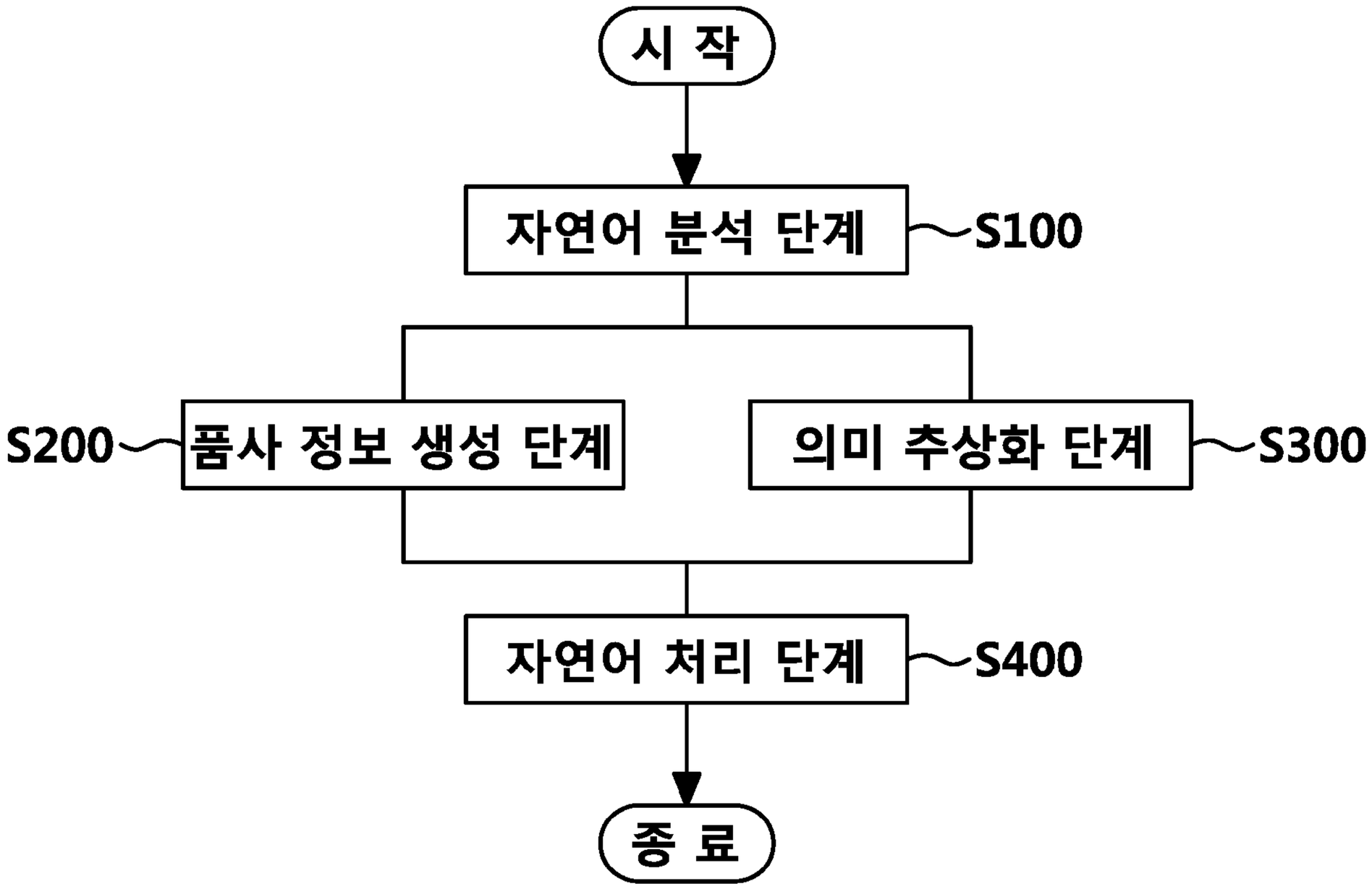
[청구항 8] 제 7 항에 있어서,
상기 상위 개념 데이터베이스는 복수의 상위 개념 정보를 포함하며,
상기 상위 개념 정보는 복수의 하위 개념 정보를 포함하는 자연어 표상 방법.

