

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국



(10) 국제공개번호
WO 2012/074338 A2

(43) 국제공개일
2012년 6월 7일 (07.06.2012)

- (51) 국제특허분류:
G06F 17/20 (2006.01)
- (21) 국제출원번호: PCT/KR2011/009333
- (22) 국제출원일: 2011년 12월 2일 (02.12.2011)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보:
10-2010-0122025 2010년 12월 2일 (02.12.2010) KR
10-2010-0132141 2010년 12월 22일 (22.12.2010) KR
10-2010-0133761 2010년 12월 23일 (23.12.2010) KR
10-2010-0138531 2010년 12월 30일 (30.12.2010) KR
10-2011-0001282 2011년 1월 6일 (06.01.2011) KR
10-2011-0014968 2011년 2월 21일 (21.02.2011) KR
- (71) 출원인 (US을(를) 제외한 모든 지정국에 대하여): 에스케이텔레콤 주식회사 (SK TELECOM CO., LTD.) [KR/KR]; 서울 중구 을지로 2가 11번지, 100-999 Seoul (KR).
- (72) 발명자; 겸
- (75) 발명자/출원인 (US에 한하여): 박용길 (PARK, Yong Gil) [KR/KR]; 경기도 성남시 분당구 서현동 효자촌동 아아파트 203-1503, 463-763 Gyeonggi-do (KR). 박근태 (PARK, Keun Tae) [KR/KR]; 경기도 성남시 분당구 서현동 시범단지한양아파트 317동 2802호, 463-776 Gyeonggi-do (KR). 이동학 (LEE, Dong Hahk) [KR/KR]; 서울시 송파구 잠실동 22번지 리젠스아파트 204동 1603호, 138-220 Seoul (KR). 최형인 (CHOI, Hyeongin) [KR/KR]; 서울시 송파구 방이동 217 대림가락아파트 7-601, 138-050 Seoul (KR). 위남숙 (WEE, Nam Sook) [KR/KR]; 서울시 송파구 방이동 217 대림가락아파트 7-601, 138-050 Seoul (KR). 이두석 (LEE, Doo Seok) [KR/KR]; 서울시 구로구 신도림동 동아아파트 201-2203, 152-773 Seoul (KR). 손정교 (SOHN, Jung Kyo) [KR/KR]; 서울시 관악구 삼성동 주공 1차아파트 107-1904, 151-793 Seoul (KR). 김행문 (KIM, Haeng Moon) [KR/KR]; 경기도 과천시 주암동 19-9, 427-070 Gyeonggi-do (KR).
- (74) 대리인: 이철희 (LEE, Chulhee); 서울 강남구 역삼동 647-13 동궁빌딩 5층, 135-911 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, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NL, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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, SZ, TZ, UG, ZM, ZW), 유라시아 (AM, AZ, BY, KG, KZ, MD, 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, ML, MR, NE, SN, TD, TG).

[다음 쪽 계속]

(54) Title: NATURAL LANGUAGE AND MATHEMATICAL FORMULA PROCESSING METHOD AND DEVICE THEREFOR

(54) 발명의 명칭: 자연어 및 수학적 처리 방법과 그를 위한 장치

(57) Abstract: Disclosed are a natural language and mathematical formula processing method and a device therefor. A natural language and mathematical formula processing device comprises: a natural language and mathematical formula input unit which receives a natural language and a mathematical formula; a natural language and mathematical formula structuring unit which analyzes combined data having a combination of said natural language and said mathematical formula, divides the analyzed data according to a particular meaning, and recombines the divided data; a natural language and mathematical formula indexing unit which indexes said combined data; an operation information extracting unit which extracts operation information using a logical condition from said combined data; an operation structuring unit which structures said operation information; and an information generating unit which generates parsing semantic information on the mathematical formula from said combined data. According to the present embodiment, the invention enables: a separate input tool to be provided such that a user is able to input a natural language and a mathematical formula; the natural language and the mathematical formula to be structured such that the natural language and the mathematical formula are managed as recombined data on the basis of the analyzed content of data in which the natural language and the mathematical formula are combined; information, which has structured a user query, to be indexed with the natural language and the mathematical formula on the basis of semantic information; the semantic information to be automatically extracted; a complicated sentence, which includes the natural language and the mathematical formula, to be expressed to automatically have a logical relationship; and semantic information to be generated additionally.

(57) 요약서: 자연어 및 수학적 처리 방법과 그를 위한 장치를 개시한다. 자연어 및 수식을 입력받는 자연어 및 수식 입력부; 상기 자연어 및 상기 수식의 조합으로 이루어진 조합 데이터를 분석하여 특정 의미에 따라 구분한 후 재조합하는 자연어 및 수식 구조화부; 상기 조합 데이터를 색인화하는 자연어 및 수식 색인화부; 상기 조합 데이터에서 논리적 조건을 이용한 동작 정보를 추출하는 동작 정보 추출부; 상기 동작 정보를 구조화하는 동작 구조화부; 및 상기 조합 데이터로부터 수학적식에 대한 파싱 시멘틱 정보를 생성하는 정보 생성부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다. 본 실시예에 의하면, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 입력

[다음 쪽 계속]

WO 2012/074338 A2



공개:

- 국제조사보고서 없이 공개하며 보고서 접수 후 이를 별도 공개함 (규칙 48.2(g))

를 제공하고, 자연어와 수식이 조합된 데이터의 분석 내용에 기초하여 자연어 및 수식을 재조합한 데이터로 관리할 수 있도록 하는 구조화하고, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화되도록 하며, 의미론적인 시멘틱 정보를 자동으로 추출하고, 자연어와 수학적식이 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하며, 추가적으로 시멘틱 정보를 생성할 수 있는 효과가 있다.

명세서

발명의 명칭: 자연어 및 수학적 처리 방법과 그를 위한 장치

기술분야

- [1] 본 실시예는 자연어 및 수학적 처리 방법과 그를 위한 장치에 관한 것이다. 더욱 상세하게는, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 입력 툴을 제공하고, 시멘틱 정보를 생성하며, 의미론적인 시멘틱 정보를 자동으로 추출하며, 자연어와 수식이 조합된 데이터의 분석 내용에 기초하여 자연어 및 수식을 재조합한 데이터로 관리할 수 있도록 하는 구조화하고, 자연어와 수학적식이 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하며, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화되도록 하는 자연어 및 수학적 처리 방법과 그를 위한 장치에 관한 것이다.

배경기술

- [2] 이 부분에 기술된 내용은 단순히 본 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.
- [3] 사람의 언어는 풍부하고 복잡하며, 복잡한 문법 및 문맥 의미를 갖는 엄청난 어휘를 포함하고 있으나 기계 또는 소프트웨어 애플리케이션은 일반적으로 특정 형식 또는 규칙에 따라 데이터를 입력할 것을 요구한다. 여기서, 자연어 입력은 사람과 상호작용하기 위한 거의 모든 소프트웨어 애플리케이션에서 이용될 수 있다. 일반적인 자연어 처리 과정은 자연어를 토큰(Token)으로 분리하고 소프트웨어 애플리케이션에 의해 제공되는 하나 이상의 동작에 매핑하며 각각의 소프트웨어 애플리케이션이 일련의 고유한 동작들을 가지도록 설정된다. 즉, 소프트웨어 개발자가 자연어 입력을 해석하는 코드를 작성하여 입력을 각각의 애플리케이션에 대한 적절한 동작에 매핑하는 방식이다.
- [4] 하지만, 이러한 자연어 처리 방식은 수식을 입력받기 위한 별도의 입력 툴을 제공하지 못하고, 수식을 인식하지 못하며, 자연어 및 수식을 색인화, 구조화하지 못하며, 실제 수학적 내부에 포함된 의미를 파악하지 못하는 문제가 있었다.

발명의 상세한 설명

기술적 과제

- [5] 전술한 문제점을 해결하기 위해 본 실시예는, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 입력 툴을 제공하고, 시멘틱 정보를 생성하며, 의미론적인 시멘틱 정보를 자동으로 추출하며, 자연어와 수식이 조합된 데이터의 분석 내용에 기초하여 자연어 및 수식을 재조합한 데이터로 관리할 수 있도록 하는 구조화하고, 자연어와 수학적식이 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하며, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화되도록 하는 자연어 및 수학적 처리 방법과 그를

위한 장치를 제공하는 데 주된 목적이 있다.

과제 해결 수단

- [6] 전술한 목적을 달성하기 위해 본 실시예의 일 측면에 의하면, 자연어 및 수식을 입력받는 자연어 및 수식 입력부; 상기 자연어 및 상기 수식의 조합으로 이루어진 조합 데이터로부터 수학적식에 대한 파싱 시멘틱 정보를 생성하는 정보 생성부; 상기 조합 데이터에서 논리적 조건을 이용한 동작 정보를 추출하는 동작 정보 추출부; 상기 조합 데이터를 분석하여 특정 의미에 따라 구분한 후 재조합하는 자연어 및 수식 구조화부; 상기 동작 정보를 구조화하는 동작 구조화부; 및 상기 조합 데이터를 색인화하는 자연어 및 수식 색인화부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [7] 또한, 제 1 실시예에 의하면, 자연어를 입력받기 위한 텍스트 입력 툴을 제공하는 제 1 자연어 입력 처리기; 수식을 입력받기 위한 수식 입력 툴을 제공하는 제 1 수식 입력 처리기; 입력된 상기 자연어 및 상기 수식을 취합한 취합 데이터를 전달하는 제 1 정보 처리부; 상기 취합 데이터를 수신하며, 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 1 파싱부; 및 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 저장하는 제 1 데이터 관리부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [8] 또한, 제 2 실시예에 의하면, 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 2 정보 입력부; 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하는 제 2 분리부; 분리된 상기 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하는 제 2 자연어 처리부; 분리된 상기 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하는 제 2 수식 처리부; 및 상기 제 1 정보, 상기 제 2 정보, 상기 자연어 및 상기 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 2 데이터 관리부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [9] 또한, 제 3 실시예에 의하면, 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 3 정보 입력부; 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하고, 분리된 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 3 시멘틱 파서부; 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 3 데이터 관리부; 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하는 제 3 쿼리 파서부; 및 상기 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 상기 시멘틱 색인 정보에 상기 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 제 3

- 색인부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [10] 또한, 제 4 실시예에 의하면, 자연어와 수식이 포함된 복합문장을 수신하는 제 4 정보 입력부; 상기 복합문장에서 상기 자연어 및 상기 수식을 분리하는 제 4 분리부; 분리된 상기 자연어를 토큰화하여 자연어토큰을 생성하는 제 4 자연어처리부; 분리된 상기 수식을 파싱하고 시멘틱의미를 추출하여 수학적토큰을 생성하는 제 4 수식처리부; 자연어와 수학적의 논리적 조건 및 상기 논리적 조건에 대응되는 동작 정보를 결합한 룰을 저장하는 제 4 룰저장부; 및 상기 생성된 자연어토큰과 수학적토큰을 상기 저장된 룰의 논리적 조건과 비교하여 상기 저장된 룰로부터 상기 복합문장의 동작 정보를 추출하는 제 4 동작추출부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [11] 또한, 제 5 실시예에 의하면, 자연어와 수식이 포함된 복합문장을 수신하는 제 5 정보 입력부; 상기 복합문장의 문장구성을 분석하여 수식데이터 및 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하는 제 5 문장분석부; 자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는 동작 정보를 추출하는 제 5 동작추출부; 및 상기 추출된 동작 정보를 상기 수식토큰에 대하여 구조화하는 제 5 동작실행부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [12] 또한, 제 6 실시예에 의하면, 수식이 표현된 수식데이터를 입력받는 제 6 정보 입력부; 상기 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하는 제 6 수식데이터 구조화부; 및 구조화된 상기 오퍼레이터에 대하여 상기 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 상기 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하는 제 6 오퍼레이터 파싱부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치를 제공한다.
- [13] 또한, 본 실시예의 다른 측면에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서, 자연어 및 수식을 입력받는 자연어 및 수식 입력 과정; 상기 자연어 및 상기 수식의 조합으로 이루어진 조합 데이터로부터 수학적식에 대한 파싱 시멘틱 정보를 생성하는 정보 생성 과정; 상기 조합 데이터에서 논리적 조건을 이용한 동작 정보를 추출하는 동작 정보 추출 과정; 상기 조합 데이터를 분석하여 특정 의미에 따라 구분한 후 재조합하는 자연어 및 수식 구조화 과정; 상기 동작 정보를 구조화하는 동작 구조화 과정; 및 상기 조합 데이터를 색인화하는 자연어 및 수식 색인화 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법을 제공한다.
- [14] 또한, 제 1 실시예에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서, 자연어를 입력받기 위한 텍스트 입력 툴을 제공하는 제 1 자연어 입력 처리 과정; 수식을 입력받기 위한 수식 입력 툴을 제공하는 제 1 수식 입력 처리 과정; 입력된 상기 자연어 및 상기 수식을 취합한 취합 데이터를

전달하는 제 1 정보 처리 과정; 상기 취합 데이터를 수신하며, 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 1 파싱 과정; 및 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 저장하는 제 1 데이터 관리 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법을 제공한다.

[15] 또한, 제 2 실시예에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적 처리하는 방법에 있어서, 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 2 정보 입력 과정; 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하는 제 2 분리 과정; 분리된 상기 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하는 제 2 자연어 처리 과정; 분리된 상기 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하는 제 2 수식 처리 과정; 및 상기 제 1 정보, 상기 제 2 정보, 상기 자연어 및 상기 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 2 데이터 관리 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법을 제공한다.

[16] 또한, 제 3 실시예에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적 처리하는 방법에 있어서, 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 3 정보 입력 과정; 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하고, 분리된 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 3 시멘틱 파서 과정; 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 3 데이터 관리 과정; 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하는 제 3 쿼리 파서 과정; 및 상기 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 상기 시멘틱 색인 정보에 상기 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 제 3 색인 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법을 제공한다.

[17] 또한, 제 4 실시예에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적 처리하는 방법에 있어서, 자연어와 수식이 포함된 복합문장을 수신하는 제 4 정보 입력 과정; 상기 복합문장에서 상기 자연어 및 상기 수식을 분리하는 제 4 분리 과정; 분리된 상기 자연어를 토큰화하여 자연어토큰을 생성하는 제 4 자연어처리 과정; 분리된 상기 수식을 파싱하고 시멘틱의 의미를 추출하여 수학적토큰을 생성하는 제 4 수식처리 과정; 자연어와 수학적의 논리적 조건 및 상기 논리적 조건에 대응되는 동작 정보를 결합한 룰을 저장하는 제 4 룰저장 과정; 및 상기 생성된 자연어토큰과 수학적토큰을 상기 저장된 룰의 논리적 조건과 비교하여 상기 저장된 룰로부터 상기 복합문장의 동작 정보를 추출하는 제 4 동작추출 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리

방법을 제공한다.

[18] 또한, 제 5 실시예에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서, 자연어와 수식이 포함된 복합문장을 수신하는 제 5 정보 입력 과정; 상기 복합문장의 문장구성을 분석하여 수식데이터 및 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하는 제 5 문장분석 과정; 자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는 동작 정보를 추출하는 제 5 동작추출 과정; 및 상기 추출된 동작 정보를 상기 수식토큰에 대하여 구조화하는 제 5 동작실행 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법을 제공한다.

[19] 또한, 제 6 실시예에 의하면, 자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서, 수식이 표현된 수식데이터를 입력받는 제 6 정보 입력 과정; 상기 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하는 제 6 수식데이터 구조화 과정; 및 구조화된 상기 오퍼레이터에 대하여 상기 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 상기 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하는 제 6 오퍼레이터 파싱 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법을 제공한다.

발명의 효과

[20] 이상에서 설명한 바와 같이 본 실시예에 의하면, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 입력 툴을 제공하고, 시멘틱 정보를 생성하며, 의미론적인 시멘틱 정보를 자동으로 추출하며, 자연어와 수식이 조합된 데이터의 분석 내용에 기초하여 자연어 및 수식을 재조합한 데이터로 관리할 수 있도록 하는 구조화하고, 자연어와 수학적식이 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하며, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화할 수 있는 효과가 있다.

[21] 또한, 제 1 실시예에 의하면, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 텍스트 입력 툴과 수식 입력 툴을 제공하며, 텍스트 입력 툴과 수식 입력 툴을 통해 입력된 자연어와 수식을 입력받는 효과가 있다. 또한, 본 실시예에 의하면, 텍스트 입력 툴과 수식 입력 툴을 통해 입력된 자연어와 수식에 대해 자연어 처리 및 수식 처리를 함께 수행한 시멘틱 정보를 저장 및 관리할 수 있는 효과가 있다.

[22] 또한, 제 2 실시예에 의하면, 자연어와 수식이 조합된 데이터에 대해 자연어 처리 및 수식 처리를 함께 수행한 분석 내용에 기초하여 자연어 및 수식을 재조합한 데이터로 관리할 수 있도록 하는 효과가 있다. 또한, 제 3 실시예에 의하면, 자연어 처리 및 수식 처리를 함께 수행한 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화되도록 하여 자연어 및 수식이 조합된 데이터의 인덱스를 통해 유사성을 분석하고, 점수화한 순위 결과를

제공할 수 있는 효과가 있다.

- [23] 또한, 제 4 실시예에 의하면, 자연어와 표준화된 수학적 식으로 구성된 수학적 문제에 포함된 의미론적인 시멘틱 정보를 자동으로 추출하는 효과가 있다. 또한, 제 5 실시예에 의하면, 자연어와 수학적 식이 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하는 효과가 있다. 또한, 제 6 실시예에 의하면, 임의의 구조화된 방식으로 입력되는 수학적 식을 파싱(Parsing) 할 때, 수학적 식이 내포하는 시멘틱 정보를 추출하는 효과가 있다.

도면의 간단한 설명

- [24] 도 1은 제 1 실시예에 따른 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도,
 [25] 도 2는 제 1 실시예에 따른 자연어 및 수식 입력 제공 방법을 설명하기 위한 순서도,
 [26] 도 3은 제 1 실시예에 따른 XML의 구성을 나타낸 예시도,
 [27] 도 4는 제 2 실시예에 따른 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도,
 [28] 도 5는 제 2 실시예에 따른 자연어 처리부를 개략적으로 나타낸 블록 구성도,
 [29] 도 6은 제 2 실시예에 따른 수식 처리부를 개략적으로 나타낸 블록 구성도,
 [30] 도 7은 제 2 실시예에 따른 자연어 및 수식 구조화 방법을 설명하기 위한 순서도,
 [31] 도 8은 제 2 실시예에 따른 수식의 트리 형태 표현을 나타낸 예시도,
 [32] 도 9는 제 2 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도,
 [33] 도 10은 제 2 실시예에 따른 자연어 및 수식을 구성하고 있는 정보를 분석하여 특정 의미에 따라 구분하는 방법에 대한 예시도,
 [34] 도 11은 제 3 실시예에 따른 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도,
 [35] 도 12는 제 3 실시예에 따른 자연어 및 수식 색인화 방법을 설명하기 위한 순서도,
 [36] 도 13은 제 3 실시예에 따른 색인화된 쿼리 정보의 순위 결과를 제공하는 방법을 설명하기 위한 순서도,
 [37] 도 14는 제 3 실시예에 따른 시멘틱 정보에 포함된 반전된 파일 구조의 예시도,
 [38] 도 15는 제 3 실시예에 따른 시멘틱 정보에 포함된 인덱스를 불 벡터로 표현한 예시도,
 [39] 도 16은 제 3 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도,
 [40] 도 17은 제 3 실시예에 따른 자연어 및 수식을 구성하고 있는 정보를 분석하여 특정 의미에 따라 구분하는 방법에 대한 예시도이다.

- [41] 도 18은 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도,
- [42] 도 19는 제 4 실시예에 따른 수학 문제가 구성된 형태를 트리 구조로 예시한 도면,
- [43] 도 20은 제 4 실시예에 따른 룰을 생성하는 절차를 예시한 도면,
- [44] 도 21은 제 4 실시예에 따른 룰저장부로 사용되는 룰엔진의 구성 및 동작 정보의 추출과정을 예시한 도면,
- [45] 도 22는 제 4 실시예에 따른 수학 오브젝트를 획득하는 절차에 대해서 간략히 도시한 도면,
- [46] 도 23은 제 4 실시예에 따른 복합문장의 시멘틱 정보 추출방법을 설명하기 위한 순서도,
- [47] 도 24는 제 4 실시예에 따른 룰 매칭에 의해 동작 정보를 추출하는 방법을 예시한 도면,
- [48] 도 25는 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도,
- [49] 도 26은 제 5 실시예에 따른 복합문장의 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도,
- [50] 도 27는 제 5 실시예에 따른 문장분석부를 개략적으로 나타낸 블록 구성도,
- [51] 도 28은 제 5 실시예에 따른 자연어 처리부를 개략적으로 나타낸 블록 구성,
- [52] 도 29는 제 5 실시예에 따른 수식 처리부를 개략적으로 나타낸 블록 구성도,
- [53] 도 30은 제 5 실시예에 따른 복합문장의 논리적 표현 변환방법을 설명하기 위한 순서도,
- [54] 도 31은 제 5 실시예에 따른 복합문장의 트리 형태 표현을 나타낸 예시도,
- [55] 도 32는 제 5 실시예에 따른 복합문장의 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도,
- [56] 도 33은 제 6 실시예에 따른 수학적 및 자연어 처리 장치를 간략하게 나타낸 블록 구성도,
- [57] 도 34 및 도 35는 제 6 실시예에 따른 수식이 표현된 수식데이터를 오퍼레이터 파싱을 수행한 결과를 예시한 도면,
- [58] 도 36은 제 6 실시예에 따른 cMathML 특성을 반영한 노드의 횡단순서를 예시한 도면,
- [59] 도 37은 제 6 실시예에 따른 파싱 시멘틱 정보(b)와 입력된 수식데이터(a)를 결합한 시멘틱 정보 결합 수식데이터를 예시한 도면,
- [60] 도 38은 제 6 실시예에 따른 노드를 횡단하는 도중에 노드간에 데이터를 전달하는 데이터구조를 예시한 도면,
- [61] 도 39는 제 6 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도,
- [62] 도 40은 제 6 실시예에 따른 수학적 시멘틱 정보 생성 방법을 설명하기 위한

순서도이다.