[청구항 9] 제 8 항에 있어서,
상기 상위 개념 정보와 상기 하위 개념 정보는 명사인 자연어 표상 방법.

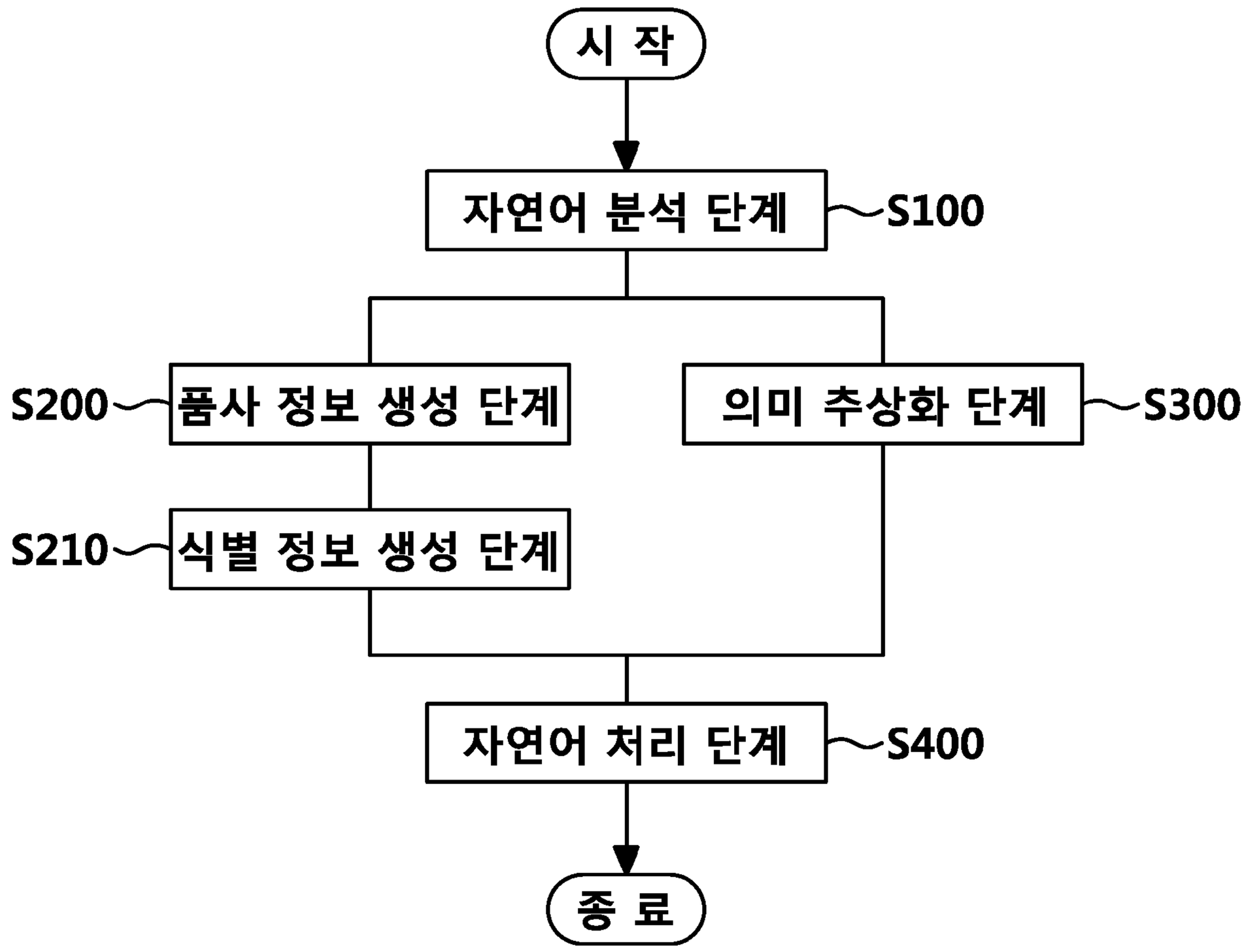
[청구항 10] 제 1 항의 자연어 표상 방법을 실행시키기 위해 매체에 저장된 컴퓨터

프로그램.