발명의 실시를 위한 형태

- [63] 이하, 첨부된 도면을 참조하여 상세하게 설명한다.
- [64] 한편, 본 실시예에 기재된 자연어 및 수식 처리 장치(100)는 제 1 실시예를 위한 자연어 및 수식 입력부, 제 2 실시예를 위한 자연어 및 수식 구조화부, 제 3 실시예를 위한 자연어 및 수식 색인화부, 제 4 실시예를 위한 동작 정보 추출부, 제 5 실시예를 위한 동작 구조화부 및 제 6 실시예를 위한 정보 생성부를 포함하여 하나의 장치로 구현될 수 있다. 여기서, 자연어 및 수식 입력부는 자연어 및 수식을 입력받고, 자연어 및 수식 구조화부는 자연어 및 상기 수식의 조합으로 이루어진 조합 데이터를 분석하여 특정 의미에 따라 구분한 후 재조합하며, 자연어 및 수식 색인화부는 조합 데이터를 색인화할 수 있다. 동작 정보 추출부는 조합 데이터에서 논리적 조건을 이용한 동작 정보를 추출하고, 동작 구조화부는 동작 정보를 구조화하며, 정보 생성부는 조합 데이터로부터 수학적식에 대한 파싱 시멘틱 정보를 생성할 수 있다.
- [65] 자연어 및 수식 입력부는 자연어를 입력받기 위한 텍스트 입력 툴을 제공하고, 수식을 입력받기 위한 수식 입력 툴을 제공하고, 입력된 자연어 및 수식을 취합한 취합 데이터를 생성하고, 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하고, 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 저장한다. 자연어 및 수식 구조화부는 조합 데이터를 입력받고, 조합 데이터에서 자연어 및 수식을 각각 분리하며, 분리된 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하며, 분리된 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하고, 제 1 정보, 제 2 정보, 자연어 및 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장한다. 자연어 및 수식 색인화부는 조합 데이터를 입력받고, 조합 데이터에서 자연어 및 수식을 각각 분리하고, 분리된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하고, 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하며, 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하며, 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 시멘틱 색인 정보에 키워드 정보를 매칭한 쿼리 색인 정보를 생성한다.
- [66] 동작 정보 추출부는 조합 데이터를 입력받고, 조합 데이터에서 자연어 및 상기 수식을 분리하며, 분리된 자연어를 토큰화하여 자연어토큰을 생성하고, 분리된 수식을 파싱하고 시멘틱의 의미를 추출하여 수학적식토큰을 생성하며, 자연어와 수학적식의 논리적 조건 및 논리적 조건에 대응되는 동작 정보를 결합한 룰을 저장하고, 생성된 자연어토큰과 수학적식토큰을 저장된 룰의 논리적 조건과 비교하여 저장된 룰로부터 조합 데이터에 대한 상기 동작 정보를 추출한다. 동작

구조화부는 조합 데이터를 입력받고, 조합 데이터의 문장구성을 분석하여 수식 및 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하고, 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작 정보를 추출하고, 추출된 동작 정보를 수식토큰에 대하여 구조화한다. 정보 생성부는 수식이 표현된 수식데이터를 입력받고, 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하며, 구조화된 오퍼레이터에 대하여 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성한다.

- [67] 한편, 본 실시예의 구현에 있어서, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 입력 툴을 제공한 후 그 나머지 동작(시멘틱 정보 생성 및 추출, 자연어 및 수식의 구조화 및 색인화 등)이 그 어떠한 순서로 수행해도 무방하나, 바람직하게는 시멘틱 정보를 생성하며, 의미론적인 시멘틱 정보를 자동으로 추출하며, 자연어와 수식이 조합된 데이터의 분석 내용에 기초하여 자연어 및 수식을 제조합한 데이터로 관리할 수 있도록 하는 구조화하고, 자연어와 수학적 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하며, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화하는 것으로 구현될 수 있을 것이다. 즉, 본 실시예는 모두 각각의 독립된 특징을 갖고 있으므로, 어떠한 프로세스를 수행한 경우에만 다음 프로세스 이어지는 것으로 한정되지 않으며, 각각 독립적인 프로세스를 수행할 수 있다.

[68]

[69] <제 1 실시예>

- [70] 이하, 도 1 내지 도 3을 통해서는 자연어 및 수식 입력 제공 방법과 그를 위한 장치인 제 1 실시예에 대해 설명토록 한다.

- [71] 제 1 실시예에 기재된 자연어 및 수식 처리 장치(100)는 자연어(Natural Language)를 입력받기 위한 텍스트(Text) 입력 툴(Tool)과 수식(Mathematical Formula)을 입력받기 위한 수식 입력 툴을 제공하는 장치를 말하며, 자연어 및 수식 처리 장치(100)는 하드웨어 또는 소프트웨어로 구현되어, 서버 또는 단말에 탑재될 수 있다.

- [72] 도 1은 제 1 실시예에 따른 자연어 및 수식 처리 장치를 개략적으로 나타낸 블럭 구성도이다.

- [73] 제 1 실시예에 따른 자연어 및 수식 처리 장치(100)는 제 1 자연어 입력 처리기(110), 제 1 수식 입력 처리기(120), 제 1 이미지 변환부(130), 제 1 정보 처리부(140), 제 1 파싱부(150) 및 제 1 데이터 관리부(160)를 포함한다. 한편, 제 1 실시예에서는 자연어 및 수식 처리 장치(100)가 제 1 자연어 입력 처리기(110), 제 1 수식 입력 처리기(120), 제 1 이미지 변환부(130), 제 1 정보 처리부(140), 제 1 파싱부(150) 및 제 1 데이터 관리부(160)만을 포함하는 것으로 기재하고 있으나, 이는 제 1 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 1 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 1 실시예의

본질적인 특성에서 벗어나지 않는 범위에서 자연어 및 수식 처리 장치(100)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용 가능할 것이다.

[74] 제 1 자연어 입력 처리기(110)는 자연어를 입력받기 위한 텍스트 입력 툴을 제공한다. 제 1 자연어 입력 처리기(110)는 자연어를 입력할 수 있는 별도의 텍스트 입력 툴을 제공한다. 한편, 자연어 및 수식 처리 장치(100)가 외부 서버와 연동하는 경우, 자연어 및 수식 처리 장치(100)의 제 1 자연어 입력 처리기(110)는 해당 서버를 통해 텍스트 입력 툴을 제공할 수 있으며, 자연어 및 수식 처리 장치(100)가 서버 형태로 구현되어 외부 단말기와 연동하는 경우, 자연어 및 수식 처리 장치(100)의 제 1 자연어 입력 처리기(110)는 해당 단말기로 텍스트 입력 툴을 제공할 수 있으며, 자연어 및 수식 처리 장치(100)가 외부 장치와 연동하지 않는 자립형(Stand Alone) 단말기 형태로 구현되는 경우, 자연어 및 수식 처리 장치(100)의 제 1 자연어 입력 처리기(110)는 구비된 표시부를 통해 텍스트 입력 툴을 제공하는 것으로 구현될 수 있다. 또한, 제 1 자연어 입력 처리기(110)에 입력되는 텍스트 정보는 수학 문제, 수식 증명 등을 포함한 수학 콘텐츠 중 텍스트에 해당하는 정보인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 또한, 제 1 자연어 입력 처리기(110)가 제공하는 텍스트 입력 툴을 통해 사용자가 직접 텍스트 정보를 입력할 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버 또는 단말기로부터 자연어에 해당하는 텍스트 정보를 입력받을 수도 있을 것이다.

[75] 제 1 수식 입력 처리기(120)는 수식을 입력받기 위한 수식 입력 툴을 제공한다. 제 1 수식 입력 처리기(120)는 수식 입력 툴을 통해 Math ML(Mathematical Markup Language) 형태의 수식을 입력받는다. 제 1 수식 입력 처리기(120)는 자바 애플릿(Java Applet), 실버 라이트(SilverLight), 액티브 X(Active X) 중 적어도 하나 이상의 기술을 지원하는 툴을 말한다. 한편, 자연어 및 수식 처리 장치(100)가 외부 서버와 연동하는 경우, 자연어 및 수식 처리 장치(100)의 제 1 수식 입력 처리기(120)는 해당 서버를 통해 수식 입력 툴을 제공할 수 있으며, 자연어 및 수식 처리 장치(100)가 서버 형태로 구현되어 외부 단말기와 연동하는 경우, 자연어 및 수식 처리 장치(100)의 제 1 수식 입력 처리기(120)는 해당 단말기로 수식 입력 툴을 제공할 수 있으며, 자연어 및 수식 처리 장치(100)가 외부 장치와 연동하지 않는 자립형(Stand Alone) 단말기 형태로 구현되는 경우, 자연어 및 수식 처리 장치(100)의 제 1 수식 입력 처리기(120)는 구비된 표시부를 통해 수식 입력 툴을 제공하는 것으로 구현될 수 있다. 또한, 제 1 수식 입력 처리기(120)에 입력되는 수식 정보는 수학 문제, 수식 증명 등을 포함한 수학 콘텐츠 중 텍스트에 해당하는 정보인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 또한, 제 1 수식 입력 처리기(120)가 제공하는 수식 입력 툴을 통해 사용자가 직접 수식 정보를 입력할 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버 또는 단말기로부터 자연어에 해당하는 수식 정보를 입력받을 수도 있을 것이다.

- [76] 제 1 이미지 변환부(130)는 제 1 수식 입력 처리기(120)를 통해 입력된 수식을 이미지로 변환한 후 수식 입력 툴을 통해 나타나도록 제어한다. 즉, 제 1 이미지 변환부(130)는 제 1 수식 입력 처리기(120)를 통해 입력된 Math ML 형태의 수식을 이미지로 변환함으로써, 수식의 해상도를 높일 수 있으며, 이를 다시 제 1 수식 입력 처리기(120)의 수식 입력 툴을 통해 나타나도록 제어함으로써, 수식을 입력한 사용자에게 높은 해상도의 수식 이미지를 제공하는 것이다. 여기서, 제 1 이미지 변환부(130)는 제 1 수식 입력 처리기(120)를 통해 입력된 수식을 조합한 형태에서 수식 이미지로 변환할 수 있을 것이다. 즉, 수식 입력 툴에서 입력된 수식을 이미지로 바로 컨버팅할 수 있는 API(Application Programming Interface)를 제공하지 않기 때문에, 제 1 이미지 변환부(130)에서 입력된 Math ML형식의 수식을 이미지로 변환하여 제공함으로써 사용자 경험(User Experience)를 높일 수 있다.
- [77] 제 1 정보 처리부(140)는 입력된 자연어 및 수식을 취합한 취합 데이터를 전달한다. 즉, 제 1 정보 처리부(140)는 제 1 자연어 입력 처리기(110)로부터 자연어를 수신하며, 제 1 수식 입력 처리기(120)로부터 수식을 입력받으며, 이를 취합하여 제 1 파싱부(150)로 전달하는 것이다. 제 1 정보 처리부(140)는 PHP(Personal Hypertext Preprocessor)를 이용하여 취합 데이터를 제 1 파싱부(150)로 전달한다. 즉, 제 1 정보 처리부(140)는 PHP를 이용하여 XML 형태의 취합 데이터를 제 1 파싱부(150)로 전달할 수 있다. 이때, 제 1 파싱부(150)는 어떠한 프로그래밍 언어로든 만들어 질 수 있으며, 소켓(Socket)을 오픈한 상태에서 복수 개의 PHP 연결이 접속할 수 있도록 기다리는 형태로 설정될 수 있다. 여기서, 제 1 파싱부(150)를 통해 출력이 되는 시멘틱 정보는 다시 XML 형태로 저장되거나 필요한 시멘틱 정보 기반으로 저장될 수 있다.
- [78] 제 1 파싱부(150)는 취합 데이터를 수신하며, 취합 데이터에 포함된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성한다. 제 1 파싱부(150)는 자연어와 수식이 조합된 스트링(String)을 자바 스크립트(JavaScript)를 이용하여 파싱(Parsing)한다. 예를 들어서, 제 1 파싱부(150)는 웹(Web)에서 입력된 자연어와 Math ML 형태의 수학이 조합된 스트링을 자바 스크립트 기술을 이용해서 파싱을 시도할 때, 자연어와 수식을 명확히 구분하여 특정 시스템에서 부합하는 형태로 구조화하는 것이다.
- [79] 제 1 파싱부(150)는 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성한다. 제 1 파싱부(150)가 수행하는 프로세스에 대해 구체적으로 설명하자면, 제 1 파싱부(150)는 자연어 및 수식이 입력되면, 자연어를 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분한다. 제 1 파싱부(150)는 자연어를 토큰화(Tokenization)한 자연어 토큰을 생성하고, 자연어 토큰을 근거로 중지 단어(Stop Word)를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링

데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성하고, 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다. 여기서, 토큰(Token)이란 연속된 문장에서 구별할 수 있는 단위를 말하며, 토큰화는 자연어를 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드(Word) 단위로 쪼개는 과정을 말한다. 토큰화에 대해 좀더 구체적으로 설명하자면, 제 1 실시예에서 토큰화는 크게 자연어 토큰화와 수식 토큰화로 구분된다. 자연어 토큰화란 조합 데이터(수학 문제)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 어절(Word)을 자연어 토큰으로 인식하는 과정을 말한다. 각 토큰의 의미를 좀 더 명확히 파악하기 위하여 토큰에 대한 형태소 분석을 추가적으로 수행할 수도 있다. 한편, 수식 토큰화란 조합 데이터(수학 문제)에 포함되는 수식을 파싱(Parsing)한 후 얻게되는 개별 단위 정보를 수식 토큰으로 인식하는 과정을 말한다.

[80] [예제 1]

[81] Find the function value $9y^3 + 8y^2 - 4y - 9$ with $y = -1$

[82] 예를 들어서, [예제 1]에서 자연어 토큰에 해당하는 정보는 'Find', 'the', 'function', 'value', 'with' 가 되며, 수식 토큰은 파싱을 통해서 정보를 추출한 후에 반환되는 값인 다항식(Polynomial), 최고 차수(Max degree=3), 항의 수(Number of terms=4), 조건(Condition) 등이 될 수 있다.

[83] 제 1 파싱부(150)는 자연어를 구성하고 있는 구성 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성하고, 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 여기서, 중지 단어란 문장이나 수식의 분석에 있어서 필요없는 토큰에 해당하는 부분을 제거하기 위해서 미리 정의해 놓은 단어들의 집합을 의미한다. 즉, [예제 1]에서 'the'(이외에도 a나 to 등)는 시스템에서 사전(Dictionary) 형태로 미리 정의되어 있다. 여기서, 사전은 단어의 집합을 포함하는 리스트를 의미한다. 즉, 자연어 토큰을 생성한 후 제 1 파싱부(150)에서는 분석에 필요없는 부분인 중지단어를 제거하는 과정을 수행하게 되는데, 중지 단어 필터링은 수학 문제가 길어질 경우(서술형 문제 등)에 분석 과정에 너무 많은 토큰이 들어가는 것을 방지해 주며, 더불어 시스템의 처리 속도를 향상 시키기 위해 동작한다.

[84] 제 1 파싱부(150)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한 후 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다. 여기서, 동작 정보는 자연어 토큰 또는 수식 토큰을 바탕으로 추출할 수 있는 요약 정보를 의미한다. 예를 들어서, [예제 1]에서 자연어 토큰 또는 수식 토큰을 바탕으로 '풀다(Solve)'라는 동작 정보를 추출할 수 있다. 여기서, 중복 제거 필터링 데이터에서 술어에 해당하는

데이터를 동작 정보와 매칭 저장하는 이유는 조합 데이터(수학 문제)를 스키마(Schema)로 정의하는 과정에서 전체 문장이 의미하는 대표 동작에 대한 정보를 획득하여 이후에 검색 또는 문제간의 연관성(Similarity)을 분석할 때 도움이 되는 도구로 활용하기 위함이다.

- [85] 제 1 파싱부(150)는 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분한다. 제 1 파싱부(150)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단(Traverse) 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 제 1 파싱부(150)는 Math ML(Mathematical Markup Language)로 작성된 수식을 XML 트리 형태로 변환한 후 DOM(Document Object Tree) 형태로 변환한다. 제 1 파싱부(150)는 수식을 구성하는 구성 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색(Depth-First Search) 방식으로 횡단을 실행한다. 한편, 횡단 과정과 깊이 우선 검색에 대해 구체적으로 설명하자면, 일반적으로 수식은 Math ML의 형태를 띄고 있으며, 이는 트리의 형태로 구성이 되며, 이러한, 트리를 횡단하는 과정을 횡단 과정이라 칭하며, 횡단 과정을 수행할 때, 깊이 우선 검색(Depth-First Search)을 사용한다. 이러한, 횡단 과정은 트리의 루트(Root)에서 시작하여 자식 노드까지 들어간 후 모든 자식 노드의 검색이 끝나면 부모 노드로 이동하기 때문에, 자식 노드에서 가지고 있는 정보 모두를 부모 노드로 전달한다. 시간 복잡도 측면에서 엣지(Edge)의 수만큼만 검색을 수행하면 됨으로 효율적이다.
- [86] 제 1 데이터 관리부(160)는 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장한다. 제 1 데이터 관리부(160)는 재조합된 데이터를 문서 데이터로 변환한다.
- [87] 도 2는 제 1 실시예에 따른 자연어 및 수식 입력 제공 방법을 설명하기 위한 순서도이다.
- [88] 자연어 및 수식 처리 장치(100)는 자연어를 입력받기 위한 텍스트 입력 툴과 수식을 입력받기 위한 수식 입력 툴을 제공하며, 텍스트 입력 툴 및 수식 입력 툴을 통해 자연어 및 수식을 입력 받는다(S210). 여기서, 자연어 및 수식 처리 장치(100)가 외부 서버와 연동하는 경우, 자연어 및 수식 처리 장치(100)는 해당 서버를 통해 텍스트 입력 툴 및 수식 입력 툴을 제공할 수 있으며, 자연어 및 수식 처리 장치(100)가 서버 형태로 구현되어 외부 단말기와 연동하는 경우, 자연어 및 수식 처리 장치(100)는 해당 단말기로 텍스트 입력 툴 및 수식 입력 툴을 제공할 수 있으며, 자연어 및 수식 처리 장치(100)가 외부 장치와 연동하지 않는 독립형(Stand Alone) 단말기 형태로 구현되는 경우, 자연어 및 수식 처리 장치(100)는 구비된 표시부를 통해 텍스트 입력 툴 및 수식 입력 툴을 제공하는 것으로 구현될 수 있다. 또한, 자연어 및 수식 처리 장치(100)에 입력되는 자연어 및 수식은 수학 문제, 수식 증명 등을 포함한 수학 콘텐츠 중 텍스트에 해당하는 정보인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 한편, 수식 입력 툴을 통해 입력되는 수식은 Math ML 형태이며, 수식 입력 툴은 자바 애플릿,

실버 라이트, 액티브 X 중 적어도 하나 이상의 기술을 지원하는 툴을 말한다.

- [89] 예를 들어서, 자연어 및 수식 처리 장치(100)가 웹에 적용되어 별도의 외부 서버와 연동하는 경우, 사용자가 웹을 통해 자연어와 수식으로 입력하게 되고, 외부 서버는 웹 요청/응답이나 에이잭스(Ajax) 기술을 통해서 입력된 자연어와 수식을 자연어 및 수식 처리 장치(100)로 전달한다. 텍스트 입력 툴과 수식 입력 툴을 이용한 자연어 및 수식에 대한 사용자 입력이 완료되면, 외부 서버에서 구동되는 PHP가 소켓(Socket) 접속을 통해서 자연어 및 수식 처리 장치(100)에 전달한다. 이때, 전달 형식은 Math ML을 포함한 트리 형태의 데이터로 전달되되, 복수 개의 자연어 및 수식이 조합된 XML 데이터로 전달될 수 있다. 단, XML은 자연어 및 수식 처리 장치(100)에서 이해할 수 있도록 표준화된 형태를 가져야 한다.
- [90] 자연어 및 수식 처리 장치(100)는 수식 입력 툴을 통해 입력된 수식을 이미지로 변환한 후 수식 입력 툴을 통해 나타나도록 제어한다(S220). 즉, 자연어 및 수식 처리 장치(100)는 수식 입력 툴을 통해 입력된 Math ML 형태의 수식을 이미지로 변환함으로써, 수식의 해상도를 높일 수 있으며, 이를 다시 제 1 수식 입력 처리기(120)의 수식 입력 툴을 통해 나타나도록 제어함으로써, 수식을 입력한 사용자에게 높은 해상도의 수식 이미지를 제공하는 것이다. 여기서, 자연어 및 수식 처리 장치(100)는 수식 입력 툴을 통해 입력된 수식을 조합한 형태에서 수식 이미지로 변환할 수 있을 것이다. 즉, 수식 입력 툴에서 입력된 수식을 이미지로 바로 컨버팅할 수 있는 API를 제공하지 않기 때문에, 제 1 이미지 변환부(130)에서 입력된 Math ML형식의 수식을 이미지로 변환하여 제공함으로써 사용자 경험을 높일 수 있다.
- [91] 자연어 및 수식 처리 장치(100)는 입력된 자연어 및 수식을 취합한다(S230). 즉, 자연어 및 수식 처리 장치(100)는 자연어 입력 툴을 통해 자연어를 수신하며, 수식 입력 툴을 통해 수식을 입력받으며, 이를 취합하는 것이다. 자연어 및 수식 처리 장치(100)는 자연어 및 수식을 취합한 취합 데이터에 포함된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성한다(S240). 자연어 및 수식 처리 장치(100)는 자연어와 수식이 조합된 스트링을 자바 스크립트를 이용하여 파싱한다.
- [92] 자연어 및 수식 처리 장치(100)는 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성한다. 자연어 및 수식 처리 장치(100)가 수행하는 프로세스에 대해 구체적으로 설명하자면, 자연어 및 수식 처리 장치(100)는 자연어 및 수식이 입력되면, 자연어를 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분한다. 자연어 및 수식 처리 장치(100)는 자연어를 토큰화한 자연어 토큰을 생성하고, 자연어 토큰을 근거로 중지 단어를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성하고, 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다.

- [93] 즉, 자연어 및 수식 처리 장치(100)는 자연어를 구성하고 있는 구성 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성하고, 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성하며, 자연어 및 수식 처리 장치(100)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한 후 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다.
- [94] 자연어 및 수식 처리 장치(100)는 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분한다. 자연어 및 수식 처리 장치(100)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 자연어 및 수식 처리 장치(100)는 Math ML로 작성된 수식을 XML 트리 형태로 변환한 후 DOM 형태로 변환한다. 제 1 파싱부(150)는 수식을 구성하는 구성 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색 방식으로 횡단을 실행한다.
- [95] 자연어 및 수식 처리 장치(100)로 전달된 자연어 및 수식이 조합된 XML 스트림(Stream)은 해당 데이터를 대기하고 있는 소켓으로 전달되고, 처리단에서 자연어와 수식으로 구분되어서 각각 처리될 수 있다. 즉, 자연어 및 수식 처리 장치(100)는 자연어가 가지고 있는 특성을 기반으로 하여, 주변의 수식과 어떻게 연결할 것인지에 대한 정보를 추출하고, 이를 바탕으로 해당 콘텐츠에서 필요한 시멘틱 정보를 추출할 수 있다. 한편, 자연어 및 수식 처리 장치(100)는 입력된 Math ML 형태의 수식을 표준 형식에 맞추어서 파싱(Parsing)하는 작업을 거쳐서 수식에 관련된 시멘틱 정보를 추출할 수 있다.
- [96] 자연어 및 수식 처리 장치(100)는 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장한다(S250). 제 1 데이터 관리부(160)는 재조합된 데이터를 문서 데이터로 변환한다. 즉, 시멘틱 정보는 시스템의 향후 사용 목적에 맞게 적당한 형태로 DB나 파일 시스템에 저장될 수 있다.
- [97] 도 2에서는 단계 S210 내지 단계 S250을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 제 1 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 1 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 1 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 2에 기재된 순서를 변경하여 실행하거나 단계 S210 내지 단계 S250 중 하나 이상의 단계를 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 2는 시계열적인 순서로 한정되는 것은 아니다.
- [98] 도 3은 제 1 실시예에 따른 XML의 구성을 나타낸 예시도이다.
- [99] 사용자가 자연어 및 수식 처리 장치(100)에서 제공하는 텍스트 입력 툴 및 수식 입력 툴을 이용하여 특정 수학 문제에 대해 입력한 자연어 및 수식을 일반적인

XML 형태로 나타낸 예시도는 도 3과 같다. 즉, 해당 수학 문제는 자연어와 수식이 조합된 형태이기 때문에, XML은 자연어와 수식을 포괄할 수 있도록 작성되어 한다. 즉, XML은 복수 개의 <Text></Text> 부분과 Math ML을 포함하는 <Mathbody></Mathbody>가 중복 사용된다.