[도1]


시 작
자연어 분석 단계
S100
S200
품사 정보 생성 단계
의미 추상화 단계
S300
자연어 처리 단계
S400
종 료

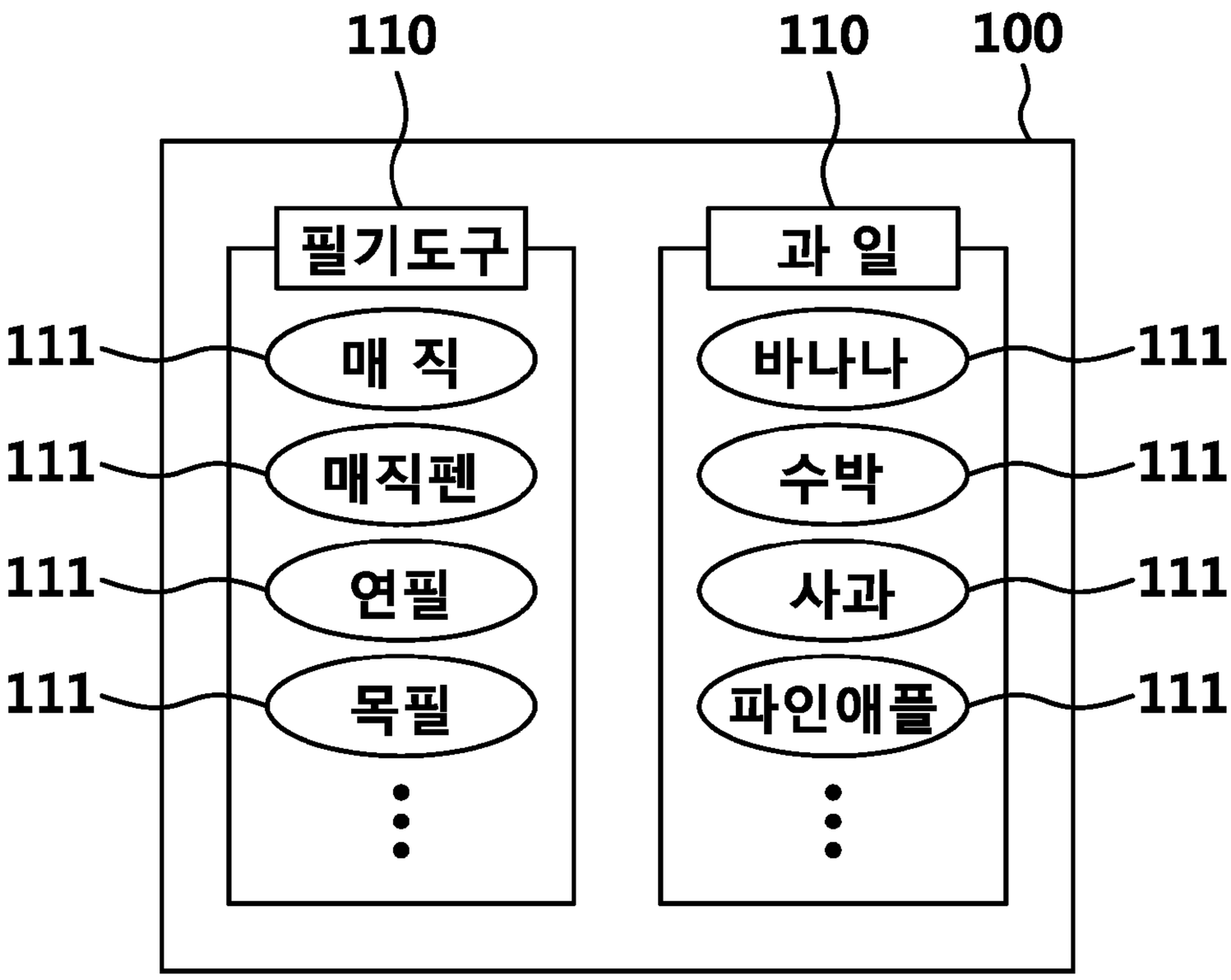
[도2]


시 작
자연어 분석 단계
S100
S200
품사 정보 생성 단계
의미 추상화 단계
S300
S210
식별 정보 생성 단계
자연어 처리 단계
S400
종 료