[100] 또한, XML은 입력된 수학 문제에 대해 정보를 특정 시스템에서 요구하는 폼에 맞게 변환할 수 있다. 즉, 자연어 및 수식 처리 장치(100)를 통해 입력된 자연어 및 수식을 기계에서 이해할 수 있는 형태로 관리할 수 있으며, 자연어 및 수식에 대해 추출된 시멘틱 정보를 저장 및 관리할 수 있다. 예를 들어서, 사용자가 '이차 방정식'에 대한 수학 문제를 입력하고자 하는 경우, 사용자는 자연어 및 수식 처리 장치(100)가 제공하는 텍스트 입력 툴과 수식 입력 툴을 통해 자연어 및 수식을 각각 입력할 수 있으며, 더 나아가 사용자가 입력한 '이차 방정식'과 연관성이 있는 정보를 추가로 제공받을 수 있을 것이다.

[101]

[102] <제 2 실시예>

[103] 이하, 도 4 내지 도 10을 통해서는 자연어 및 수식 구조화 방법과 그를 위한 장치인 제 2 실시예에 대해 설명토록 한다.

[104] 제 2 실시예에 기재된 자연어 및 수식 처리 장치(100)는 자연어(Natural Language) 및 수식(Mathematical Formula)의 조합으로 이루어진 조합 데이터에서 자연어 및 수식 별로 구조화(DB화) 하기 위한 장치를 말하며, 자연어 및 수식 처리 장치(100)는 하드웨어 또는 소프트웨어로 구현되어, 서버 또는 단말에 탑재될 수 있다.

[105] 도 4는 제 2 실시예에 따른 자연어 및 수식 처리 장치를 개략적으로 나타낸 블럭 구성도이다.

[106] 제 2 실시예에 따른 자연어 및 수식 처리 장치(100)는 제 2 정보 입력부(410), 제 2 분리부(420), 제 2 자연어 처리부(430), 제 2 수식 처리부(440) 및 제 2 데이터 관리부(450)를 포함한다. 한편, 제 2 실시예에서는 자연어 및 수식 처리 장치(100)가 제 2 정보 입력부(410), 제 2 분리부(420), 제 2 자연어 처리부(430), 제 2 수식 처리부(440) 및 제 2 데이터 관리부(450)만을 포함하는 것으로 기재하고 있으나, 이는 제 2 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 2 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 2 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 자연어 및 수식 처리 장치(100)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용 가능할 것이다.

[107] 제 2 정보 입력부(410)는 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는다. 여기서, 조합 데이터는 수학 문제, 수식 증명 등을 포함한 수학 콘텐츠인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 또한, 자연어 및 수식의 조합으로 이루어진 조합 데이터는 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터

자연어 및 수식의 조합으로 이루어진 문서 데이터를 입력받을 수도 있을 것이다. 제 2 분리부(420)는 조합 데이터에서 자연어 및 수식을 각각 분리한다. 즉, 제 2 분리부(420)는 제 2 정보 입력부(410)를 통해 자연어 및 수식의 조합으로 이루어진 조합 데이터가 입력되면, 조합 데이터에 포함된 자연어와 수식을 각각 분리하여 인식하는 것이다.

[108] 제 2 자연어 처리부(430)는 분리된 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분한다. 한편, 제 2 자연어 처리부(430)가 특정 의미를 파악하기 위해 수행하는 동작에 대해 구체적으로 설명하자면, 제 2 자연어 처리부(430)는 자연어를 구성하고 있는 제 1 정보를 분석한 후 문장의 구조 및 포함된 키워드 중 적어도 하나 이상의 정보를 이용하여 특정 의미를 파악할 수 있다. 즉, 제 2 자연어 처리부(430)는 기 설정된 룰(Rule) 기반으로 동작하여 특정 의미를 파악할 수 있으며, 제 2 자연어 처리부(430)가 자연어를 구성하고 있는 제 1 정보를 분석하여 특정 의미에 따라 구분하는 구체적인 방법에 대해서는 도 10을 통해 설명하도록 한다.

[109] 제 2 자연어 처리부(430)는 자연어를 토큰화(Tokenization)한 자연어 토큰을 생성한다. 여기서, 토큰(Token)이란 연속된 문장에서 구별할 수 있는 단위를 말하며, 토큰화는 자연어를 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드(Word) 단위로 쪼개는 과정을 말한다. 토큰화에 대해 좀더 구체적으로 설명하자면, 제 2 실시예에서 토큰화는 크게 자연어 토큰화와 수식 토큰화로 구분된다. 자연어 토큰화란 조합 데이터(수학 문제)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 어절(Word)을 자연어 토큰으로 인식하는 과정을 말한다. 각 토큰의 의미를 좀 더 명확히 파악하기 위하여 토큰에 대한 형태소 분석을 추가적으로 수행할 수도 있다. 한편, 수식 토큰화란 조합 데이터(수학 문제)에 포함되는 수식을 파싱(Parsing)한 후 연계되는 개별 단위 정보를 수식 토큰으로 인식하는 과정을 말한다.

[110] [예제 1]

[111] Find the function value $9y^3 + 8y^2 - 4y - 9$ with $y = -1$

[112] 예를 들어서, [예제 1]에서 자연어 토큰에 해당하는 정보는 'Find', 'the', 'function', 'value', 'with' 가 되며, 수식 토큰은 파싱을 통해서 정보를 추출한 후에 반환되는 값인 다항식(Polynomial), 최고 차수(Max degree=3), 항의 수(Number of terms=4), 조건(Condition) 등이 될 수 있다.

[113] 제 2 자연어 처리부(430)는 자연어 토큰을 근거로 중지 단어(Stop Word)를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성한다. 여기서, 중지 단어란 문장이나 수식의 분석에 있어서 필요없는 토큰에 해당하는 부분을 제거하기 위해서 미리 정의해 놓은 단어들의 집합을 의미한다. 즉, [예제 1]에서 'the'(이외에도 a나 to 등)는 시스템에서 사전(Dictionary) 형태로 미리 정의되어

있다. 여기서, 사전은 단어의 집합을 포함하는 리스트를 의미한다. 즉, 자연어 토큰을 생성한 후 제 2 자연어 처리부(430)에서는 분석에 필요없는 부분인 중지단어를 제거하는 과정을 수행하게 되는데, 중지 단어 필터링은 수학 문제가 길어질 경우(서술형 문제 등)에 분석 과정에 너무 많은 토큰이 들어가는 것을 방지해 주며, 더불어 시스템의 처리 속도를 향상 시키기 위해 동작한다.

- [114] 제 2 자연어 처리부(430)는 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작(Action) 정보를 매칭한다. 여기서, 동작 정보는 자연어 토큰 또는 수식 토큰을 바탕으로 추출할 수 있는 요약 정보를 의미한다. 예를 들어, [예제 1]에서 자연어 토큰 또는 수식 토큰을 바탕으로 '풀다(Solve)'라는 동작 정보를 추출할 수 있다. 여기서, 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 동작 정보와 매칭 저장하는 이유는 조합 데이터(수학 문제)를 스키마(Schema)로 정의하는 과정에서 전체 문장이 의미하는 대표 동작에 대한 정보를 획득하여 이후에 검색 또는 문제간의 연관성(Similarity)을 분석할 때 도움이 되는 도구로 활용하기 위함이다.
- [115] 제 2 자연어 처리부(430)는 자연어를 구성하고 있는 제 1 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 제 2 자연어 처리부(430)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 제 2 자연어 처리부(430)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 제 2 자연어 처리부(430)는 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다.
- [116] 제 2 수식 처리부(440)는 분리된 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분한다. 한편, 제 2 수식 처리부(440)가 특정 의미를 파악하기 위해 수행하는 동작에 대해 구체적으로 설명하자면, 제 2 수식 처리부(440)는 수식을 구성하고 있는 제 2 정보를 분석한 후 수식의 종류 정보를 이용하여 특정 의미를 파악할 수 있다. 즉, 제 2 수식 처리부(440)는 기 설정된 룰(Rule) 기반으로 동작하여 특정 의미를 파악할 수 있으며, 제 2 수식 처리부(440)가 수식을 구성하고 있는 제 2 정보를 분석하여 특정 의미에 따라 구분하는 구체적인 방법에 대해서는 도 10을 통해 설명하도록 한다.
- [117] 제 2 수식 처리부(440)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단(Traverse) 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 제 2 수식 처리부(440)는 Math ML(Mathematical Markup Language)로 작성된 수식을 XML 트리 형태로 변환한 후 DOM(Document Object Model) 형태로 변환한다. 제 2 수식 처리부(440)는 수식을 구성하는 제 2 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색(Depth-First Search) 방식으로 횡단을 실행한다. 한편, 횡단 과정과 깊이 우선 검색에 대해 구체적으로 설명하자면, 일반적으로 수식은 Math ML의 형태를 띄고 있으며,

이는 트리의 형태로 구성이 되며, 이러한, 트리를 횡단하는 과정을 횡단 과정이라 칭하며, 횡단 과정을 수행할 때, 깊이 우선 검색(Depth-First Search)을 사용한다. 이러한, 횡단 과정은 트리의 루트(Root)에서 시작하여 자식 노드까지 들어간 후 모든 자식 노드의 검색이 끝나면 부모 노드로 이동하기 때문에, 자식 노드에서 가지고 있는 정보 모두를 부모 노드로 전달한다. 시간 복잡도 측면에서 엣지(Edge)의 수만큼만 검색을 수행하면 됨으로 효율적이다.

[118] 제 2 데이터 관리부(450)는 제 2 자연어 처리부(430)를 통해 분석된 제 1 정보, 제 2 수식 처리부(440)를 통해 분석된 제 2 정보, 제 2 분리부(420)를 통해 인식된 자연어 및 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장한다. 제 2 데이터 관리부(450)는 재조합된 데이터를 문서 데이터로 변환한다. 한편, 제 2 데이터 관리부(450)는 제 1 정보, 제 2 정보, 자연어 및 수식이 하나의 XML(eXtended Markup Language) 트리로 저장되도록 XML을 정의 할 수 있으나, 제 2 실시예에서 그에 대한 구체적인 예시는 생략토록 한다. 다만, 제 1 정보, 제 2 정보, 자연어 및 수식이 정의된 XML에 대해 개략적으로 설명하자면, 정의된 XML은 형태는 크게 두 부분으로 구분될 수 있는데, 첫 번째는 '문제 묘사' 부분과, 두 번째는 자연어와 수식에서 추출한 정보를 기초로 구성되는 '시멘틱' 부분으로 구분될 수 있다. 여기서, '시멘틱' 부분은 새로운 수학 문제의 형태 발견에 따라 향후에 추가되거나 변경될 수 있을 것이다.

[119] 또한, 수학 문제가 정의된 XML에 대해 설명하자면, 수학 문제는 트리 형태로 구성되고, 그에 필요한 정보가 전체 트리에서 시멘틱 부분으로 집결되는 형태의 구조를 갖도록 구성되어, 향후에 수학 문제에 대한 검색(인덱싱) 등에 이용될 수 있다. 즉, 트리 형태로 구성된 수학 문제에 의하면, 자연어와 표준화된 수식으로 표현된 수학 콘텐츠를 자연어 및 수식 처리 장치(100)가 이해(인식)할 수 있는 형태로 변환하고, 자연어와 수식이 갖는 의미에 근거하여 시멘틱 정보를 추출하여 XML 트리 형태로 구조화할 수 있는 것이다.

[120] 한편, 자연어 및 수식 처리 장치(100)는 자연어 및 수식을 구조화하기 위한 하드웨어 또는 소프트웨어 등의 컴퓨팅 자원을 저장하고, 클라이언트가 필요로 하는 컴퓨팅 자원을 클라우드 컴퓨팅(Cloud Computing)으로 해당 단말기로 제공할 수 있을 것이다. 이와 관련된 구체적인 설명은 도 9을 통해 하도록 한다.

[121] 도 5는 제 2 실시예에 따른 자연어 처리부를 개략적으로 나타낸 블럭 구성도이다.

[122] 제 2 실시예에 따른 제 2 자연어 처리부(430)는 제 2 자연어 토큰화부(510), 제 2 중지 단어 필터링부(520), 제 2 중복 제거 필터링부(530) 및 제 2 동작 매칭부(540)를 포함한다. 한편, 제 2 실시예에서는 제 2 자연어 처리부(430)가 제 2 자연어 토큰화부(510), 제 2 중지 단어 필터링부(520), 제 2 중복 제거 필터링부(530) 및 제 2 동작 매칭부(540)만을 포함하는 것으로 기재하고 있으나, 이는 제 2 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 2 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 2 실시예의

본질적인 특성에서 벗어나지 않는 범위에서 제 2 자연어 처리부(430)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용 가능할 것이다.

[123] 제 2 자연어 토큰화부(510)는 자연어를 토큰화한 자연어 토큰을 생성한다. 제 2 자연어 토큰화부(510)는 자연어를 구성하고 있는 제 1 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 여기서, 자연어 토큰이란 조합 데이터(수학 문제)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 단어를 말한다. 예를 들어서, 자연어 및 수식 처리 장치(100)는 제 2 자연어 토큰화부(510)를 이용하여 조합 데이터에 포함된 자연어 노드를 개별적으로 입력받거나 자연어 노드 전체를 한번에 입력받을 수 있다. 여기서, 자연어는 노드 자체가 복수 개의 단어로 구성된 문장의 성질을 가지지나 완벽한 문장으로 한정되는 것은 아니다. 즉, 이런 자연어 노드를 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드 단위로 쪼개게 되는데, 이러한 과정을 토큰화 과정이라고 한다. 한편, 자연어 노드란 조합 데이터(수학 문제)를 스키마로 구성할 때, 자연어와 수식이 순서와 상관없이 혼재된 형태를 띠게 되는데, 이때, 자연어에 해당하는 부분을 자연어 노드라 한다. 즉, 하나의 문제(즉, 스키마)에 복수 개의 자연어 부분이 포함될 수 있다. [예제 1]는 자연어 노드를 두 개를 포함하는데, 'Find the function value'와 'with'가 자연어 노드가 되는 것이다. 따라서, 시스템에 입력될 경우에는 자연어 노드를 시스템이 이해할 수 있는 단위로 쪼개는 토큰화 과정을 수행하는 것이다. 여기서, 자연어 토큰이란 조합 데이터(수학 문제)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 단어를 말한다.

[124] 제 2 중지 단어 필터링부(520)는 자연어 토큰을 근거로 중지 단어를 필터링한 중지 단어 필터링 데이터를 생성한다. 제 2 중지 단어 필터링부(520)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 여기서, 중지 단어란 문장이나 수식의 분석에 있어서 필요없는 토큰에 해당하는 부분을 제거하기 위해서 미리 정의해 놓은 단어들의 집합을 의미한다. 즉, [예제 1]에서 'the'(이외에도 a나 to 등)는 시스템에서 사전(Dictionary) 형태로 미리 정의되어 있다. 여기서, 사전은 단어의 집합을 포함하는 리스트를 의미한다. 즉, 자연어 토큰을 생성한 후 제 2 자연어 처리부(430)에서는 분석에 필요없는 부분인 중지단어를 제거하는 과정을 수행하게 되는데, 중지 단어 필터링은 수학 문제가 길어질 경우(서술형 문제 등)에 분석 과정에 너무 많은 토큰이 들어가는 것을 방지해 주며, 더불어 시스템의 처리 속도를 향상 시키기 위해 동작한다. 즉, 자연어 및 수식 처리 장치(100)는 제 2 중지 단어 필터링부(520)를 이용하여 토큰화 과정이 수행된 후 자연어를 구성하고 있는 각각의 제 1 정보가 복수 개의 토큰으로 분리되어 자연어 및 수식 처리 장치(100)에 입력되면, 다음 단계로 중지 단어 제거 과정을 거친다. 이 과정에서는 시멘틱 의미를 추출하기 위해 필요 없는 토큰들을 제거하게 된다. 예를 들어서, '이', '저', '여기' 및 '저기' 등이 중지

단어로 설정될 수 있으나 반드시 이에 한정되는 것은 아니며, 의미상 필요 없는 토큰을 설정하는 것은 각 시스템에 따라 다르게 설정될 수 있다.

- [125] 제 2 중복 제거 필터링부(530)는 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성한다. 제 2 중복 제거 필터링부(530)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 즉, 자연어 및 수식 처리 장치(100)는 제 2 중복 제거 필터링부(530)를 이용하여 중지 단어를 필터링한 후 중복을 제거하는 과정을 수행하며, 중복 제거 필터링을 통해 중복된 단어를 제거함으로써 자연어 및 수식 처리 장치(100)의 처리 부하를 낮출 수 있다.
- [126] 제 2 동작 매칭부(540)는 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다. 제 2 동작 매칭부(540)는 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다. 여기서, 동작 정보는 자연어 토큰 또는 수식 토큰을 바탕으로 추출할 수 있는 요약 정보를 의미한다. 예를 들어서, [예제 1]에서 자연어 토큰 또는 수식 토큰을 바탕으로 '풀다(Solve)'라는 동작 정보를 추출할 수 있다. 여기서, 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 동작 정보와 매칭 저장하는 이유는 조합 데이터(수학 문제)를 스키마(Schema)로 정의하는 과정에서 전체 문장이 의미하는 대표 동작에 대한 정보를 획득하여 이후에 검색 또는 문제간의 연관성(Similarity)을 분석할 때 도움이 되는 도구로 활용하기 위함이다. 자연어 및 수식 처리 장치(100)는 제 2 동작 매칭부(540)를 이용하여 선처리 작업(Pre-Processing)을 거쳐 조합 데이터의 특성을 분석하여 기 정의된 의미가 부여된 동작을 토큰과 비교하여 매칭 저장하게 되는 것이다. 즉, 자연어 및 수식 처리 장치(100)는 제 2 동작 매칭부(540)를 이용하여 제 2 자연어 처리부(430)에서 획득한 결과에 근거하여 조합 데이터에 포함된 수식을 '조건'이나 '정의' 등으로 묶거나, 수학 콘텐츠 자체가 의미하는 시멘틱 의미를 파악하는 데 이용할 수 있다.
- [127] 도 6은 제 2 실시예에 따른 수식 처리부를 개략적으로 나타낸 블록 구성도이다.
- [128] 제 2 실시예에 따른 제 2 수식 처리부(440)는 제 2 트리 변환부(610), 제 2 시멘틱 파서부(620) 및 제 2 수식 토큰화부(630)를 포함한다. 한편, 제 2 실시예에서는 제 2 수식 처리부(440)가 제 2 트리 변환부(610), 제 2 시멘틱 파서부(620) 및 제 2 수식 토큰화부(630)만을 포함하는 것으로 기재하고 있으나, 이는 제 2 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 2 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 2 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 제 2 수식 처리부(440)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용 가능할 것이다. 여기서, 시멘틱이란 해당 장치에서 특정 정보의 뜻을 이해하고 논리적 추론 가능하도록 하는 의미이다.
- [129] 자연어 및 수식 처리 장치(100)는 제 2 정보 입력부(410)를 통해 표준화된

형식으로 작성된 개별 수식을 입력받고, 이를 제 2 수식 처리부(440)로 전달하게 된다. 즉, 제 2 수식 처리부(440)로 전달된 수식은 W3C(World Wide Web Consortium)에서 정의된 표준인 Math ML(Mathematical Markup Language)을 기준으로 XML 태그의 형태를 이룬다. 단, 제 2 수식 처리부(440)로 전달된 수식은 Math ML 인 것이 바람직하나 반드시 이에 한정되는 것은 아니다.

- [130] 제 2 트리 변환부(610)는 수식을 트리 형태로 변환한다. 제 2 트리 변환부(610)는 Math ML로 작성된 수식을 XML 트리 형태로 변환한 후 DOM 형태로 변환한다. 자연어 및 수식 처리 장치(100)는 제 2 트리 변환부(610)를 이용하여 수식을 Math ML 형태의 XML 트리로 변환하고, 이 트리는 DOM로 변경되어 프로그램에서 접근 가능한 트리 형태로 변환된다.
- [131] 제 2 시멘틱 파서부(620)는 트리 형태로 변환된 수식에 횡단 과정을 수행한다. 제 2 시멘틱 파서부(620)는 수식을 구성하는 제 2 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색 방식으로 횡단을 실행한다. 자연어 및 수식 처리 장치(100)는 제 2 시멘틱 파서부(620)를 이용하여 수식이 가진 시멘틱 의미를 파악하기 위해서 횡단 과정을 거치는데, 제 2 시멘틱 파서부(620)는 가장 낮은 노드에서 점차 상위 노드로 정보를 전달하는 형태인 깊이 우선 검색으로 횡단을 실행한다. 이에 따라 결과적으로, 제 2 시멘틱 파서부(620)를 통해 수집된 제 2 정보는 모두 최상위 노드에 집결되고, 이러한 정보를 바탕으로 수식의 토큰을 만드는 과정을 거치게 된다. 횡단 과정과 깊이 우선 검색에 대해 구체적으로 설명하자면, 일반적으로 수식은 Math ML의 형태를 띄고 있으며, 이는 트리의 형태로 구성이 되며, 이러한, 트리를 횡단하는 과정을 횡단 과정이라 칭하며, 횡단 과정을 수행할 때, 깊이 우선 검색(Depth-First Search)을 사용한다. 이러한, 횡단 과정은 트리의 루트(Root)에서 시작하여 자식 노드까지 우선 들어간 후 모든 자식 노드의 검색이 끝나면 부모 노드로 이동하기 때문에, 자식 노드에서 가지고 있는 정보를 모두를 부모 노드로 전달한다. 시간 복잡도 측면에서 엣지(Edge)의 수만큼만 검색을 수행하면 됨으로 효율적이다.
- [132] 제 2 수식 토큰화부(630)는 횡단 과정이 수행된 수식에 토큰화를 수행한 수식 토큰을 생성한다. 여기서, 수식 토큰이란 조합 데이터(수학 문제)에 포함된 수식을 파싱(Parsing)한 후 얻게 되는 개별 단위 정보를 말한다. 즉, 토큰화된 수식 토큰은 수학 언어(Mathematics Natural Language)로 이루어진 토큰을 말한다. 한편, 수식 토큰은 자연어 토큰과는 다르게 취급된다. 즉, 제 2 자연어 처리부(430)에서는 자연어 토큰을 바탕으로 동작을 매칭하는 반면, 제 2 수식 처리부(440)에서는 수식 토큰이 결과물이 되며, 향후 수식 토큰은 검색을 통해서 수학 콘텐츠를 찾는 등의 작업에 이용될 수 있다.
- [133] 도 7는 제 2 실시예에 따른 자연어 및 수식 구조화 방법을 설명하기 위한 순서도이다.
- [134] 자연어 및 수식 처리 장치(100)는 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는다(S710). 여기서, 자연어 및 수식의 조합으로 이루어진 조합

데이터는 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 자연어 및 수식의 조합으로 이루어진 문서 데이터를 입력받을 수도 있을 것이다. 자연어 및 수식 처리 장치(100)는 조합 데이터에서 자연어 및 수식을 각각 분리한다(S720). 즉, 자연어 및 수식 처리 장치(100)는 자연어 및 수식의 조합으로 이루어진 조합 데이터가 입력되면, 조합 데이터에 포함된 자연어와 수식을 각각 분리하여 인식하는 것이다.