[도3]

| 품사 | 형태소 | 형태소 식별 정보 |
|---|---|---|
| 명사 | 사과 / NNG | 사과 |
| 용언 | 갈 / VV | 갈## |
| 접두사 | 수 / XPN | 수## |
| 접미사 | 가 / XSN | ##가 |
| 어미 | 다 / EP | ##다 |

[도4]



# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/KR2020/013525**

## A. CLASSIFICATION OF SUBJECT MATTER

**G06F 40/279**(2020.01)i; **G06F 40/205**(2020.01)i; **G06F 40/268**(2020.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F 40/279(2020.01); G06F 17/21(2006.01); G06F 17/27(2006.01); G06F 17/28(2006.01); G06N 3/00(2006.01); G06N 3/04(2006.01)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above
Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & keywords: 자연어(natural-language), 품사(word-class), 형태소(morpheme), 식별(percipience), 의미(meaning)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | JP 2008-293359 A (SHARP CORP.) 04 December 2008 (2008-12-04)<br>See paragraphs [0036]-[0049]; claim 1; and figure 2. | 1-10 |
| A | JP 2015-169947 A (NIPPON TELEGR & TELEPH CORP.) 28 September 2015 (2015-09-28)<br>See paragraphs [0057]-[0073]; and claim 1. | 1-10 |
| A | JP 2014-013549 A (BRAINPAD INC.) 23 January 2014 (2014-01-23)<br>See paragraphs [0029]-[0033]; claim 1; and figure 1. | 1-10 |
| A | KR 10-2004-0051351 A (ELECTRONICS AND TELECOMMUNICATIONS RESEARCH INSTITUTE) 18 June 2004 (2004-06-18)<br>See page 2, line 15 - page 3, line 6; claim 1; and figure 2. | 1-10 |
| A | KR 10-2017-0000201 A (ACRYL INC.) 02 January 2017 (2017-01-02)<br>See paragraphs [0027]-[0033]; claim 1; and figure 1. | 1-10 |

☑ Further documents are listed in the continuation of Box C. ☑ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| Date of the actual completion of the international search | Date of mailing of the international search report |
|---|---|
| **19 January 2021** | **20 January 2021** |
| Name and mailing address of the ISA/KR<br>**Korean Intellectual Property Office**<br>**Government Complex-Daejeon Building 4, 189 Cheongsa-ro, Seo-gu, Daejeon 35208**<br>Facsimile No. **+82-42-481-8578** | Authorized officer<br><br>Telephone No. |

# INTERNATIONAL SEARCH REPORT

International application No.
**PCT/KR2020/013525**

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X | 이주상 등. 의미 정보와 BERT를 결합한 개념 언어 모델, 제31회 한글 및 한국어 정보처리 학술대회 논문집 (2019년). ISSN 2005-3053. 12 October 2019, pp. 3-7 (LEE, Ju-Sang et al. A Concept Language Model combining Word Sense Information and BERT. Proceedings of the 31st Annual Conference on Human & Cognitive Language Technology (2019).).<br><URL: http://hclt.kr/symp/?lnb=conference>.<br>(This document is a document declaring exceptions to lack of novelty by the applicant.) | 1-10 |

# INTERNATIONAL SEARCH REPORT
**Information on patent family members**

International application No.
**PCT/KR2020/013525**

| Patent document cited in search report | Publication date (day/month/year) | Patent family member(s) | Publication date (day/month/year) |
|---|---|---|---|
| JP 2008-293359 A | 04 December 2008 | JP 5137107 B2 | 06 February 2013 |
| JP 2015-169947 A | 28 September 2015 | JP 6145059 B2 | 07 June 2017 |
| JP 2014-013549 A | 23 January 2014 | None | |
| KR 10-2004-0051351 A | 18 June 2004 | KR 10-0487716 B1 | 03 May 2005 |
| KR 10-2017-0000201 A | 02 January 2017 | KR 10-1702055 B1 | 13 February 2017 |

# 국 제 조 사 보 고 서

국제출원번호

**PCT/KR2020/013525**

**A.** 발명이 속하는 기술분류(국제특허분류(IPC))

**G06F 40/279**(2020.01)i; **G06F 40/205**(2020.01)i; **G06F 40/268**(2020.01)i

**B.** 조사된 분야

조사된 최소문헌(국제특허분류를 기재)