[135] 자연어 및 수식 처리 장치(100)는 분리된 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하는 프로세스를 처리한다(S730). 즉, 자연어 및 수식 처리 장치(100)는 자연어를 토큰화한 자연어 토큰을 생성하고, 자연어 토큰을 근거로 중지 단어를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성하고, 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다. 자연어 및 수식 처리 장치(100)는 자연어를 구성하고 있는 제 1 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 자연어 및 수식 처리 장치(100)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 자연어 및 수식 처리 장치(100)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 자연어 및 수식 처리 장치(100)는 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다.

[136] 자연어 및 수식 처리 장치(100)는 분리된 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하는 프로세스를 처리한다(S740). 자연어 및 수식 처리 장치(100)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 자연어 및 수식 처리 장치(100)는 Math ML로 작성된 수식을 XML 트리 형태로 변환한 후 DOM 형태로 변환한다. 자연어 및 수식 처리 장치(100)는 수식을 구성하는 제 2 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색 방식으로 횡단을 실행한다.

[137] 자연어 및 수식 처리 장치(100)는 제 1 정보, 제 2 정보, 자연어 및 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장한다(S750). 자연어 및 수식 처리 장치(100)는 재조합된 데이터를 문서 데이터로 변환한다. 즉, 단계 S710 내지 단계 S750을 수행함으로써, 자연어 및 수식 처리 장치(100)를 통해 자연어 및 수식이 재조합된 데이터로 저장되어 관리될 수 있으며, 향후 저장된 재조합 데이터를 이용하여 수식을 검색하거나, 수식에 따른 시멘틱을 추출할 수 있을 것이다.

[138] 도 7에서는 단계 S710 내지 단계 S750을 순차적으로 실행하는 것으로 기재하고

있으나, 이는 제 2 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 2 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 2 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 7에 기재된 순서를 변경하여 실행하거나 단계 S710 내지 단계 S750 중 하나 이상의 단계를 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 7는 시계열적인 순서로 한정되는 것은 아니다.

- [139] 도 8는 제 2 실시예에 따른 수식의 트리 형태 표현을 나타낸 예시도이다.
- [140] 도 8에 도시된 바와 같이, 하나의 수학 콘텐츠가 가질 수 있는 구조를 표현하자면, 루트(Root) 노드에 연결된 자식 노드들은 중요 의미 중 하나인 어순 정보를 그대로 유지한 채 자연어와 수식으로 분리된 형태를 갖게 된다. 또한, 각 자연어는 문장의 연결 순서에 따라 특별한 의미를 가진다. 즉, 많은 수학 콘텐츠는 자연어를 기준으로 수식을 엮는 구조가 대부분이다. 예를 들어서, 하나의 자연어의 뒤에 따라오는 수식이 특정 조건으로서 연결되거나, 뒤에 오는 수식이 정의되었는지 등의 구조를 가질 수 있는 것이다. 각 노드의 자연어가 가지는 의미와 연결 관계 뿐만 아니라, 자연어를 통합하여 시멘틱 의미를 추출할 수 있다. 즉, 수학 콘텐츠에서 요구하는 것이 해당 수식을 푸는 것인지, 설명하는 것인지 등의 동작을 구분하기 위해서는 전체 자연어를 한데 묶어서 의미를 파악하는 것으로 문제의 방향성을 파악하는데 이용될 수 있다.
- [141] 도 9은 제 2 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도이다.
- [142] 제 2 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하기 위해서는 단말기(910), 통신망(920) 및 제 2 클라우드 컴퓨팅 장치(930)를 포함한 시스템이 필요하다.
- [143] 여기서, 단말기(910)는 사용자의 명령 또는 조작에 따라 통신망(920)을 경유하여 각종 데이터를 송수신할 수 있는 단말기를 말하는 것이며, 태블릿 PC(Tablet PC), 랩톱(Laptop), 개인용 컴퓨터(PC: Personal Computer), 스마트폰(Smart Phone), 개인휴대용 정보단말기(PDA: Personal Digital Assistant) 및 이동통신 단말기(Mobile Communication Terminal) 등 중 어느 하나일 수 있다. 또한, 단말기(910)는 통신망(920)을 통하여 데이터 읽고 쓰기 및 저장, 네트워크, 콘텐츠 사용 등의 서비스를 이용할 수 있는 클라우드 컴퓨팅(Cloud Computing)을 지원하는 클라우드 컴퓨팅 단말기가 될 수 있다. 즉, 단말기(910)는 통신망(920)을 경유하여 제 2 클라우드 컴퓨팅 장치(930)에 접속하기 위한 프로그램을 저장하기 위한 메모리, 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비하고 있는 장치를 의미한다. 즉, 단말기(910)는 통신망(920)에 연결되어 제 2 클라우드 컴퓨팅 장치(930)와 서버-클라이언트 통신이 가능하다면 그 어떠한 단말기도 가능하며, 노트북 컴퓨터, 이동통신 단말기, PDA 등 여하한 통신 컴퓨팅 장치를 모두 포함하는 넓은 개념이다. 한편, 단말기(910)는 터치 스크린을 구비한 형태로 제작되는 것이 바람직하나 반드시

이에 한정되는 것은 아니다.

- [144] 단말기(910)는 제 2 클라우드 컴퓨팅 장치(930)를 통해 클라우드 컴퓨팅(Cloud Computing) 방식으로 자연어 및 수식을 구조할 수 있다. 즉, 단말기(910)는 제 2 클라우드 컴퓨팅 장치(930)로부터 클라우드 컴퓨팅 방식으로 자연어 및 수식을 구조화하기 위해 제 2 클라우드 컴퓨팅 장치(930)에 저장된 저장 매체와의 입출력 인터페이스를 제공하는 별도의 입출력 인터페이스부를 포함할 수 있으며, 입출력 인터페이스부를 통해 제 2 클라우드 컴퓨팅 장치(930)에 저장된 저장 매체에 대한 데이터 읽기 및 쓰기가 수행되도록 하는 인터페이스 제어부를 포함할 수 있다. 이에 대해 좀 더 구체적으로 설명하자면, 단말기(910)는 입출력 인터페이스부를 통해 자연어 및 수식의 조합으로 이루어진 조합 데이터를 제 2 클라우드 컴퓨팅 장치(930)로 입력할 수 있고, 제 2 클라우드 컴퓨팅 장치(930)를 통해 조합 데이터에서 자연어 및 수식을 각각 분리되고, 분리된 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하고, 분리된 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하며, 제 1 정보, 제 2 정보, 자연어 및 수식 중 적어도 하나 이상의 정보를 제조합한 제조합 데이터를 생성/저장할 수 있도록 하므로, 실질적으로 단말기(910)에서는 어떠한 어플리케이션의 설치없이도 자연어 및 수식 구조화할 수 있는 것이다.
- [145] 통신망(920)은 인터넷망, 인트라넷망, 이동통신망, 위성 통신망 등 다양한 유무선 통신 기술을 이용하여 인터넷 프로토콜로 데이터를 송수신할 수 있는 망을 말하며, 단말기(910)와 제 2 클라우드 컴퓨팅 장치(930) 간에 데이터를 중계하는 기능을 수행한다. 또한, 통신망(920)은 제 2 클라우드 컴퓨팅 장치(930)와 결합되어 하드웨어, 소프트웨어 등의 컴퓨팅 자원을 저장하고, 클라이언트가 필요로 하는 컴퓨팅 자원을 해당 단말기(910)로 제공할 수 있는 클라우드 컴퓨팅망을 포함할 수 있다.
- [146] 제 2 클라우드 컴퓨팅 장치(930)는 자연어 및 수식 처리 장치(100)를 기반으로 구현될 수 있으며, 클라우드 컴퓨팅으로 단말기(910)를 통해 자연어 및 수식을 구조화할 수 있도록 하기 위해, 단말기(910)로 하여금 제 2 클라우드 컴퓨팅 장치(930)에 저장된 저장 매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 하되, 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력되면, 조합 데이터에서 자연어 및 수식을 각각 분리하며, 분리된 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하고, 분리된 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하며, 제 1 정보, 제 2 정보, 자연어 및 수식 중 적어도 하나 이상의 정보를 제조합한 제조합 데이터를 생성하는 컴퓨터로 읽을 수 있는 기록매체를 저장하며, 해당 기록매체의 일부 데이터만을 단말기(910)로 전송하여, 단말기(910)에서 어플리케이션의 설치없이 자연어 및 수식 구조화할 수 있도록 하는 클라우드 컴퓨팅을 제공할 수 있다. 즉, 제 2 클라우드 컴퓨팅 장치(930)는 클라우드 컴퓨팅 방식으로 자연어 및 수식을 구조화하기 위해 저장 매체를 저장하는 저장부와 단말기(910)로 하여금 저장

매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 하는 클라우드 컴퓨팅부를 추가로 구비할 수 있다.

- [147] 도 10은 제 2 실시예에 따른 자연어 및 수식을 구성하고 있는 정보를 분석하여 특정 의미에 따라 구분하는 방법에 대한 예시도이다.
- [148] 제 2 자연어 처리부(430) 및 제 2 수식 처리부(440)가 특정 의미를 파악하기 위해 수행하는 동작에 대해 구체적으로 설명하자면, 제 2 자연어 처리부(430) 및 제 2 수식 처리부(440)는 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석한 후 문장의 구조, 포함된 키워드 및 수식의 종류 정보 중 적어도 하나 이상의 정보를 이용하여 특정 의미를 파악할 수 있으며, 파악된 특정 의미로 구분된 시멘틱 정보를 생성할 수 있다.
- [149] 제 2 자연어 처리부(430) 및 제 2 수식 처리부(440)는 기 설정된 룰 기반으로 동작하여 특정 의미를 파악할 수 있는 데, 이를 구체적으로 설명하자면, 도 10의 (A)에 도시된 바와 같이, 자연어 및 수식의 조합으로 이루어진 네 개의 수학 문장(P1, P2, P3, P4)이 제 2 정보 입력부(410)를 통해 입력되는 경우, 도 10의 (B)에 도시된 바와 같이, 제 2 자연어 처리부(430) 및 제 2 수식 처리부(440)에 의해 자연어를 구성하고 있는 제 1 정보와 수식을 구성하고 있는 제 2 정보를 분석한(파싱된) 결과가 생성될 수 있다.
- [150] 예를 들어서, P1의 경우, 제 2 자연어 처리부(430)에 의해 자연어를 구성하고 있는 제 1 정보를 분석한 결과 수식명(Name)이 "Find"이고 그 타입은 동사(VB)임을 나타내고, 제 2 수식 처리부(440)에 의해 수식을 구성하고 있는 제 2 정보를 분석한 결과 방정식(Equation)이 맞고(True), 다항식(Polynomial)이 맞음(True)을 나타내며 이를 도 10의 (C)에 도시된 바와 같이, 저장된 룰의 논리적 조건과 비교하면 룰 R1, R2, R3 중에서 R1과 매칭됨을 알 수 있다. 따라서 도 10의 (D)에 도시된 바와 같이, 매칭된 룰로부터 해당 논리적 조건을 만족하는 동작 정보인 "Solve"를 동작 정보로서 추출할 수 있다. 즉, 이러한 경우, P1이 나타내는 특정 의미를 동작 인덱스로 인식하여 구분할 수 있는 것이다.
- [151] 제 2 자연어 처리부(430) 또는 제 2 수식 처리부(440)는 기 저장된 룰의 논리적조건을 만족하는 모든 동작 정보를 추출할 수 있다. 만일, 자연어 토큰과 수학적 토큰 조합이 이루는 논리적 조건이 저장된 룰의 여러 가지 논리적 조건을 만족할 수도 있으며, 이 경우는 하나의 수학 문제가 여러 개의 동작 정보를 포함하고 있는 경우이며, 자연어 토큰과 수학적 토큰 조합이 어떠한 논리적 조건도 만족하지 않는 경우는 해당 복합문장은 룰 생성시 수학기문장(조합 데이터)의 분석에서 누락됐거나 분석과정에 포함되지 않은 항목 또는 잘못된 수학기문장인 경우로 판단할 수 있다. 또한, 제 2 자연어 처리부(430) 또는 제 2 수식 처리부(440)는 자연어 파싱의 결과 생성된 자연어토큰의 대상이 되는 수식을 수학기문장토큰 중에서 매칭시킬 수 있다.

[152]

[153] <제 3 실시예>

- [154] 이하, 도 11 내지 도 17을 통해서는 자연어 및 수식 색인화 방법과 그를 위한 장치인 제 3 실시예에 대해 설명토록 한다.
- [155] 제 3 실시예에 기재된 자연어 및 수식 처리 장치(100)는 자연어(Natural Language) 및 수식(Mathematical Formula)의 조합으로 이루어진 조합 데이터에서 자연어 및 수식 별로 구조화(DB화)할 때, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보를 함께 색인화하기 위한 장치를 말하며, 자연어 및 수식 처리 장치(100)는 하드웨어 또는 소프트웨어로 구현되어, 서버 또는 단말에 탑재될 수 있다.
- [156] 도 11은 제 3 실시예에 따른 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도이다.
- [157] 제 3 실시예에 따른 자연어 및 수식 처리 장치(100)는 제 3 정보 입력부(11110), 제 3 시멘틱 파서부(1120), 제 3 데이터 관리부(1130), 제 3 색인부(1140), 제 3 사용자 쿼리 입력부(1150), 제 3 쿼리 파서부(1160), 제 3 점수화부(1170), 제 3 결과 페이지 제공부(1180), 제 3 저장부(1190) 및 제 3 클라우드 컴퓨팅부(1192)를 포함한다. 한편, 제 3 실시예에서는 자연어 및 수식 처리 장치(100)가 제 3 정보 입력부(11110), 제 3 시멘틱 파서부(1120), 제 3 데이터 관리부(1130), 제 3 색인부(1140), 제 3 사용자 쿼리 입력부(1150), 제 3 쿼리 파서부(1160), 제 3 점수화부(1170), 제 3 결과 페이지 제공부(1180), 제 3 저장부(1190) 및 제 3 클라우드 컴퓨팅부(1192)만을 포함하는 것으로 기재하고 있으나, 이는 제 3 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 3 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 3 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 자연어 및 수식 처리 장치(100)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용 가능할 것이다.
- [158] 제 3 정보 입력부(11110)는 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는다. 여기서, 조합 데이터는 수학 문제, 수식 증명 등을 포함한 수학 콘텐츠인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 또한, 자연어 및 수식의 조합으로 이루어진 조합 데이터는 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 자연어 및 수식의 조합으로 이루어진 문서 데이터를 입력받을 수도 있을 것이다.
- [159] 제 3 시멘틱 파서부(1120)는 조합 데이터에서 자연어 및 수식을 각각 분리하고, 분리된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성한다. 여기서, 시멘틱 정보는 동작 인덱스, 시멘틱 인덱스, 문제 목록 인덱스 중 적어도 하나 이상의 인덱스를 포함하며, 문제 목록은 문제 ID로 정렬될 수 있다. 한편, 제 3 시멘틱 파서부(1120)가 특정 의미를 파악하기 위해 수행하는 동작에 대해 구체적으로 설명하자면, 제 3 시멘틱 파서부(1120)는 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석한 후 문장의 구조, 포함된 키워드 및 수식의 종류 정보 중 적어도 하나 이상의 정보를 이용하여 특정 의미를 파악한다. 즉, 제 3 시멘틱

파서부(1120)는 기 설정된 룰(Rule) 기반으로 동작하여 특정 의미를 파악할 수 있으며, 제 3 시멘틱 파서부(1120)가 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 구체적인 방법에 대해서는 도 17을 통해 설명하도록 한다.

- [160] 또한, 제 3 시멘틱 파서부(1120)가 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하기 위해 수행하는 동작에 대해 구체적으로 설명하자면, 제 3 시멘틱 파서부(1120)는 조합 데이터에서 자연어 및 수식을 각각 분리한다. 즉, 제 3 시멘틱 파서부(1120)는 제 3 정보 입력부(1110)를 통해 자연어 및 수식의 조합으로 이루어진 조합 데이터가 입력되면, 조합 데이터에 포함된 자연어와 수식을 각각 분리하여 인식하는 것이다. 제 3 시멘틱 파서부(1120)는 분리된 자연어를 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분한다. 제 3 시멘틱 파서부(1120)는 자연어를 토큰화(Tokenization)한 자연어 토큰을 생성한다. 여기서, 토큰(Token)이란 연속된 문장에서 구별할 수 있는 단위를 말하며, 토큰화는 자연어를 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드(Word) 단위로 쪼개는 과정을 말한다. 토큰화에 대해 좀더 구체적으로 설명하자면, 제 3 실시예에서 토큰화는 크게 자연어 토큰화와 수식 토큰화로 구분된다. 자연어 토큰화란 조합 데이터(수학 문제)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 어절(Word)을 자연어 토큰으로 인식하는 과정을 말한다. 각 토큰의 의미를 좀 더 명확히 파악하기 위하여 토큰에 대한 형태소 분석을 추가적으로 수행할 수도 있다. 한편, 수식 토큰화란 조합 데이터(수학 문제)에 포함되는 수식을 파싱(Parsing)한 후 얻게되는 개별 단위 정보를 수식 토큰으로 인식하는 과정을 말한다.

[161] [예제 1]

[162] Find the function value $9y^3 + 8y^2 - 4y - 9$ with $y = -1$

[163] 예를 들어서, [예제 1]에서 자연어 토큰에 해당하는 정보는 'Find', 'the', 'function', 'value', 'with' 가 되며, 수식 토큰은 파싱을 통해서 정보를 추출한 후에 반환되는 값인 다항식(Polynomial), 최고 차수(Max degree=3), 항의 수(Number of terms=4), 조건(Condition) 등이 될 수 있다.

[164] 제 3 시멘틱 파서부(1120)는 자연어 토큰을 근거로 중지 단어(Stop Word)를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성한다. 여기서, 중지 단어란 문장이나 수식의 분석에 있어서 필요없는 토큰에 해당하는 부분을 제거하기 위해서 미리 정의해 놓은 단어들의 집합을 의미한다. 즉, [예제 1]에서 'the'(이외에도 a나 to 등)는 시스템에서 사전(Dictionary) 형태로 미리 정의되어 있다. 여기서, 사전은 단어의 집합을 포함하는 리스트를 의미한다. 즉, 자연어 토큰을 생성한 후 제 3 시멘틱 파서부(1120)에서는 분석에 필요없는 부분인 중지단어를 제거하는 과정을 수행하게 되는데, 중지 단어 필터링은 수학 문제가

길어질 경우(서술형 문제 등)에 분석 과정에 너무 많은 토큰이 들어가는 것을 방지해 주며, 더불어 시스템의 처리 속도를 향상 시키기 위해 동작한다.

- [165] 제 3 시멘틱 파서부(1120)는 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다. 여기서, 동작 정보는 자연어 토큰 또는 수식 토큰을 바탕으로 추출할 수 있는 요약 정보를 의미한다. 예를 들어서, [예제 1]에서 자연어 토큰 또는 수식 토큰을 바탕으로 '풀다(Solve)'라는 동작 정보를 추출할 수 있다. 여기서, 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 동작 정보와 매칭 저장하는 이유는 조합 데이터(수학 문제)를 스키마(Schema)로 정의하는 과정에서 전체 문장이 의미하는 대표 동작에 대한 정보를 획득하여 이후에 검색 또는 문제간의 연관성(Similarity)을 분석할 때 도움이 되는 도구로 활용하기 위함이다.
- [166] 제 3 시멘틱 파서부(1120)는 자연어를 구성하고 있는 구성 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 제 3 시멘틱 파서부(1120)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 제 3 시멘틱 파서부(1120)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 제 3 시멘틱 파서부(1120)는 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다.
- [167] 제 3 시멘틱 파서부(1120)는 분리된 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분한다. 제 3 시멘틱 파서부(1120)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단(Traverse) 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 제 3 시멘틱 파서부(1120)는 Math ML(Mathematical Markup Language)로 작성된 수식을 XML 트리 형태로 변환한 후 DOM(Document Object Model) 형태로 변환한다. 제 3 시멘틱 파서부(1120)는 수식을 구성하는 구성 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색(Depth-First Search) 방식으로 횡단을 실행한다. 한편, 횡단 과정과 깊이 우선 검색에 대해 구체적으로 설명하자면, 일반적으로 수식은 Math ML의 형태를 띄고 있으며, 이는 트리의 형태로 구성이 되며, 이러한, 트리를 횡단하는 과정을 횡단 과정이라 칭하며, 횡단 과정을 수행할 때, 깊이 우선 검색(Depth-First Search)을 사용한다. 이러한, 횡단 과정은 트리의 루트(Root)에서 시작하여 자식 노드까지 들어간 후 모든 자식 노드의 검색이 끝나면 부모 노드로 이동하기 때문에, 자식 노드에서 가지고 있는 정보 모두를 부모 노드로 전달한다. 시간 복잡도 측면에서 엣지(Edge)의 수만큼만 검색을 수행하면 됨으로 효율적이다.
- [168] 제 3 데이터 관리부(1130)는 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장한다. 제 3 데이터 관리부(1130)는 재조합된 데이터를 문서 데이터로 변환한다. 제 3

색인부(1140)는 제 3 시멘틱 파서부(1120) 및 제 3 데이터 관리부(1130)를 통해 수신된 시멘틱 정보에 번호를 부여하는 색인화(Indexing)를 수행하며, 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 시멘틱 색인 정보에 키워드 정보를 매칭한 쿼리 색인 정보를 생성한다.

[169] 즉, 제 3 정보 입력부(11110)를 통해 입력된 조합 데이터에 포함된 XML 형태의 구조의 콘텐츠(Content) 기반의 Math ML인 수식이 제 3 시멘틱 파서부(1120)에 입력되며, XML 입력을 바탕으로 자연어와 수식의 시멘틱 정보를 추출하고, 제 3 데이터 관리부(1130)에 의해 다시 XML 결과로서 도출하는 것이다. 즉, 시멘틱 정보를 포함한 XML 결과는 제 3 색인부(1140)에 의해 색인화 과정을 거친 후 색인된다.

[170] 제 3 사용자 쿼리 입력부(1150)는 입력된 사용자 쿼리(Query)를 제 3 쿼리 파서부(1160)로 전달한다. 여기서, 사용자 쿼리는 일종의 검색 쿼리로서, 사용자가 검색하고자 입력한 키워드를 포함한다. 제 3 쿼리 파서부(1160)는 입력된 사용자 쿼리에 포함된 키워드(Key Word)를 추출하여 구조화한다. 제 3 점수화부(1170)는 키워드와 시멘틱 색인 정보의 연관성에 근거하여 쿼리 색인 정보를 점수화(Scoring)한다. 제 3 점수화부(1170)는 점수화를 수행하기 위해 코사인 유사성(Cosine Similarity)을 이용한다. 또한, 제 3 점수화부(1170)는 [수학식 1]을 이용하여 점수화를 수행할 수 있다.

[171] 수학식 1

$$\cos(q, p) = \frac{q \cdot p}{\|q\| \|p\|} = \frac{q}{\|q\|} \cdot \frac{p}{\|p\|} = \frac{\sum_{i=1}^{|V|} q_i p_i}{\sqrt{\sum_{i=1}^{|V|} q_i^2} \sqrt{\sum_{i=1}^{|V|} p_i^2}}$$

[172] (p: 문제 벡터, q: 쿼리 벡터, p_i: 불(Boolean)/쿼리 q에서 i의 가중치, p_i: 불(Boolean)/쿼리 p에서 i의 가중치, v: 벡터가 가지는 원소의 개수)

[173] 제 3 결과 페이지 제공부(1180)는 제 3 점수화부(1170)에 의해 점수화된 쿼리 색인 정보의 순위(Ranking) 결과 페이지를 제공한다. 여기서, 제 3 결과 페이지 제공부(1180)는 순위 결과 페이지를 요청하는 서버 또는 단말기로 순위 결과 페이지를 제공할 수 있으나 반드시 이에 한정되는 것은 아니며, 자연어 및 수식 처리 장치(100)가 자립형(Stand Alone) 장치로 구현되는 경우, 구비된 표시부를 통해 해당 순위 결과 페이지를 나타낼 수 있을 것이다.

[174] 즉, 제 3 사용자 쿼리 입력부(1150)를 통해 입력된 사용자 쿼리는 제 3 쿼리 파서부(1160)에서 파싱되고, 제 3 색인부(1140)로 전달되고, 제 3 점수화부(1170)는 기 저장된 수학 콘텐츠에 대한 인덱스와 사용자 쿼리의 인덱스를 토대로 연관성을 비교하여 점수화를 수행하며, 제 3 결과 페이지 제공부(1180)는 사용자 결과 페이지에 순위를 출력하는 것이다.

[175] 한편, 자연어 및 수식 처리 장치(100)는 클라이언트에 해당하는 단말기에

애플리케이션의 설치없이 자연어와 수식이 조합된 데이터를 구조화할 때 사용자 쿼리를 구조화한 정보가 함께 색인화되도록 하는 클라우드 컴퓨팅을 제공하기 위해 별도의 제 3 저장부(1190)와 제 3 클라우드 컴퓨팅부(1192)를 포함할 수 있다. 여기서, 제 3 저장부(1190)는 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 경우, 조합 데이터에서 자연어 및 수식을 각각 분리하고, 분리된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하고, 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하며, 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하며, 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 시멘틱 색인 정보에 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 저장 매체를 저장하며, 제 3 클라우드 컴퓨팅부(1192)는 클라이언트에 해당하는 단말기로 하여금 제 3 저장부(1190)에 저장된 저장 매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 한다.

- [176] 즉, 자연어 및 수식 처리 장치(100)는 제 3 저장부(1190)와 제 3 클라우드 컴퓨팅부(1192)를 통해 자연어와 수식이 조합된 데이터를 구조화할 때 사용자 쿼리를 구조화한 정보가 함께 색인화되도록 하는 하드웨어 또는 소프트웨어 등의 컴퓨팅 자원을 저장하고, 클라이언트가 필요로 하는 컴퓨팅 자원을 클라우드 컴퓨팅(Cloud Computing)으로 해당 단말기로 제공할 수 있다. 이와 관련된 구체적인 설명은 도 16을 통해 하도록 한다.
- [177] 도 12는 제 3 실시예에 따른 자연어 및 수식 색인화 방법을 설명하기 위한 순서도이다.
- [178] 자연어 및 수식 처리 장치(100)는 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는다(S1210). 여기서, 자연어 및 수식의 조합으로 이루어진 조합 데이터는 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 자연어 및 수식의 조합으로 이루어진 문서 데이터를 입력받을 수도 있을 것이다.
- [179] 자연어 및 수식 처리 장치(100)는 조합 데이터에서 자연어 및 수식을 각각 분리하고, 분리된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성한다(S1220). 더 구체적으로 설명하자면, 자연어 및 수식 처리 장치(100)는 조합 데이터에서 자연어 및 수식을 각각 분리한다. 즉, 자연어 및 수식 처리 장치(100)는 자연어 및 수식의 조합으로 이루어진 조합 데이터가 입력되면, 조합 데이터에 포함된 자연어와 수식을 각각 분리하여 인식하는 것이다. 자연어 및 수식 처리 장치(100)는 분리된 자연어를 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 프로세스를 처리한다. 즉, 자연어 및 수식 처리 장치(100)는 자연어를 토큰화한 자연어 토큰을 생성하고, 자연어 토큰을 근거로 중지 단어를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성하고, 중복 제거 필터링 데이터에 기

정의된 의미가 부여된 동작 정보를 매칭한다. 자연어 및 수식 처리 장치(100)는 자연어를 구성하고 있는 구성 정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 자연어 및 수식 처리 장치(100)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 자연어 및 수식 처리 장치(100)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 자연어 및 수식 처리 장치(100)는 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭 저장한다. 자연어 및 수식 처리 장치(100)는 분리된 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 프로세스를 처리한다.

- [180] 자연어 및 수식 처리 장치(100)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 자연어 및 수식 처리 장치(100)는 Math ML로 작성된 수식을 XML 트리 형태로 변환한 후 DOM 형태로 변환한다. 자연어 및 수식 처리 장치(100)는 수식을 구성하는 구성 정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색 방식으로 횡단을 실행한다.
- [181] 자연어 및 수식 처리 장치(100)는 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 제조합하여 제조합 데이터로 저장한다(S1230). 자연어 및 수식 처리 장치(100)는 제조합된 데이터를 문서 데이터로 변환한다. 자연어 및 수식 처리 장치(100)는 시멘틱 정보를 색인화 한다(S1240). 예를 들어, 자연어 및 수식 처리 장치(100)는 시멘틱 정보에 번호를 부여하는 색인화를 수행하는 것이다.
- [182] 도 12에서는 단계 S1210 내지 단계 S1240을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 제 3 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 3 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 3 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 12에 기재된 순서를 변경하여 실행하거나 단계 S1210 내지 단계 S1240 중 하나 이상의 단계를 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 12는 시계열적인 순서로 한정되는 것은 아니다.
- [183] 전술한 바와 같이 도 12에 기재된 제 3 실시예에 따른 자연어 및 수식 색인화 방법은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 제 3 실시예에 따른 자연어 및 수식 색인화 방법을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 이러한 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한

컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 제 3 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드 및 코드 세그먼트들은 제 3 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.

- [184] 도 13은 제 3 실시예에 따른 색인화된 쿼리 정보의 순위 결과를 제공하는 방법을 설명하기 위한 순서도이다.
- [185] 자연어 및 수식 처리 장치(100)는 사용자 쿼리를 입력받는다(S1310). 여기서, 사용자 쿼리는 일종의 검색 쿼리로서, 사용자가 검색하고자 입력한 키워드를 포함한다. 자연어 및 수식 처리 장치(100)는 입력된 사용자 쿼리에 포함된 키워드(Key Word)를 추출하여 구조화한다(S1320). 자연어 및 수식 처리 장치(100)는 시멘틱 정보를 색인화한 시멘틱 색인 정보에 키워드 정보를 매칭한 쿼리 색인 정보를 생성한다(S1330).
- [186] 자연어 및 수식 처리 장치(100)는 키워드와 시멘틱 색인 정보의 연관성에 근거하여 쿼리 색인 정보를 점수화한다(S1340). 제 3 점수화부(1170)는 점수화를 수행하기 위해 코사인 유사성을 이용한다. 또한, 제 3 점수화부(1170)는 [수학식 1]을 이용하여 점수화를 수행할 수 있다. 자연어 및 수식 처리 장치(100)는 제 3 점수화부(1170)에 의해 점수화된 쿼리 색인 정보의 순위 결과 페이지를 제공한다(S1350). 여기서, 제 3 결과 페이지 제공부(1180)는 순위 결과 페이지를 요청하는 서버 또는 단말기로 순위 결과 페이지를 제공할 수 있으나 반드시 이에 한정되는 것은 아니며, 자연어 및 수식 처리 장치(100)가 자립형(Stand Alone) 장치로 구현되는 경우, 구비된 표시부를 통해 해당 순위 결과 페이지를 나타낼 수 있을 것이다.
- [187] 도 13에서는 단계 S1310 내지 단계 S1350을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 제 3 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 3 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 3 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 13에 기재된 순서를 변경하여 실행하거나 단계 S1310 내지 단계 S1350 중 하나 이상의 단계를 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 13은 시계열적인 순서로 한정되는 것은 아니다.
- [188] 도 14는 제 3 실시예에 따른 시멘틱 정보에 포함된 반전된 파일 구조의 예시도이다.
- [189] 자연어 및 수식 처리 장치(100)의 제 3 시멘틱 파서부(1120)를 통해 생성된 시멘틱 정보에 포함된 반전된 파일 구조(Inverted File Structure)의 인덱스는 도 14에 도시된 바와 같다. 한편, 제 3 실시예에서는 반전된 파일 구조에서 이용된 시멘틱 정보의 XML의 형태에서는 언급하지 않도록 한다. 다만, 함수 형태, 동작, 시멘틱 키워드는 모두 계층 구조의 형태로 저장되어 있는 것으로 가정한다. 즉, 시멘틱 정보는 동작 인덱스, 시멘틱 인덱스, 문제 목록 인덱스 중 적어도 하나

이상의 인덱스를 포함하며, 문제 목록은 문제 ID로 정렬된다. 이에 따라, 따라서 두 개의 목록은 선형 시간에 의해 결합(Merge)될 수 있다

[190] 도 15는 제 3 실시예에 따른 시멘틱 정보에 포함된 인덱스를 불 벡터로 표현한 예시도이다.

[191] 자연어 및 수식 처리 장치(100)의 제 3 점수화부(1170)는 점수화를 수행하기 위해 코사인 유사성(Cosine Similarity)을 이용할 수 있다. 즉, 시멘틱 정보에 포함된 인덱스를 불 벡터(Boolean Vector)로 표현하자면, 도 15에 도시된 바와 같다. 여기서, '0'의 값은 해당 행에서 일치하는 '용어'나 '키워드'가 존재하지 않거나, 그 열에 있는 문제와 관련이 없는 것을 의미한다. 반면, '1'의 값은 그 행에 일치하는 '용어'나 '키워드'가 존재하거나, 그 열에 있는 문제와 관련성이 없는 것을 나타낸다. 이러한 행렬을 이용하는 경우, 두 개의 문제 벡터 p 와 쿼리 벡터 q 사이의 코사인 각을 산출할 수 있으며, 코사인 각을 산출하는 식은 [수학식 1]과 같다.

[192] 즉, [수학식 1]에서 $\cos(q,p)$ 는 q 와 p 의 코사인 유사성, 또는 q 와 p 의 코사인 각도라 한다. 코사인은 '0°', '180°'에서 단조 감소하는 함수이기 때문에, 해당 값이 작거나 크면 두 문제는 유사성을 가진다고 할 수 있다. 또한, 불(Boolean) 형태 대신 가중치가 적용될 수도 있다. 예를 들어서, 시멘틱 정보 중 중요한 의미를 갖는 동작(Action)이나 수학 오브젝트(Math Object) 등에는 더 많은 가중치를 부여할 수 있다. 또한, 함수 중에서 상대적으로 빈번하지 않은 함수는 빈번한 함수에 비해 더 작은 가중치가 부여되도록 할 수 있다. 이러한 내용은 다음과 같이 공식화될 수 있다.

[193] 즉, 문제 빈도는(Problem Frequency)는 '용어'와 '키워드'가 주어진 문제의 개수이며, 해당 값은 용어 정보의 반대의 값이며, 해당 값을 나타내기 위해서, 반대 문제 빈도(ipf, Inverse Problem Frequency)가 사용된다. 여기서, ipf는 N/pf 로 계산될 수 있고, 이때 N 은 전체 문제의 수를 나타낸다. 사용자 쿼리와 분석된 자연어 및 수식의 조합으로 이루어진 조합 데이터(수학 콘텐츠)들의 인덱스를 통해서, 유사성을 분석하고, 순위를 산정하여 얻어진 순서대로 표시부를 통해 출력할 수 있다. 이에 따라, 사용자의 쿼리에 가장 가까운 수식 포함 문서부터 그와 유사한 문서를 포함하여 확인을 할 수 있는 것이다.

[194] 도 16은 제 3 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도이다.

[195] 제 3 실시예에 따른 클라우드 컴퓨팅을 이용한 데이터를 제공하기 위해서는 단말기(910), 통신망(920) 및 제 3 클라우드 컴퓨팅 장치(1600)를 포함한 시스템이 필요하다.

[196] 여기서, 단말기(910)는 사용자의 명령 또는 조작에 따라 통신망(920)을 경유하여 각종 데이터를 송수신할 수 있는 단말기를 말하는 것이며, 태블릿 PC, 랩톱, 개인용 컴퓨터, 스마트폰, 개인휴대용 정보단말기 및 이동통신 단말기 등 중 어느 하나일 수 있다. 또한, 단말기(910)는 통신망(920)을 통하여 데이터 읽고

쓰기 및 저장, 네트워크, 콘텐츠 사용 등의 서비스를 이용할 수 있는 클라우드 컴퓨팅을 지원하는 클라우드 컴퓨팅 단말기가 될 수 있다. 즉, 단말기(910)는 통신망(920)을 경유하여 제 3 클라우드 컴퓨팅 장치(1600)에 접속하기 위한 프로그램을 저장하기 위한 메모리, 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비하고 있는 장치를 의미한다. 즉, 단말기(910)는 통신망(920)에 연결되어 제 3 클라우드 컴퓨팅 장치(1600)와 서버-클라이언트 통신이 가능하다면 그 어떠한 단말기도 가능하며, 노트북 컴퓨터, 이동통신 단말기, PDA 등 여하한 통신 컴퓨팅 장치를 모두 포함하는 넓은 개념이다. 한편, 단말기(910)는 터치 스크린을 구비한 형태로 제작되는 것이 바람직하나 반드시 이에 한정되는 것은 아니다.

- [197] 단말기(910)는 제 3 클라우드 컴퓨팅 장치(1600)를 통해 클라우드 컴퓨팅 방식으로 자연어와 수식이 조합된 데이터를 구조화할 때 사용자 쿼리를 구조화한 정보가 함께 색인화되도록 한다. 즉, 단말기(910)는 제 3 클라우드 컴퓨팅 장치(1600)로부터 클라우드 컴퓨팅 방식으로 자연어 및 수식을 구조화하기 위해 제 3 클라우드 컴퓨팅 장치(1600)에 저장된 저장 매체와의 입출력 인터페이스를 제공하는 별도의 입출력 인터페이스부를 포함할 수 있으며, 입출력 인터페이스부를 통해 제 3 클라우드 컴퓨팅 장치(1600)에 저장된 저장 매체에 대한 데이터 읽기 및 쓰기가 수행되도록 하는 인터페이스 제어부를 포함할 수 있다. 이에 대해 좀 더 구체적으로 설명하자면, 단말기(910)는 입출력 인터페이스부를 통해 자연어 및 수식의 조합으로 이루어진 조합 데이터를 제 3 클라우드 컴퓨팅 장치(1600)로 입력할 수 있고, 이를 통해 제 3 클라우드 컴퓨팅 장치(1600)가 시멘틱 색인 정보에 키워드 정보를 매칭한 쿼리 색인 정보를 생성/저장되도록 하므로, 실질적으로 단말기(910)에서는 어떠한 애플리케이션의 설치없이도 자연어와 수식이 조합된 데이터를 구조화할 때 사용자 쿼리를 구조화한 정보가 함께 색인화되도록 한다.
- [198] 통신망(920)은 인터넷망, 인트라넷망, 이동통신망, 위성 통신망 등 다양한 유무선 통신 기술을 이용하여 인터넷 프로토콜로 데이터를 송수신할 수 있는 망을 말하며, 단말기(910)와 제 3 클라우드 컴퓨팅 장치(1600) 간에 데이터를 중계하는 기능을 수행한다. 또한, 통신망(920)은 제 3 클라우드 컴퓨팅 장치(1600)와 결합되어 하드웨어, 소프트웨어 등의 컴퓨팅 자원을 저장하고, 클라이언트가 필요로 하는 컴퓨팅 자원을 해당 단말기(910)로 제공할 수 있는 클라우드 컴퓨팅망을 포함할 수 있다.
- [199] 제 3 클라우드 컴퓨팅 장치(1600)는 자연어 및 수식 처리 장치(100)를 기반으로 구현될 수 있으며, 클라우드 컴퓨팅으로 단말기(910)를 통해 자연어와 수식이 조합된 데이터를 구조화할 때 사용자 쿼리를 구조화한 정보가 함께 색인화되도록 하기 위해, 단말기(910)로 하여금 제 3 클라우드 컴퓨팅 장치(1600)에 저장된 저장 매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 하되, 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력되면, 조합

데이터에서 자연어 및 수식을 각각 분리하고, 분리된 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하고, 구성 정보, 자연어, 수식 및 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하며, 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하며, 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 시멘틱 색인 정보에 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 컴퓨터로 읽을 수 있는 기록매체를 저장하며, 해당 기록매체의 일부 데이터만을 단말기(910)로 전송하여, 단말기(910)에서 애플리케이션의 설치없이 자연어와 수식이 조합된 데이터를 구조화할 때 사용자 쿼리를 구조화한 정보가 함께 색인화되도록 하는 클라우드 컴퓨팅을 제공할 수 있다.

- [200] 도 17은 제 3 실시예에 따른 자연어 및 수식을 구성하고 있는 정보를 분석하여 특정 의미에 따라 구분하는 방법에 대한 예시도이다.
- [201] 제 3 시멘틱 파서부(1120)가 특정 의미를 파악하기 위해 수행하는 동작에 대해 구체적으로 설명하자면, 제 3 시멘틱 파서부(1120)는 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석한 후 문장의 구조, 포함된 키워드 및 수식의 종류 정보 중 적어도 하나 이상의 정보를 이용하여 특정 의미를 파악할 수 있으며, 파악된 특정 의미로 구분된 시멘틱 정보를 생성할 수 있다.
- [202] 제 3 시멘틱 파서부(1120)는 기 설정된 룰 기반으로 동작하여 특정 의미를 파악할 수 있는 데, 이를 구체적으로 설명하자면, 도 17의 (A)에 도시된 바와 같이, 자연어 및 수식의 조합으로 이루어진 네 개의 수학 문장(P1, P2, P3, P4)이 제 3 정보 입력부(11110)를 통해 입력되는 경우, 도 17의 (B)에 도시된 바와 같이, 제 3 시멘틱 파서부(1120)에 의해 자연어와 수식을 구성하고 있는 각각의 구성 정보를 분석한(파싱된) 결과가 생성될 수 있다.
- [203] 예를 들어서, P1의 경우, 제 3 시멘틱 파서부(1120)에 의해 자연어를 구성하고 있는 구성 정보를 분석한 결과 수식명(Name)이 "Find"이고 그 타입은 동사(VB)임을 나타내고, 제 3 시멘틱 파서부(1120)에 의해 수식을 구성하고 있는 구성 정보를 분석한 결과 방정식(Equation)이 맞고(True), 다항식(Polynomial)이 맞음(True)을 나타내며 이를 도 17의 (C)에 도시된 바와 같이, 저장된 룰의 논리적 조건과 비교하면 룰 R1, R2, R3 중에서 R1과 매칭됨을 알 수 있다. 따라서 도 17의 (D)에 도시된 바와 같이, 매칭된 룰로부터 해당 논리적 조건을 만족하는 동작 정보인 "Solve"를 동작 정보로서 추출할 수 있다. 즉, 이러한 경우, P1이 나타내는 특정 의미를 동작 정보로 인식하여, 생성되는 시멘틱 정보는 동작 인덱스로 포함할 수 있는 것이다.
- [204] 제 3 시멘틱 파서부(1120)는 기 저장된 룰의 논리적조건을 만족하는 모든 동작 정보를 추출할 수 있다. 만일, 자연어 토큰과 수학적 토큰 조합이 이루는 논리적 조건이 저장된 룰의 여러 가지 논리적 조건을 만족할 수도 있으며, 이 경우는 하나의 수학 문제가 여러 개의 동작 정보를 포함하고 있는 경우이며, 자연어 토큰과 수학적 토큰 조합이 어떠한 논리적 조건도 만족하지 않는 경우는 해당

복합문장은 룰 생성시 수학기문장(조합 데이터)의 분석에서 누락됐거나 분석과정에 포함되지 않은 항목 또는 잘못된 수학기문장인 경우로 판단할 수 있다. 또한, 제 3 시멘틱 파서부(1120)는 자연어 파싱의 결과 생성된 자연어토큰의 대상이 되는 수식을 수학기토큰 중에서 매칭시킬 수 있다.

[205]

[206] <제 4 실시예>

[207] 이하, 도 18 내지 도 25을 통해서는 자연어와 수학기식이 포함된 복합문장의 시멘틱 정보 추출 방법과 그를 위한 장치인 제 4 실시예에 대해 설명토록 한다.

[208] 도 18은 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도이다.

[209] 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치(100)는 제 4 정보입력부(1810), 제 4 분리부(1820), 제 4 자연어처리부(1830), 제 4 수식 처리부(1840), 제 4 동작추출부(1850), 제 4 오브젝트생성부(1860) 및 제 4 룰저장부(1870)를 포함하여 구성될 수 있다.