G06F 40/279(2020.01); G06F 17/21(2006.01); G06F 17/27(2006.01); G06F 17/28(2006.01); G06N 3/00(2006.01); G06N 3/04(2006.01)

조사된 기술분야에 속하는 최소문헌 이외의 문헌

한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC
일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC

국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우))

eKOMPASS(특허청 내부 검색시스템) & 키워드: 자연어(natural-language), 품사(word-class), 형태소(morpheme), 식별(percipience), 의미(meaning)

**C.** 관련 문헌

| 카테고리* | 인용문헌명 및 관련 구절(해당하는 경우)의 기재 | 관련 청구항 |
|---|---|---|
| X | JP 2008-293359 A (SHARP CORP.) 2008.12.04<br>단락 [0036]-[0049]; 청구항 1; 및 도면 2 | 1-10 |
| A | JP 2015-169947 A (NIPPON TELEGR & TELEPH CORP.) 2015.09.28<br>단락 [0057]-[0073]; 및 청구항 1 | 1-10 |
| A | JP 2014-013549 A (BRAINPAD INC.) 2014.01.23<br>단락 [0029]-[0033]; 청구항 1; 및 도면 1 | 1-10 |
| A | KR 10-2004-0051351 A (한국전자통신연구원) 2004.06.18<br>페이지 2, 라인 15 - 페이지 3, 라인 6; 청구항 1; 및 도면 2 | 1-10 |
| A | KR 10-2017-0000201 A ((주)아크릴) 2017.01.02<br>단락 [0027]-[0033]; 청구항 1; 및 도면 1 | 1-10 |

☑ 추가 문헌이 C(계속)에 기재되어 있습니다.

☑ 대응특허에 관한 별지를 참조하십시오.

* 인용된 문헌의 특별 카테고리:

"A" 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌

"D" 본 국제출원에서 출원인이 인용한 문헌

"E" 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후에 공개된 선출원 또는 특허 문헌

"L" 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌

"O" 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌

"P" 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌

"T" 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌

"X" 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다.

"Y" 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다.

"&" 동일한 대응특허문헌에 속하는 문헌

| 국제조사의 실제 완료일 | 국제조사보고서 발송일 |
|---|---|
| **2021년01월19일(19.01.2021)** | **2021년01월20일(20.01.2021)** |


ISA/KR의 명칭 및 우편주소

**대한민국 특허청**
**(35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사)**

팩스 번호 **+82-42-481-8578**

심사관

**양정록**

전화번호 **+82-42-481-5709**

# 국 제 조 사 보 고 서

| 국제출원번호 |
| --- |
| **PCT/KR2020/013525** |

**C. 관련 문헌**

| 카테고리* | 인용문헌명 및 관련 구절(해당하는 경우)의 기재 | 관련 청구항 |
| --- | --- | --- |
| X | 이주상 등, 의미 정보와 BERT를 결합한 개념 언어 모델, 제31회 한글 및 한국어 정보처리 학술대회 논문집 (2019년), ISSN 2005-3053, 2019.10.12, pp. 3-7<br><URL: http://hclt.kr/symp/?lnb=conference><br>(본 문헌은 출원인에 의해 `신규성 상실의 예외`가 선언된 문헌임) | 1-10 |

# 국 제 조 사 보 고 서
## 대응특허에 관한 정보

국제출원번호
**PCT/KR2020/013525**

| 국제조사보고서에서 인용된 특허문헌 | 공개일 | 대응특허문헌 | 공개일 |
|---|---|---|---|
| JP 2008-293359 A | 2008/12/04 | JP 5137107 B2 | 2013/02/06 |
| JP 2015-169947 A | 2015/09/28 | JP 6145059 B2 | 2017/06/07 |
| JP 2014-013549 A | 2014/01/23 | 없음 | |
| KR 10-2004-0051351 A | 2004/06/18 | KR 10-0487716 B1 | 2005/05/03 |
| KR 10-2017-0000201 A | 2017/01/02 | KR 10-1702055 B1 | 2017/02/13 |