[210] 제 4 정보입력부(1810)는 자연어와 수식이 포함된 복합문장을 수신한다. 제 4 분리부(1820)는 복합문장에서 자연어 및 수식을 분리한다. 제 4 자연어처리부(1830)는 분리된 자연어를 토큰화하여 자연어토큰을 생성한다. 제 4 수식 처리부(1840)는 분리된 수식을 파싱하고 시멘틱의미를 추출하여 수학기토큰을 생성한다. 제 4 룰저장부(1870)는 자연어와 수학식의 조합과 대응되는 동작정보를 결합한 룰을 저장한다. 제 4 동작추출부(1850)는 생성된 자연어토큰과 수학기토큰을 저장된 룰 내의 자연어와 수학식의 조합과 비교하여 제 4 룰저장부(1870)에 저장된 룰로부터 복합문장의 동작정보를 추출한다. 제 4 오브젝트생성부(1860)는 자연어토큰의 대상이 되는 수식을 제 4 수식 처리부(1840)에서 생성한 수학기토큰 중에서 매칭시켜 수학 오브젝트를 생성한다.

[211] 수학 오브젝트를 생성함에 있어서 자연어와 더불어 수식이 포함된 복합문장으로 구성된 수학기문장의 실제 의미를 추출하여 표현하기 위하여 다음의 단계를 수행한다.

[212] 1. 수학기식과 자연어의 토큰 관계를 룰로 구성하는 단계

[213] 2. 자연어와 수학식을 표현한 문장을 읽어 들여 수학기문장이 의미하는 동작정보를 찾는 단계.

[214] 3. 수학 오브젝트를 구성하는 단계.

[215] 수학 문장이 가지는 시멘틱 정보는 동작정보와 수학 오브젝트를 포함할 수 있다. 또한, 동작정보(Action)는 수학기문제가 기본적으로 풀어야 하는 목적을 표현한다. 예를 들어, 해당 수학기문장이 문제를 푸는 것인지, 개념을 설명하는 것인지 등에 관해서 실제 푸는 사람이 동작을 취할 수 있도록 하는 정보에 기반해서 문제에서 추출한 정보다. 이 정보는 자연어와 수학식의 토큰을 통해 전처리(pre-processing)되어 정의된 룰에 의해서 생성될 수 있다.

- [216] 수학 오브젝트(Math Object)는 수학 문제에 포함된 세분화된 각 엔티티(entity)를 표현하는 데 사용된다. 즉, 이 수학 문제를 풀기 위해서는 어떤 기법(technique)이나 사실(fact)이 필요한지를 나타낼 수 있고, 수학 문제 내부에 어떤 형태의 함수가 들어가는지 등을 나타낼 수도 있다. 이 오브젝트의 개념은 수학 문제의 다양성을 지원하기 위한 확장성에 도움이 될 수 있다. 이 정보는 자연어에서 얻은 정보와 수학적식에서 얻은 정보가 각각 수학 오브젝트화 될 수 있다.
- [217] 도 19는 수학 문제가 구성된 형태를 트리 구조로 예시한 도면이다. 도 19에 도시된 바와 같이, 하나의 수학 콘텐츠가 가질 수 있는 구조를 트리로 표현하면, 해당 수학콘텐츠(루트(Root) 노드)를 구성하는 자식 노드들은 중요 의미 중 하나인 어순 정보를 그대로 유지한 채 자연어와 수식으로 분리된 형태를 갖게 된다. 또한, 각 자연어는 문장의 연결 순서에 따라 특별한 의미를 가진다. 예를 들어서, 하나의 자연어의 뒤에 따라오는 수식이 특정 조건으로서 연결되거나, 뒤에 오는 수식이 정의되었는지 등의 의미를 가질 수 있다.
- [218] 수학 문제에서 자동으로 위와 같은 정보를 얻기 위해서는, 자연어와 표준화된 수학적식을 각각 구분하여 토큰화(tokenization)를 할 필요가 있다. 이런 자연어와 수학적식을 분석하는 프로그램의 입력은 도 18과 같이 이 두 가지가 혼재된 형태가 될 것이다. 즉, 일반 자연어와 W3C(World Wide Web Consortium)에서 표준화된 MathML 표준을 따르는 XML이 입력이 될 수 있다.
- [219] 제 4 정보입력부(1810)는 자연어 및 수식의 조합으로 이루어진 조합 데이터(복합문장)를 입력받는다. 여기서, 조합 데이터는 수학 문제, 수식 증명 등을 포함한 수학 콘텐츠인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 또한, 자연어 및 수식의 조합으로 이루어진 조합 데이터는 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 자연어 및 수식의 조합으로 이루어진 문서 데이터를 입력받을 수도 있을 것이다.
- [220] 제 4 분리부(1820)는 조합 데이터에서 자연어 및 수식을 분리한다. 즉, 제 4 분리부(1820)는 제 4 정보입력부(1810)를 통해 자연어 및 수식의 조합으로 이루어진 조합 데이터가 입력되면, 조합 데이터에 포함된 자연어와 수식을 분리하여 인식한다. 여기서, 수식은 콘텐츠 기반의 MathML(cMathML) 형태로 생성될 수 있다.
- [221] 제 4 자연어처리부(1830)는 자연어를 토큰화한 자연어 토큰을 생성하고, 생성된 자연어 토큰에서 중지 단어(Stop Word)를 필터링한 중지단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성하고, 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다. 여기서, 토큰(Token)이란 연속된 문장에서 구별할 수 있는 단위를 말하며, 토큰화는 자연어를 복합문장의 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드(Word) 단위로 쪼개는 과정을

말한다. 제 4 자연어처리부(1830)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 제 4 자연어처리부(1830)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 제 4 자연어처리부(1830)는 중복 제거 필터링 데이터에서 술어에 해당하는 데이터를 기 정의된 의미가 부여된 동작 정보와 매칭하여 자연어토큰을 추출할 수 있다.

[222] 토큰화에 대해 좀더 구체적으로 설명하면, 제 4 실시예에서 토큰화는 크게 자연어 토큰화와 수식 토큰화로 구분될 수 있는데, 이 중에서 자연어 토큰화란 조합 데이터(수학 문제 또는 복합문장)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 단어를 자연어 토큰으로 인식하는 과정을 말한다. 한편, 수식 토큰화란 조합 데이터에 포함되는 수식을 파싱(Parsing)한 후 얻게 되는 개별 단위 정보를 수식 토큰으로 인식하는 과정을 말한다.

[223] [예제 1]

[224] Find the function value $9y^3 + 8y^2 - 4y - 9$ with $y = -1$

[225] 예를 들어서, [예제 1]에서 자연어 토큰에 해당하는 정보는 'Find', 'the', 'function', 'value', 'with' 가 되며, 수식 토큰은 파싱을 통해서 정보를 추출한 후에 반환되는 값인 다항식(Polynomial), 최고 차수(Maxdegree=3), 항의 수(Numofterm=4), 컨디션(Condition, $y=-1$) 등이 될 수 있다.

[226] 또한, 중지단어 필터링에 대하여 상세히 설명하면, 중지 단어란 문장이나 수식의 분석에 있어서 필요 없는 토큰에 해당하는 부분을 제거하기 위해서 미리 정의해 놓은 단어들의 집합을 의미한다. 즉, [예제 1]의 단어들 중에서 'the'와 같은 단어(이외에도 a나 to 등)는 중지단어로서 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치(100)에서 사전(Dictionary) 형태로 미리 정의되어 있다. 여기서, 사전은 단어의 집합을 포함하는 리스트를 의미한다. 즉, 제 4 자연어처리부(1830)는 자연어 토큰을 생성한 후 분석에 필요 없는 부분인 중지단어를 제거하는 과정을 수행하게 되는데, 중지 단어 필터링은 수학 문제가 길어질 경우(서술형 문제 등)에 분석 과정에 너무 많은 토큰이 들어가는 것을 방지해 주며, 더불어 수학문제의 처리 속도를 향상시키기 위해 동작한다. 그리고, 중복제거 필터링은, 예를 들어, "이 방정식에서 하나의 해는 3인데, 방정식이 가지고 있는 다른 해를 구하십시오." 라는 수학문제가 있는 경우 자연어를 토큰화하고 나면 "방정식"과 "해"라는 토큰이 각각 두 개씩 추출될 수 있다. 이 경우 중복되는 2개의 "방정식"이라는 토큰과 2개의 "해"라는 토큰 중에서 각각 하나씩 제거하여, 제거된 데이터를 이용하여 동작정보 추출할 수 있다.

[227] 제 4 수식 처리부(1840)는 복합문장으로부터 분리된 수식을 파싱하고

시멘틱의미를 추출하여 수학적토큰을 생성한다. 제 4 수식 처리부(1840)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단(Traverse) 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 제 4 수식 처리부(1840)는 Math ML(Mathematical Markup Language)로 작성된 수식을 XML 트리 형태로 변환한 후 DOM(Document Object Model) 형태로 변환할 수 있다. 제 4 수식 처리부(1840)는 수식을 구성하는 정보의 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색(Depth-First Search) 방식으로 횡단을 실행하여 시멘틱 의미를 추출한다.

- [228] 횡단 과정과 깊이 우선 검색에 대해 구체적으로 설명하면, 일반적으로 수식은 Math ML의 형태를 띄고 있으며, 이는 트리의 형태로 구성이 되며, 이러한 트리로부터 정보를 추출하기 위하여 이러한 트리의 노드를 검색해 나가는 과정을 횡단 과정이라 칭하며, 횡단 과정을 수행할 때, 깊이 우선 검색(Depth-First Search)을 사용할 수 있다. 깊이 우선 검색 횡단 과정은 트리의 루트(Root)에서 시작하여 자식 노드까지 들어간 후 모든 자식 노드의 검색이 끝나면 부모 노드로 이동하기 때문에, 자식 노드에서 가지고 있는 정보 모두를 부모 노드로 전달하며, 시간 복잡도 측면에서 노드 사이를 연결하는 연결선인 엣지(Edge)의 수만큼만 검색을 수행하면 되므로 효율적이다. 여기서 깊이 우선 검색을 예시하였으나 제 4 실시예가 이에 한정되지는 않는다.
- [229] 도 20은 룰을 생성하는 절차를 예시한 도면이다.
- [230] 제 4 룰저장부(1870)는 자연어토큰과 수학적토큰의 조합과 대응되는 동작정보를 결합한 룰을 저장한다.
- [231] 여기서 제 4 룰저장부(1870)에 저장되는 룰은 각각 하나 이상의 자연어토큰과 수학적토큰의 논리적인 조건과 이러한 논리적인 조건에 대응하여 생성될 수 있는 동작 정보를 포함할 수 있다.
- [232] 룰을 저장하기 위해서는 수학 문제를 기반으로 해서 어떠한 자연어 토큰과 수학적시멘틱 토큰 조합이 있는지 파악하는 절차를 수행한다(S2010). 이는 룰의 논리적 조건(예컨대, Binary 트리 형태의 자료구조 상에서 LHS(Left Hand Side)로 저장될 수 있음)이 된다. 논리적 조건은 여러 개의 토큰으로 구성돼 있으면서 토큰들의 논리적인 관계를 정의할 수 있다. 즉, 두 개의 토큰이 동시에 만족해야만 하는 그리고(and) 조건, 두 개 중에 하나만 만족하면 되는 또는(or) 조건 등을 이용하여 다수의 자연어 토큰 및 수학적토큰을 논리적관계로 정의할 수 있다. 다음 단계로, 정의된 자연어 토큰과 수학적 토큰으로 구성된 논리적 조건에 대응하는 동작정보(Action)(예컨대 Binary 트리 형태의 자료구조 상에서 RHS(Right Hand Side)로 저장될 수 있음)를 정의하며(S2020), 따라서 이와 같은 정의에 의해 동작정보를 추출하고자 하는 수학문장이 제 4 룰저장부(1870)에 저장된 어느 룰의 논리적 조건을 만족하면 그에 대응되는 동작정보를 생성하는 형태가 될 수 있다. 이와 같이 정의된 룰을 파일로 생성하고(S2030), 생성된 파일을 룰엔진에 XML 형태로 입력함으로써 제 4 룰저장부(1870)에 저장될 수

있다(S2040).

- [233] 제 4 동작추출부(1850)는 제 4 자연어처리부(1830) 및 제 4 수식 처리부(1840)에서 각각 생성된 자연어토큰과 수학적토큰을 제 4 룰저장부(1870)에 저장된 룰의 자연어와 수학적의 논리적 조건과 비교하여 저장된 어느 룰의 논리적 조건을 만족하면 그에 대응되는 동작정보를 추출하여 해당 복합문장의 동작정보로 생성한다.
- [234] 도 24은 룰 매칭에 의해 동작정보를 추출하는 방법을 예시한 도면이다.
- [235] 도 24에 예시하듯이, (A)와 같이 수학기문장이 4개(P1, P2, P3, P4)가 존재하는 경우, (B)와 같이 제 4 자연어처리부(1830) 및 제 4 수식 처리부(1840)에 의해 파싱된 결과가 생성될 수 있다. 예를 들어, P1의 경우, 제 4 자연어처리부(1830)에 의해 파싱된 결과 수식명(Name)이 "Find"이고 그 타입은 동사(VB)임을 나타내고, 제 4 수식 처리부(1840)에 의해 파싱된 결과가 방정식(Equation)이 맞고(True), 다항식(Polynomial)이 맞음(True)을 나타내며 이를 (C)와 같이 저장된 룰의 논리적 조건과 비교하면 룰 R1, R2, R3 중에서 R1과 매칭됨을 알 수 있다. 따라서 (D)와 같이 매칭된 룰로부터 해당 논리적 조건을 만족하는 동작정보인 "Solve"를 동작정보로서 추출될 수 있다.
- [236] 제 4 동작추출부(1850)는 제 4 룰저장부(1870)에 저장된 룰의 논리적조건을 만족하는 모든 동작정보를 추출할 수 있다. 만일, 자연어 토큰과 수학적 토큰 조합이 이루는 논리적 조건이 저장된 룰의 여러 가지 논리적 조건을 만족할 수도 있으며, 이 경우는 하나의 수학기문제가 여러 개의 동작정보를 포함하고 있는 경우이며, 자연어 토큰과 수학적 토큰 조합이 어떠한 논리적 조건도 만족하지 않는 경우는 해당 복합문장을 룰 생성시 수학기문장의 분석에서 누락되거나 분석과정에 포함되지 않은 항목 또는 잘못된 수학기문장인 경우로 판단할 수 있다.
- [237] 제 4 오브젝트생성부(1860)는 자연어 파싱의 결과 생성된 자연어토큰의 대상이 되는 수식을 수학적토큰 중에서 매칭시킨다.
- [238] 도 21는 제 4 룰저장부(1870)로 사용되는 룰엔진의 구성 및 동작정보의 추출과정을 예시한 도면이다.
- [239] 도 21에서 제 4 자연어처리부(1830)에서 추출된 자연어의 토큰과 제 4 수식 처리부(1840)에서 추출된 수학적의 시멘틱 의미를 갖는 수학적 토큰은 해당 수학기문제가 가진 전체 동작의 의미를 추출하기 위해서 사용된다. 앞서 설명한 것과 같이 수학기문제의 전처리 과정(Pre-processing)을 통해서 어떤 자연어 토큰과 어떤 수학적 시멘틱 토큰이 들어왔을 때, 추출될 동작정보는 미리 XML로 입력되어(S2110) 룰(rule)로 정의되어 저장된다(S2120). 분석하고자 하는 복합문장은 자연어토큰과 수학적 토큰으로 분리 파싱되며(S2130, S2140), 각 토큰들은 사실(Fact)으로써 제 4 동작추출부(1850)에 입력되고(S2150), 제 4 동작추출부(1850)는 룰을 검색할 수 있는 룰 엔진을 구동하여 룰이 정의되어 저장된(예컨대, XML 형태로) 제 4 룰저장부(1870)를 참조한다(S2160). 룰 엔진은 입력된 사실과 저장된 룰을 비교하여 논리적 조건을 만족하는 해당 룰의

- 동작정보를 발생한다(S2170).
- [240] 도 22는 수학 오브젝트를 획득하는 절차에 대해서 간략히 도시한 도면이다.
- [241] 도 22의 왼쪽 부분의 흐름도(S2240, S2250, S2260)는 자연어의 내부에 수학 문제에서 알고 있어야만 하는 기법(Technique), 정의(Definition), 정리(Theorem) 등의 지식(Knowledge)에 해당하는 정보를 추출한다. 이러한 정보는 문제 분석을 통해서 필요한 정보가 더 있는 경우, 원하는 형태의 범주(Category)를 만들고 추가할 수 있다.
- [242] 도 22에서 오른쪽 부분의 흐름도(S2210, S2220, S2230)는 W3C에서 표준화된 MathML의 형태로 들어오는 수학식의 파싱(parsing)을 통해서 시멘틱 정보를 추출하는 과정을 나타낸다. 즉, 제 4 수식 처리부(1840)는 수학적토큰이 입력되면(S2210), 일반 DOM(Document Object Model)으로 XML을 트리 형태로 만든 후에 깊이 우선 검색(Depth First Search)를 통해서 가장 낮은 노드의 정보를 파악하고 상위 노드로 전달하는 식으로 정보를 수집하여 수학적식을 파싱하고(S2220) 시멘틱 정보를 추출한다(S2230). 수학적식의 시멘틱 정보 추출에 관한 기술은 제 4 실시예의 범위를 벗어나기 때문에 자세한 언급을 생략한다.
- [243] 자연어가 입력되면(S2240), 자연어를 파싱하여 자연어토큰을 생성하고(S2250), 생성된 자연어토큰의 대상이 되는 수식을 제 4 수식 처리부(1840)에서 생성된 수학적토큰 중에서 매칭시키는 과정을 수행하여 해당 수학 오브젝트를 추출하여(S2260) 자연어토큰과 결합된 형태로 수학적 오브젝트를 저장한다(S2270).
- [244] 여기서, 수학적 오브젝트는 저장하는 방법에 따라서 다양한 형태로 저장이 가능하며, 이는 병렬, 직렬, 내포(nested) 형태 등으로 표현될 수 있다. 즉, 하나의 수학적 오브젝트 내에 다수의 수학적 오브젝트가 직렬로 배열되거나 병렬로 배열되거나 하나의 수학적 오브젝트 내에 다른 수학적 오브젝트가 포함되는 형태가 될 수 있다.
- [245] 제 4 실시예에는 수학 문제의 동작정보와 수학 오브젝트는 수학 문제가 무엇을 하는 것인지, 어떤 내용을 포함하고 있는지에 대한 정보를 모두 포함하고 있다. 이런 수학 문제 시멘틱 정보를 바탕으로 활용할 수 있는 범위는 아주 광범위하다. 예를 들면, 어떤 사람이 이차 방정식을 푸는 문제를 연습하고 싶다고 하면, 현재의 수학 문제처럼 자연어를 비교하고, MathML 형태의 XML을 모두 파싱하고 원하는 정보가 있는지 확인하는 절차를 거치는 대신에, 미리 추출한 정보를 바탕으로 원하는 정보를 빠른 시간에 제공할 수 있는 것이다. 또한, 검색된 문제 간의 연관성(ranking)을 파악하는 과정에서도 사용될 수 있고, 이런 동작은 사용자로 하여금 최적의 검색 결과를 얻을 수 있도록 도와줄 수 있다.
- [246] 도 23은 제 4 실시예에 따른 복합문장의 시멘틱 정보 추출방법을 설명하기 위한 순서도이다.
- [247] 제 4 실시예에 따른 복합문장의 시멘틱 정보 추출방법은 자연어)와 수식이

포함된 복합문장을 수신하는 정보입력단계(S2310), 복합문장에서 자연어 및 수식을 분리하는 분리단계(S2320), 분리된 자연어를 토큰화하여 자연어토큰을 생성하는 자연어처리단계(S2330), 분리된 수식을 파싱하고 시멘틱의미를 추출하여 수학적토큰을 생성하는 수식처리단계(S2340), 생성된 자연어토큰과 수학적토큰을 자연어와 수학적의 논리적 조건 및 상기 논리적 조건에 대응되는 동작정보를 결합한 룰을 비교하여 복합문장의 동작정보를 추출하는 동작추출단계(S2350) 및 생성된 자연어토큰의 대상이 되는 수식을 생성된 수학적토큰 중에서 매칭시키는 오브젝트생성단계(S2360)를 포함한다.

[248] 여기서, 정보입력단계(S2310)는 제 4 정보입력부(1810)의 동작에 대응되며, 분리단계(S2320)는 제 4 분리부(1820)의 동작에 대응되며, 자연어처리단계(S2330)는 제 4 자연어처리부(1830)의 동작에 대응되며, 수식처리단계(S2340)는 제 4 수식 처리부(1840)의 동작에 대응되며, 동작추출단계(S2350)은 제 4 동작추출부(1850)의 동작에, 오브젝트생성단계(S2360)는 제 4 오브젝트생성부(1860)의 동작에 각각 대응되므로 상세한 설명은 생략한다.

[249] 전술한 바와 같이 도 23에 기재된 제 4 실시예에 따른 복합문장의 시멘틱 정보 추출방법은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 제 4 실시예에 따른 복합문장의 시멘틱 정보 추출방법을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 이러한 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 제 4 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드 및 코드 세그먼트들은 제 4 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.

[250] 도 25은 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도이다.

[251] 제 4 실시예에 따른 복합문장의 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하기 위해서는 단말기(910), 통신망(920) 및 제 4 클라우드 컴퓨팅 장치(2500)를 포함한 시스템이 필요하다.

[252] 여기서, 단말기(910)는 사용자의 명령 또는 조작에 따라 통신망(920)을 경유하여 각종 데이터를 송수신할 수 있는 단말기를 말하는 것이며, 태블릿 PC, 랩톱, 개인용 컴퓨터, 스마트폰, 개인휴대용 정보단말기 및 무선 통신 단말기 등 중 어느 하나일 수 있다. 또한, 단말기(910)는 통신망(920)을 통하여 데이터 읽거나 입력 및 저장, 네트워크, 콘텐츠 사용 등의 서비스를 이용할 수 있는

클라우드 컴퓨팅을 지원하는 클라우드 컴퓨팅 단말기가 될 수 있다. 즉, 단말기(910)는 통신망(920)을 경유하여 제 4 클라우드 컴퓨팅 장치(2500)에 접속하기 위한 프로그램을 저장하기 위한 메모리, 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비하고 있는 장치를 의미한다. 즉, 단말기(910)로는 통신망(920)에 연결되어 제 4 클라우드 컴퓨팅 장치(2500)와 클라이언트-서버 통신이 가능하다면 그 어떠한 단말기도 가능하며, 노트북 컴퓨터, 이동통신 단말기, PDA 등 여하한 통신 컴퓨팅 장치를 모두 포함하는 넓은 개념이다. 한편, 단말기(910)는 터치 스크린을 구비한 형태로 제작되는 것이 바람직하나 반드시 이에 한정되는 것은 아니다.

[253] 단말기(910)는 제 4 클라우드 컴퓨팅 장치(2500)로 복합문장을 입력하고 제 4 클라우드 컴퓨팅 장치(2500)는 통해 클라우드 컴퓨팅 방식으로 복합문장의 복합문장의 시멘틱 정보를 추출하고 단말기(910)으로 제공할 수 있다. 즉, 단말기(910)는 제 4 클라우드 컴퓨팅 장치(2500)로부터 클라우드 컴퓨팅 방식으로 데이터의 입출력을 위하여 제 4 클라우드 컴퓨팅 장치(2500)와의 입출력 인터페이스를 제공하는 별도의 입출력 인터페이스부를 포함할 수 있으며, 입출력 인터페이스부를 통해 제 4 클라우드 컴퓨팅 장치(2500)에 저장된 저장 매체에 대한 데이터 읽기 및 쓰기가 수행되도록 하는 인터페이스 제어부를 포함할 수 있다. 이에 대해 좀 더 구체적으로 설명하면, 단말기(910)는 입출력 인터페이스부를 통해 자연어 및 수식의 조합으로 이루어진 복합문장을 제 4 클라우드 컴퓨팅 장치(2500)로 입력할 수 있고, 제 4 클라우드 컴퓨팅 장치(2500)는 자연어와 수식이 포함된 복합문장을 수신하고 복합문장에서 자연어 및 상기 수식을 분리한 후, 분리된 자연어를 토큰화하여 자연어토큰을 생성하고 분리된 수식을 파싱하고 시멘틱의미를 추출하여 수학적토큰을 생성하고 자연어와 수학적식의 논리적 조건 및 논리적 조건에 대응되는 동작정보를 결합한 룰을 이용하여 생성된 자연어토큰과 수학적토큰을 저장된 룰의 논리적 조건과 비교하여 룰로부터 복합문장의 동작정보를 추출하므로, 실질적으로 단말기(910)에서는 어떠한 어플리케이션의 탑재 없이도 복합문장의 시멘틱 정보를 추출할 수 있다.

[254] 통신망(920)은 인터넷망, 인트라넷망, 이동통신망, 위성 통신망 등 다양한 유무선 통신 기술을 이용하여 인터넷 프로토콜로 데이터를 송수신할 수 있는 망을 말하며, 단말기(910)와 제 4 클라우드 컴퓨팅 장치(2500) 간에 데이터를 중계하는 기능을 수행한다.

[255] 제 4 클라우드 컴퓨팅 장치(2500)는 자연어 및 수식 처리 장치(100)를 기반으로 구현될 수 있으며, 클라우드 컴퓨팅 방법으로 단말기(910)에게 복합문장의 시멘틱 정보를 추출할 수 있도록, 단말기(910)로 하여금 제 4 클라우드 컴퓨팅 장치(2500)에 저장된 저장 매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 하되, 자연어 및 수식의 조합으로 이루어진 복합문장이 입력되면, 복합문장에서 자연어 및 수식을 각각 분리하며, 분리된 자연어 및 수식을 구성하고 있는

각각의 정보를 분석하여 시멘틱 의미를 추출하고, 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작정보를 추출하여 기록매체에 저장하며, 해당 기록매체의 데이터를 단말기(910)로 전송함으로써, 단말기(910)에서 어플리케이션의 탑재 없이 복합문장의 논리적 표현을 변환할 수 있도록 하는 클라우드 컴퓨팅을 제공할 수 있다. 즉, 제 4 클라우드 컴퓨팅 장치(2500)는 클라우드 컴퓨팅 방식으로 복합문장의 시멘틱 정보를 추출한 결과를 저장하는 제 4 시멘틱 정보 추출부(2510)와 단말기(910)로 하여금 제 4 시멘틱 정보 추출부(2510)에 의해 저장 매체에 저장된 데이터의 읽기 및 쓰기가 수행되도록 하는 제 4 클라우드 컴퓨팅부(2520)를 구비할 수 있다.

[256]

[257] <제 5 실시예>

[258] 이하, 도 26 내지 도 32을 통해서는 자연어와 수학적식이 포함된 복합문장의 논리적 표현 변환방법과 그를 위한 장치인 제 5 실시예에 대해 설명토록 한다.

[259] 도 26은 제 5 실시예에 따른 복합문장의 자연어 및 수식 처리 장치를 개략적으로 나타낸 블록 구성도이다.

[260] 제 5 실시예에 따른 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 정보 입력부(2610), 제 5 문장분석부(2620), 제 5 동작추출부(2630) 및 제 5 동작실행부(2640)를 포함하여 구성될 수 있다. 제 5 정보 입력부(2610)는 자연어와 수식이 포함된 복합문장을 수신한다. 제 5 문장분석부(2620)는 복합문장의 문장구성을 분석하여 수식데이터 및 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성한다. 제 5 동작추출부(2630)는 자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는 동작 정보를 추출한다. 제 5 동작실행부(2640)는 추출된 동작 정보를 수식토큰에 대하여 구조화한다. 여기서, 구조화의 의미는 추출된 동작 정보를 수식토큰과 결합하여 구조화한다는 의미이다.

[261] 도 27은 제 5 실시예에 따른 제 5 문장분석부(2620)를 개략적으로 나타낸 블록 구성도이다.

[262] 제 5 문장분석부(2620)는 조합 데이터에서 자연어 및 수식을 각각 분리하는 제 5 분리부(2710), 분리된 자연어를 구성하고 있는 각각의 자연어정보를 분석하여 시멘틱의미를 추출하는 제 5 자연어 처리부(2720) 및 분리된 수식을 구성하고 있는 각각의 수식정보를 분석하여 시멘틱 의미를 추출하는 제 5 수식 처리부(2730)를 포함한다.

[263] 제 5 정보 입력부(2610)는 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는다. 여기서, 조합 데이터는 수학 문제, 수식 증명 등을 포함한 수학 컨텐츠인 것이 바람직하나 반드시 이에 한정되는 것은 아니다. 또한, 자연어 및 수식의 조합으로 이루어진 조합 데이터는 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 자연어 및 수식의 조합으로 이루어진 문서 데이터를 입력받을 수도 있을 것이다.

제 5 분리부(2710)는 조합 데이터에서 자연어 및 수식을 각각 분리한다. 즉, 제 5 분리부(2710)는 제 5 정보 입력부(2610)를 통해 자연어 및 수식의 조합으로 이루어진 조합 데이터가 입력되면, 조합 데이터에 포함된 자연어와 수식을 각각 분리하여 인식한다.

- [264] 제 5 자연어 처리부(2720)는 분리된 자연어를 구성하고 있는 자연어정보를 분석하여 시멘틱 의미를 추출한다. 제 5 자연어 처리부(2720)는 자연어를 토큰화(Tokenization)한 자연어 토큰을 생성하고, 자연어 토큰을 근거로 기설정된 중지 단어(Stop Word)를 필터링한 중지 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성한다. 여기서, 토큰(Token)이란 연속된 문장에서 구별할 수 있는 단위를 말하며, 토큰화는 자연어를 복합문장의 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드(Word) 단위로 쪼개는 과정을 말한다. 토큰화에 대해 좀더 구체적으로 설명하면, 제 5 실시예에서 토큰화는 크게 자연어 토큰화와 수식 토큰화로 구분된다. 자연어 토큰화란 조합 데이터(수학 문제 또는 복합문장)에 포함된 자연어를 공백(Space)을 기준으로 분리한 결과물에 해당하는 각각의 단어를 자연어 토큰으로 인식하는 과정을 말한다. 한편, 수식 토큰화란 조합 데이터에 포함되는 수식을 파싱(Parsing)한 후 얻게 되는 개별 단위 정보를 수식 토큰으로 인식하는 과정을 말한다.

[265] [예제 1]

[266] Find the function value $9y^3 + 8y^2 - 4y - 9$ with $y = -1$

[267] 예를 들어서, [예제 1]에서 자연어 토큰에 해당하는 정보는 'Find', 'the', 'function', 'value', 'with' 가 되며, 수식 토큰은 파싱을 통해서 정보를 추출한 후에 반환되는 값인 다항식(Polynomial), 최고 차수(Maxdegree=3), 항의 수(Numofterm=4), 컨디션(Condition, y=-1) 등이 될 수 있다.

[268] 또한, 중지단어 필터링에 대하여 상세히 설명하면, 중지 단어란 문장이나 수식의 분석에 있어서 필요 없는 토큰에 해당하는 부분을 제거하기 위해서 미리 정의해 놓은 단어들의 집합을 의미하며, 제 5 자연어 처리부(2720)는 자연어토큰 중에서 불필요한 토큰을 정의한 중지단어 리스트를 참조하여 수행될 수 있다. 즉, [예제 1]의 단어들 중에서 'the'와 같은 단어(이외에도 a나 to 등)는 중지단어로서 시스템에서 사전(Dictionary) 형태로 미리 정의되어 있다. 여기서, 사전은 단어의 집합을 포함하는 리스트를 의미한다. 즉, 제 5 자연어 처리부(2720)는 자연어 토큰을 생성한 후 분석에 필요 없는 부분인 중지단어를 제거하는 과정을 수행하게 되는데, 중지 단어 필터링은 수학 문제가 길어질 경우(서술형 문제 등)에 분석 과정에 너무 많은 토큰이 들어가는 것을 방지해 주며, 더불어 시스템의 처리 속도를 향상시키기 위해 동작한다. 제 5 자연어 처리부(2720)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다.

- [269] 제 5 동작추출부(2630)는 제 5 자연어 처리부(2720)에서 중복 제거 필터링 데이터에 기 정의된 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작 정보를 추출한다. 동작 정보(action)는 입력된 복합문장이 문제를 푸는 것인지, 개념을 설명하는 것인지 등에 관해서 실제 푸는 사람이 동작을 취할 수 있도록 하는 정보에 기반해서 문제에서 추출한 정보이다. 즉, 동작 정보는 수학 문제에 포함되는 토큰을 바탕으로 추출할 수 있는 요약 정보를 의미한다. 예를 들어 위의 [예제 1]의 수학콘텐츠에서 자연어 토큰과 수학 토큰을 바탕으로 solve라는 동작 정보를 추출할 수 있다. 이렇게 함으로써 수학기문제를 스키마로 정의하는 과정에서 전체 문제가 의미하는 대표 동작에 대한 정보를 가질 수 있다. 이는 이후에 검색이나 문제간의 연관성(similarity)을 분석할 때 도움이 되는 도구가 될 수 있다.
- [270] 제 5 수식 처리부(2730)는 분리된 수식을 구성하고 있는 각각의 수식정보를 분석하여 시멘틱 의미를 추출한다. 제 5 수식 처리부(2730)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단(Traverse) 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 제 5 수식 처리부(2730)는 Math ML(Mathematical Markup Language)로 작성된 수식을 XML 트리 형태로 변환한 후 DOM(Document Object Model) 형태로 변환한다. 제 5 수식 처리부(2730)는 수식을 구성하는 수식정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색(Depth-First Search) 방식으로 횡단을 실행한다. 한편, 횡단 과정과 깊이 우선 검색에 대해 구체적으로 설명하면, 일반적으로 수식은 Math ML의 형태를 띄고 있으며, 이는 트리의 형태로 구성이 되며, 이러한 트리로부터 정보를 추출하기 위하여 이러한 트리의 노드를 검색해 나가는 과정을 횡단 과정이라 칭하며, 횡단 과정을 수행할 때, 깊이 우선 검색(Depth-First Search)을 사용할 수 있다. 깊이 우선 검색 횡단 과정은 트리의 루트(Root)에서 시작하여 자식 노드까지 들어간 후 모든 자식 노드의 검색이 끝나면 부모 노드로 이동하기 때문에, 자식 노드에서 가지고 있는 정보 모두를 부모 노드로 전달하며, 시간 복잡도 측면에서 노드 사이를 연결하는 연결선인 엣지(Edge)의 수만큼만 검색을 수행하면 되므로 효율적이다.
- [271] 도 28은 제 5 실시예에 따른 자연어 처리부를 개략적으로 나타낸 블럭 구성도이다.
- [272] 제 5 실시예에 따른 제 5 자연어 처리부(2720)는 제 5 자연어 토큰화부(2810), 제 5 중지 단어 필터링부(2820) 및 제 5 중복 제거 필터링부(2830)를 포함한다. 한편, 제 5 실시예에서는 제 5 자연어 처리부(2720)가 제 5 자연어 토큰화부(2810), 제 5 중지 단어 필터링부(2820), 제 5 중복 제거 필터링부(2830)만을 포함하는 것으로 기재하고 있으나, 이는 제 5 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 5 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 5 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 제 5 자연어 처리부(2720)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용

가능할 것이다.

- [273] 제 5 자연어 토큰화부(2810)는 자연어를 토큰화한 자연어 토큰을 생성한다. 제 5 자연어 토큰화부(2810)는 자연어를 구성하고 있는 자연어정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 예를 들어서, 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 자연어 토큰화부(2810)를 이용하여 조합 데이터에 포함된 자연어 노드를 개별적으로 입력받거나 자연어 노드 전체를 한번에 입력받을 수 있다. 여기서, 자연어는 노드 자체가 복수 개의 단어로 구성된 문장의 성질을 가지거나 완벽한 문장으로 한정되는 것은 아니다. 즉, 이런 자연어 노드를 복합문장의 자연어 및 수식 처리 장치(100)가 이해할 수 있는 단위인 워드 단위로 쪼개게 되는데, 이러한 과정을 토큰화 과정이라고 한다.
- [274] 제 5 중지 단어 필터링부(2820)는 자연어 토큰을 근거로 중지 단어를 필터링한 중지 단어 필터링 데이터를 생성한다. 제 5 중지 단어 필터링부(2820)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 즉, 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 중지 단어 필터링부(2820)를 이용하여 토큰화 과정이 수행된 후 자연어를 구성하고 있는 자연어정보가 복수 개의 토큰으로 분리되어 복합문장의 자연어 및 수식 처리 장치(100)에 입력되면, 다음 단계로 중지 단어 제거 과정을 거친다. 이 과정에서는 시멘틱 의미를 추출하기 위해 필요 없는 토큰들을 제거하게 된다. 예를 들어서, '이', '저', '여기' 및 '저기' 등이 중지 단어로 설정될 수 있으나 반드시 이에 한정되는 것은 아니며, 의미상 필요 없는 토큰을 설정하는 것은 각 시스템에 따라 다르게 설정될 수 있다.
- [275] 제 5 중복 제거 필터링부(2830)는 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성한다. 제 5 중복 제거 필터링부(2830)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다. 즉, 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 중복 제거 필터링부(2830)를 이용하여 중지 단어를 필터링한 후 중복을 제거하는 과정을 수행하며, 중복 제거 필터링을 통해 중복된 단어를 제거함으로써 복합문장의 자연어 및 수식 처리 장치(100)의 처리 부하를 낮출 수 있다.
- [276] 제 5 동작추출부(2630)는 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작 정보를 추출한다. 여기서 자연어토큰 규칙은 자연어토큰의 동작 정보를 정의한 규칙을 의미하며, 자연어에서의 다양한 표현을 어떤 시멘틱 의미(자연어토큰의 의미)로 정의하고, 자연어토큰의 방향성 및 자연어토큰의 영향이 미치는 지점을 포함할 수 있다. 여기서 방향성이란 수학적컨텐츠 내의 자연어토큰이 해당 자연어토큰의 앞의 수학적식과 연관되는지, 아니면 뒤의 수학적식과 연관되는지 또는 독립적인지를 나타내는 것을 의미한다.
- [277] 도 29는 제 5 실시예에 따른 수식 처리부를 개략적으로 나타낸 블럭 구성도이다.

- [278] 제 5 실시예에 따른 제 5 수식 처리부(2730)는 제 5 트리 변환부(2910), 제 5 시멘틱 파싱부(2920) 및 제 5 수식 토큰화부(2930)를 포함한다. 한편, 제 5 실시예에서는 제 5 수식 처리부(2730)가 제 5 트리 변환부(2910), 제 5 시멘틱 파싱부(2920) 및 제 5 수식 토큰화부(2930)만을 포함하는 것으로 기재하고 있으나, 이는 제 5 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 5 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 5 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 제 5 수식 처리부(2730)에 포함되는 구성 요소에 대하여 다양하게 수정 및 변형하여 적용 가능할 것이다. 여기서, 시멘틱이란 해당 장치에서 특정 정보의 뜻을 이해하고 논리적 추론 가능하도록 하는 정보를 의미한다.
- [279] 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 정보 입력부(2610)를 통해 표준화된 형식으로 작성된 개별 수식을 입력받고, 이를 제 5 수식 처리부(2730)로 전달하게 된다. 즉, 제 5 수식 처리부(2730)로 전달된 수식은 W3C(World Wide Web Consortium)에서 정의된 표준인 Math ML(Mathematical Markup Language)을 기준으로 XML 태그의 형태를 이룬다. 단, 제 5 수식 처리부(2730)로 전달된 수식은 Math ML 인 것이 바람직하나 반드시 이에 한정되는 것은 아니다.
- [280] 제 5 트리 변환부(2910)는 수식을 트리 형태로 변환한다. 제 5 트리 변환부(2910)는 Math ML로 작성된 수식을 XML 트리 형태로 변환한 후 DOM 형태로 변환한다. 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 트리 변환부(2910)를 이용하여 수식을 Math ML 형태의 XML 트리로 변환하고, 이 트리는 DOM(Document Object Model)로 변경되어 프로그램에서 접근 가능한 트리 형태로 변환된다.
- [281] 제 5 시멘틱 파싱부(2920)는 트리 형태로 변환된 수식에 횡단 과정을 수행한다. 제 5 시멘틱 파싱부(2920)는 수식을 구성하는 수식정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색 방식으로 횡단을 실행한다. 복합문장의 자연어 및 수식 처리 장치(100)는 제 5 시멘틱 파싱부(2920)를 이용하여 수식이 가진 시멘틱 의미를 파악하기 위해서 횡단 과정을 거치는데, 제 5 시멘틱 파싱부(2920)는 가장 낮은 노드에서 점차 상위 노드로 정보를 전달하는 형태인 깊이 우선 검색으로 횡단을 실행한다. 이에 따라 결과적으로, 제 5 시멘틱 파싱부(2920)를 통해 수집된 수식을 구성하는 정보는 모두 최상위 노드에 집결되고, 이러한 정보를 바탕으로 수식의 토큰을 만드는 과정을 거치게 된다.
- [282] 제 5 수식 토큰화부(2930)는 횡단 과정이 수행된 수식에 토큰화를 수행한다. 즉, 토큰화된 수식 토큰은 수학 언어(Mathematics Natural Language)로 이루어진 토큰을 말한다. 한편, 수식 토큰은 자연어 토큰과는 다르게 취급된다. 즉, 제 5 자연어 처리부(2720)에서는 자연어 토큰을 바탕으로 동작 정보를 매칭하는 반면, 제 5 수식 처리부(2730)에서는 수식 토큰이 결과물이 되며, 향후 수식 토큰은 검색을 통해서 수학 콘텐츠를 찾는 등의 작업에 이용될 수 있다.
- [283] 제 5 동작실행부(2640)는 제 5 동작추출부(2630)에서 추출된 동작 정보를

수식토큰에 대하여 결합하여 구조화하여 소정의 정의된 스키마 형태(예컨대 XML로 구조화)로 출력하거나 또는 저장매체에 저장할 수 있다.

- [284] 도 30는 제 5 실시예에 따른 복합문장의 논리적 표현 변환방법을 설명하기 위한 순서도이다.
- [285] 복합문장의 자연어 및 수식 처리 장치(100)는 자연어 및 수식의 조합으로 이루어진 복합문장을 입력받는다(S3010). 여기서, 자연어 및 수식의 조합으로 이루어진 복합문장은 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 자연어 및 수식의 조합으로 이루어진 복합문장을 입력받을 수도 있을 것이다. 복합문장의 자연어 및 수식 처리 장치(100)는 복합문장에서 자연어 및 수식을 각각 분리한다(S3020). 즉, 복합문장의 자연어 및 수식 처리 장치(100)는 자연어 및 수식의 조합으로 이루어진 복합문장이 입력되면, 복합문장에 포함된 자연어와 수식을 각각 분리하여 인식한다.
- [286] 복합문장의 자연어 및 수식 처리 장치(100)는 분리된 자연어를 구성하고 있는 자연어정보를 분석하여 시멘틱 의미를 추출하는 프로세스를 처리한다(S3030). 즉, 복합문장의 자연어 및 수식 처리 장치(100)는 자연어를 토큰화한 자연어 토큰을 생성하고, 자연어 토큰을 근거로 중지 단어를 필터링한 단어 필터링 데이터를 생성하며, 중지 단어 필터링 데이터에서 중복 제거 필터링을 수행한 중복 제거 필터링 데이터를 생성하고, 중복 제거 필터링 데이터에 기 정의된 의미가 부여된 동작 정보를 매칭한다. 복합문장의 자연어 및 수식 처리 장치(100)는 자연어를 구성하고 있는 자연어정보에 대해 토큰화를 수행하여 자연어 토큰을 생성한다. 복합문장의 자연어 및 수식 처리 장치(100)는 자연어 토큰에서 기 설정된 중지 단어로 판별된 자연어 토큰을 선별하여 제거하는 중지 단어 필터링을 수행하여 중지 단어 필터링 데이터를 생성한다. 복합문장의 자연어 및 수식 처리 장치(100)는 중지 단어 필터링 데이터에서 중복되는 데이터를 선별하여 제거하는 중복 제거 필터링을 수행하여 중복 제거 필터링 데이터를 생성한다.
- [287] 복합문장의 자연어 및 수식 처리 장치(100)는 분리된 수식을 구성하고 있는 각각의 수식정보를 분석하여 시멘틱 의미를 추출하는 프로세스를 처리한다(S3040). 복합문장의 자연어 및 수식 처리 장치(100)는 수식을 트리 형태로 변환하고, 트리 형태로 변환된 수식에 횡단 과정을 수행하고, 횡단 과정이 수행된 수식에 토큰화를 수행한다. 복합문장의 자연어 및 수식 처리 장치(100)는 Math ML로 작성된 수식을 XML 트리 형태로 변환한 후 DOM 형태로 변환한다. 복합문장의 자연어 및 수식 처리 장치(100)는 수식을 구성하는 수식정보를 최하단 노드에서 점차 상위 노드로 전달되도록 하는 깊이 우선 검색 방식으로 횡단을 실행한다.
- [288] 복합문장의 자연어 및 수식 처리 장치(100)는 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작 정보를 추출하고(S3050), 추출된 동작

정보를 수식토큰에 대하여 구조화하여 소정의 정의된 스키마 형태로 출력하거나 또는 저장매체에 저장할 수 있다(S3060).

- [289] 도 30에서는 단계 S3010 내지 단계 S3060을 순차적으로 실행하는 것으로 기재하고 있으나, 이는 제 5 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 제 5 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 제 5 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 도 30에 기재된 순서를 변경하여 실행하거나 단계 S3010 내지 단계 S3060 중 하나 이상의 단계를 병렬적으로 실행하는 것으로 다양하게 수정 및 변형하여 적용 가능할 것이므로, 도 30는 시계열적인 순서로 한정되는 것은 아니다.
- [290] 전술한 바와 같이 도 30에 기재된 제 5 실시예에 따른 복합문장의 논리적 표현 변환방법은 프로그램으로 구현되고 컴퓨터로 읽을 수 있는 기록매체에 기록될 수 있다. 제 5 실시예에 따른 복합문장의 논리적 표현 변환방법을 구현하기 위한 프로그램이 기록되고 컴퓨터가 읽을 수 있는 기록매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 이러한 컴퓨터가 읽을 수 있는 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어, 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터가 읽을 수 있는 기록매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수도 있다. 또한, 제 5 실시예를 구현하기 위한 기능적인(Functional) 프로그램, 코드 및 코드 세그먼트들은 제 5 실시예가 속하는 기술분야의 프로그래머들에 의해 용이하게 추론될 수 있을 것이다.
- [291] 도 31은 제 5 실시예에 따른 복합문장의 트리 형태 표현을 나타낸 예시도이다.
- [292] 도 31에 도시된 바와 같이, 하나의 수학 컨텐츠가 가질 수 있는 구조를 표현하면, 루트(Root) 노드에 연결된 자식 노드들은 중요 의미 중 하나인 어순 정보를 그대로 유지한 채 자연어와 수식으로 분리된 형태를 갖게 된다. 또한, 각 자연어는 문장의 연결 순서에 따라 특별한 의미를 가진다. 즉, 많은 수학 컨텐츠는 자연어를 기준으로 수식을 엮는 구조가 대부분이다. 예를 들어서, 하나의 자연어의 뒤에 따라오는 수식이 특정 조건으로서 연결되거나, 뒤에 오는 수식이 정의되었는지 등의 구조를 가질 수 있는 것이다. 각 노드의 자연어가 가지는 의미와 연결 관계뿐만 아니라, 자연어를 통합하여 시멘틱 의미를 추출할 수 있다. 즉, 수학 컨텐츠에서 요구하는 것이 해당 수식을 푸는 것인지, 설명하는 것인지 등의 동작을 구분하기 위해서는 전체 자연어를 한데 묶어서 의미를 파악하는 것으로 문제의 방향성을 파악하는데 이용될 수 있다.
- [293] 도 32은 제 5 실시예에 따른 복합문장의 논리적 표현 변환 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도이다.
- [294] 제 5 실시예에 따른 복합문장의 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하기 위해서는 단말기(910), 통신망(920) 및

복합문장의 제 5 클라우드 컴퓨팅 장치(3200)를 포함한 시스템이 필요하다.

- [295] 여기서, 단말기(910)는 사용자의 명령 또는 조작에 따라 통신망(920)을 경유하여 각종 데이터를 송수신할 수 있는 단말기를 말하는 것이며, 태블릿 PC, 랩톱, 개인용 컴퓨터, 스마트폰, 개인휴대용 정보단말기 및 무선 통신 단말기 등 중 어느 하나일 수 있다. 또한, 단말기(910)는 통신망(920)을 통하여 데이터 읽거나 입력 및 저장, 네트워크, 콘텐츠 사용 등의 서비스를 이용할 수 있는 클라우드 컴퓨팅을 지원하는 클라우드 컴퓨팅 단말기가 될 수 있다. 즉, 단말기(910)는 통신망(920)을 경유하여 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)에 접속하기 위한 프로그램을 저장하기 위한 메모리, 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비하고 있는 장치를 의미한다. 즉, 단말기(910)로는 통신망(920)에 연결되어 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)와 클라이언트-서버 통신이 가능하다면 그 어떠한 단말기도 가능하며, 노트북 컴퓨터, 이동통신 단말기, PDA 등 여하한 통신 컴퓨팅 장치를 모두 포함하는 넓은 개념이다. 한편, 단말기(910)는 터치 스크린을 구비한 형태로 제작되는 것이 바람직하나 반드시 이에 한정되는 것은 아니다.
- [296] 단말기(910)는 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)로 복합문장을 입력하고 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)는 통해 클라우드 컴퓨팅 방식으로 복합문장의 논리적 관계를 갖도록 그 표현을 변환하고 단말기(910)으로 제공할 수 있다. 즉, 단말기(910)는 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)로부터 클라우드 컴퓨팅 방식으로 데이터의 입출력을 위하여 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)와의 입출력 인터페이스를 제공하는 별도의 입출력 인터페이스부를 포함할 수 있으며, 입출력 인터페이스부를 통해 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)에 저장된 저장 매체에 대한 데이터 읽기 및 쓰기가 수행되도록 하는 인터페이스 제어부를 포함할 수 있다. 이에 대해 좀 더 구체적으로 설명하면, 단말기(910)는 입출력 인터페이스부를 통해 자연어 및 수식의 조합으로 이루어진 복합문장을 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)로 입력할 수 있고, 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)는 복합문장에서 자연어 및 수식을 각각 분리하고, 분리된 자연어를 구성하고 있는 각각의 자연어정보 및 수식정보를 분석하여 시멘틱 의미를 추출하고, 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작 정보를 추출하여 추출된 동작 정보를 수식토큰에 대하여 구조화하므로, 실질적으로 단말기(910)에서는 어떠한 어플리케이션의 탑재 없이도 복합문장의 논리적 표현을 변환할 수 있다.
- [297] 통신망(920)은 인터넷망, 인트라넷망, 이동통신망, 위성 통신망 등 다양한 유무선 통신 기술을 이용하여 인터넷 프로토콜로 데이터를 송수신할 수 있는 망을 말하며, 단말기(910)와 제 5 클라우드 컴퓨팅 장치(3200) 간에 데이터를 중계하는 기능을 수행한다.
- [298] 제 5 클라우드 컴퓨팅 장치(3200)는 자연어 및 수식 처리 장치(100)를 기반으로

구현될 수 있으며, 클라우드 컴퓨팅 방법으로 단말기(910)에게 복합문장의 논리적 표현 변환 결과를 제공할 수 있도록, 단말기(910)로 하여금 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)에 저장된 저장 매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 하되, 자연어 및 수식의 조합으로 이루어진 복합문장이 입력되면, 복합문장에서 자연어 및 수식을 각각 분리하며, 분리된 자연어 및 수식을 구성하고 있는 각각의 정보를 분석하여 시멘틱 의미를 추출하고, 자연어토큰 규칙을 참조하여 자연어토큰의 의미에 대응되는 동작 정보를 추출하여 추출된 동작 정보를 수식토큰에 대하여 구조화하여 컴퓨터로 읽을 수 있는 기록매체에 저장하며, 해당 기록매체의 데이터를 단말기(910)로 전송함으로써, 단말기(910)에서 어플리케이션의 탑재 없이 복합문장의 논리적 표현을 변환할 수 있도록 하는 클라우드 컴퓨팅을 제공할 수 있다. 즉, 복합문장의 제 5 클라우드 컴퓨팅 장치(3200)는 클라우드 컴퓨팅 방식으로 복합문장의 논리적 표현을 변환한 결과를 저장하는 제 5 논리적 표현 변환부(3210)와 단말기(910)로 하여금 제 5 논리적 표현 변환부(3210)에 의해 저장 매체에 저장된 데이터의 읽기 및 쓰기가 수행되도록 하는 제 5 클라우드 컴퓨팅부(3220)를 구비할 수 있다.

[299]

[300] <제 6 실시예>

[301] 이하, 도 32 내지 도 40을 통해서는 수학적 시멘틱 정보 생성 방법과 그를 위한 장치인 제 6 실시예에 대해 설명토록 한다.

[302] 도 33은 제 6 실시예에 따른 수학적 및 자연어 처리 장치를 간략하게 나타낸 블록 구성도이다.

[303] 제 6 실시예에 따른 자연어 및 수식 처리 장치(100)는 제 6 정보 입력부(3310), 제 6 수식데이터 구조화부(3320), 제 6 오퍼레이터 파싱부(3330) 및 제 6 시멘틱 정보 결합부(3340)를 포함하여 구성될 수 있다. 경우에 따라 제 6 시멘틱 정보 결합부(3340)는 생략될 수도 있다.

[304] 제 6 정보 입력부(3310)는 수식(또는 수학적, Mathematical Formula)이 표현된 수식데이터를 입력받아 제 6 수식데이터 구조화부(3320)로 전달한다.

[305] 제 6 수식데이터 구조화부(3320)는 제 6 정보 입력부(3310)로부터 전달받은 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화한다.

[306] 제 6 오퍼레이터 파싱부(3330)는 제 6 수식데이터 구조화부(3320)에서 구조화된 오퍼레이터에 대하여 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성한다.

[307] 제 6 시멘틱 정보 결합부(3340)는 제 6 오퍼레이터 파싱부(3330)에서 생성된 파싱 시멘틱 정보 및 입력된 수식데이터를 결합하여 시멘틱 정보 결합 수식데이터를 생성한다.

[308] W³C(World Wide Web Consortium)에서 그 스키마(Schema)가 정의되어

표준화된 콘텐츠 기반의 MathML(content MathML, 이하 cMathML이라 칭함)은 기존에 표현 기반의 MathML(presentation MathML, 이하 pMathML이라 칭함)이 가지고 있던 한계인 의미적인 부분을 추가할 수 있다. cMathML 자체는 pMathML에서 가지고 있던 의미적으로 명확하지 않은 부분을 해결하기 위하여 더 많은 태그가 존재하며, pMathML과 동일하게 그 내부에 포함된 수학식의 의미를 파악하기 위해서 그 때마다 프로그램이 파싱 과정을 거쳐서 제한적으로 의미 파악이 가능하다.

- [309] 제 6 정보 입력부(3310)는 W³C에서 그 스키마(Schema)가 정의되어 표준화된 콘텐츠 기반의 MathML(예컨대, cMathML)의 형태로 수식데이터를 입력받을 수 있다. 여기서 수식데이터로 cMathML을 예로 들었으니 제 6 실시예이 이에 한정되지는 않으며 다양한 방식으로 수식데이터가 일정한 양식으로 구조화되어 입력될 수 있다. 또한, 입력되는 수식데이터가 텍(TeX), OpenMath 등의 형태로 입력되는 경우, 제 6 정보 입력부(3310)는 입력된 텍(TeX), OpenMath 등의 형태의 데이터를 MathML 형태로 변환하여 제 6 수식데이터 구조화부(3320)로 전달할 수 있다. 또한, 수식데이터의 입력은 사용자의 조작 또는 명령에 의해 직접 입력될 수 있으나 반드시 이에 한정되는 것은 아니며, 별도의 외부 서버로부터 수식을 표현하고 있는 문서 데이터를 입력받을 수도 있을 것이다.
- [310] 한편, cMathML과 같은 XML 구조의 문서를 프로그래밍적으로 구조화하는 방법으로 DOM(Document Object Model)을 사용할 수 있다. DOM은 XML의 구조의 문서를 엘리먼트(Element)로 구분하여, 트리 구조로 만들어 주는 역할을 한다.
- [311] 즉, 제 6 수식데이터 구조화부(3320)는 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 MathML 형식의 수학적 입력이 DOM으로 처리된 트리구조를 제공한다.
- [312] 제 6 오퍼레이터 파싱부(3330)는 트리구조로 구조화된 오퍼레이터에 대하여 시멘틱의미를 추출하고 해당 오퍼레이터로부터 추출된 시멘틱 의미와 해당 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성할 수 있다. 제 6 오퍼레이터 파싱부(3330)는 기설정된 시멘틱의미DB(150)를 참조하여 해당 오퍼레이터의 시멘틱의미를 추출할 수도 있다.
- [313] 도 34 및 도 35는 수식이 표현된 수식데이터를 오퍼레이터 파싱을 수행한 결과를 예시한 도면이다.
- [314] 도 34에서 (A) 도시하듯이, 수식이 " $x^2+2x+6=0$ "인 경우에, (B)와 같이 cMathML 형태로 수식데이터가 표현되어 제 6 정보 입력부(3310)로 입력되면, 제 6 수식데이터 구조화부(3320)는 (B)와 같은 cMathML 형태의 수식데이터를 (C)와 같이 트리구조로 구조화할 수 있다.
- [315] 도 34에서, 제 6 수식데이터 구조화부(3320)가 구조화하는 (C)와 같은 트리구조에서 한 부모 노드 아래의 형제(Sibling) 노드 중에서 가장 좌측에는 'Plus', 'Power', 'Times', 'Eq' 등의 오퍼레이터 노드가 위치하며, 오퍼레이터 노드의

파라미터는 오퍼레이터 노드의 형제 노드로 위치하며, 다른 오퍼레이터 노드를 자식노드로 갖는 형제 노드인 경우에는 <Apply>와 같은 태그가 위치한다.

[316] 도 36은 cMathML 특성을 반영한 노드의 횡단순서를 예시한 도면이다.

[317] 도 36에 예시한 바와 같이, 수식데이터를 구조화하는 트리구조는 전위횡단(Pre-order Traversing) 기법으로 횡단될 수 있다. cMathML은 특징적으로 항을 표현함에 있어서 <apply></apply>를 기본으로 사용하고, 이는 어떤 노드의 자식 노드 중의 하나는 이 태그를 포함한다는 것을 의미한다. 따라서, 트리를 파싱할 때, <apply>를 포함한 노드를 제외한 노드에 대해서 먼저 정보를 추출한 후에 수집한 정보를 <apply>가 있는 노드로 전달한다. 그리고, <apply> 노드는 상위 노드로 정보를 전달하고, 상위 노드는 다시 동일한 레벨에 있는 <apply> 노드로 정보를 전달하는 식으로 최상위 노드에 도달할 때까지 계속해서 정보를 반복적으로 전달한다. 마지막으로 루트(root) 노드에 도달하게 되면, 모든 정보는 수집된 것이며, 루트 노드에서 원하는 시멘틱 정보를 얻을 수 있다. 여기서 트리구조가 전위횡단 기법으로 횡단될 수 있음을 설명하였으나 제 6 실시예에 이에 한정되지는 않는다.

[318] 이때, 제 6 오퍼레이터 파싱부(3330)는 트리구조를 횡단하면서 각 노드의 정보를 획득하고, 횡단하는 도중에 방문하는 노드에 존재하는 'Plus', 'Power', 'Times' 등의 오퍼레이터의 시멘틱 의미를 추출한다. 트리구조 상의 표현과 생성된 파싱결과 표현이 서로 다른 경우에는 트리구조 상의 표현과 대응되는 파싱결과 표현을 저장하는 시멘틱의미DB(150)를 구비하고, 제 6 오퍼레이터 파싱부(3330)가 오퍼레이터의 시멘틱의미를 추출함에 있어서 시멘틱의미DB(150)를 참조할 수도 있고, 트리구조 상의 표현과 생성된 파싱결과 표현이 서로 다른 경우에는 구조화된 트리구조에 포함되는 경우에는 'Plus', 'Power', 'Times' 등의 정보를 그대로 참조할 수도 있다.

[319] 제 6 오퍼레이터 파싱부(3330)는 오퍼레이터의 시멘틱의미를 추출하고, 구조화된 트리구조에서 오퍼레이터에 연관된 파라미터를 추출하여 이를 오퍼레이터의 시멘틱의미와 결합하여 (D)와 같은 파싱 시멘틱 정보를 생성한다. 즉, 형제 노드 중에서 오퍼레이터의 파라미터(예컨대, 'Power'의 형제노드는 'Ci' 및 'Cn'이며, 'Ci' 및 'Cn'은 각각 'x' 및 '2'라는 자식노드에 연결됨으로써 'x' 및 '2'가 오퍼레이터인 'Power'에 결합된다.)에 대하여는 "Power[x, 2]", "Times[2, x]"와 같이 오퍼레이터와 결합하여 표현한다.

[320] 한편, 제 6 오퍼레이터 파싱부(3330)는 트리구조를 파싱하면서 수식의 연산의 종류, 변수개수, 항의 차수 등을 포함하는 시멘틱 정보를 추출할 수 있다. 즉, 제 6 오퍼레이터 파싱부(3330)가 추출하는 시멘틱 정보는 어느 한 노드만을 방문함으로써 추출되는 것이 아니고, 모든 노드를 방문하여, 예컨대 오퍼레이터에 대한 변수의 개수 및 항의 차수 등에 대한 정보를 노드 방문이 완료될 때까지 저장하고 있다가 노드 방문이 완료된 후에 해당 수식데이터의 종류 및 특징 등을 나타내는 종합 시멘틱 정보를 추출하여 파싱 시멘틱 정보에

포함할 수 있다.

- [321] 도 35의 경우에는, (A) 도시하듯이, 수식이 [수학식 2]와 같은 경우에, (B)와 같이 cMathML 형태로 수식데이터가 표현되어 제 6 정보 입력부(3310)로 입력되면, 제 6 수식데이터 구조화부(3320)는 (B)와 같은 cMathML 형태의 수식데이터를 (C)와 같이 트리구조로 구조화할 수 있다.
- [322] 수학식 2

$$A := \{1, 2\}$$

$$B := \{3, 4\}$$

$$A \cup B$$

- [323] 도 35에서, 제 6 수식데이터 구조화부(3320)가 구조화하는 (C)와 같은 트리구조에서 한 부모 노드 아래의 형제(Sibling) 노드 중에서 가장 좌측에는 'Union', 'Set', 'Ci' 등의 오퍼레이터 노드가 위치하며, 오퍼레이터 노드의 파라미터는 오퍼레이터 노드의 형제 노드로 위치하며, 다른 오퍼레이터 노드를 자식노드로 갖는 형제 노드인 경우에는 <Apply>, <Declare>와 같은 태그가 위치한다.
- [324] 이때, 제 6 오퍼레이터 파싱부(3330)는 트리구조를 횡단하면서 각 노드의 정보를 획득하고, 횡단하는 도중에 방문하는 노드에 존재하는 'Union', 'Set', 'Ci' 등의 오퍼레이터의 시멘틱 의미를 추출한다.
- [325] 제 6 오퍼레이터 파싱부(3330)는 (C)와 같은 트리구조를 횡단하면서 오퍼레이터의 시멘틱의미를 추출하고, 구조화된 트리구조에서 오퍼레이터에 연관된 파라미터를 추출하여 이를 오퍼레이터의 시멘틱의미와 결합하여 (D)와 같은 파싱 시멘틱 정보(Semantic Result)를 생성한다. 즉, 형제 노드 중에서 오퍼레이터의 파라미터(예컨대, 'Union'의 형제노드는 두 개의 'Ci', 두 개의 'Ci'는 각각 'A' 및 'B'라는 자식노드에 연결됨으로써 'A' 및 'B'가 오퍼레이터인 'Ci'에 결합된다.)에 대하여는 "Union[A, B]"와 같이 오퍼레이터와 결합하여 표현한다. 또한, 파라미터 역시 트리구조에서 "Declare" 태그를 참조하여 그 시멘틱의미를 추출할 수 있다.
- [326] 도 37은 파싱 시멘틱 정보(b)와 입력된 수식데이터(a)를 결합한 시멘틱 정보 결합 수식데이터를 예시한 도면이다.
- [327] 도 37에 예시하듯이, 제 6 시멘틱 정보 결합부(3340)는 도 34에서와 같은 수식데이터(a)와 제 6 오퍼레이터 파싱부(3330)에서 생성된 파싱 시멘틱 정보(b)

- 결합하여 시멘틱 정보 결합 수식데이터를 생성한다.
- [328] 즉, 생성되는 시멘틱 정보 결합 수식데이터(a+b)는 입력되는 수식데이터(a)와 같이 XML 형태의 기설정된 스키마를 갖는 구조로 생성될 수 있으며, 도 37에 도시한 바와 같이 파싱 시멘틱 정보(b)가 <Semantic></Semantic> 태그로 XML 형태의 수식데이터(a) 다음에 삽입되는 구조로 생성될 수 있다.
- [329] 도 38은 노드를 횡단하는 도중에 노드간에 데이터를 전달하는 데이터구조를 예시한 도면이다.
- [330] 도 38은 방정식의 저장을 위한 데이터 구조 템플릿을 예시한 것이며, 이는 다른 데이터 저장 구조에도 쉽게 확장이 될 수 있다. 수학식을 큰 그룹으로 나누면 다항식(Polynomial), 행렬(Matrix), 집합(Set), 벡터(Vector), 릴레이션(Relationship), 적분(Integration), 미분(Differentiation) 등이 될 수 있다. 이 그룹들은 위의 템플릿과 유사한 데이터 구조를 가질 수 있으며, 새로운 구조가 추가될 경우, 이 템플릿을 바탕으로 확장할 수 있다.
- [331] 도 34의 경우와 같이, 트리 구조 내에서 복수의 노드를 자식노드로 갖는 경우에, 자식노드의 오퍼레이터 노드 및 파라미터 노드에 대한 정보를 저장할 수 있다. 예를 들어, 도 38과 같이, 노드에 대한 정보에는 변수 세트와 같은 저장구조가 포함될 수 있으며, 변수세트에는 변수 이름과 차수에 해당하는 정보가 들어갈 수 있다. 저장되는 변수 세트에는 한 개 이상의 변수가 들어갈 수 있으며, 저장되는 변수 세트에는 또 다른 변수 세트가 내부에 포함됨으로써 저장구조가 내포된(nested) 구조를 가질 수도 있다.
- [332] 도 39는 제 6 실시예에 따른 자연어 및 수식 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하는 시스템에 대한 예시도이다.
- [333] 제 6 실시예에 따른 수학적 및 자연어 처리 장치가 클라우드 컴퓨팅으로 데이터를 제공하기 위해서는 단말기(910), 통신망(920) 및 제 6 클라우드 컴퓨팅 장치(3900)를 포함한 시스템이 필요하다.
- [334] 여기서, 단말기(910)는 사용자의 명령 또는 조작에 따라 통신망(920)을 경유하여 각종 데이터를 송수신할 수 있는 단말기를 말하는 것이며, 태블릿 PC, 랩톱, 개인용 컴퓨터, 스마트폰, 개인휴대용 정보단말기 및 무선 통신 단말기 등 중 어느 하나일 수 있다. 또한, 단말기(910)는 통신망(920)을 통하여 데이터 읽거나 입력 및 저장, 네트워크, 콘텐츠 사용 등의 서비스를 이용할 수 있는 클라우드 컴퓨팅을 지원하는 클라우드 컴퓨팅 단말기가 될 수 있다. 즉, 단말기(910)는 통신망(920)을 경유하여 제 6 클라우드 컴퓨팅 장치(3900)에 접속하기 위한 프로그램을 저장하기 위한 메모리, 프로그램을 실행하여 연산 및 제어하기 위한 마이크로프로세서 등을 구비하고 있는 장치를 의미한다. 즉, 단말기(910)로는 통신망(920)에 연결되어 제 6 클라우드 컴퓨팅 장치(3900)와 클라이언트-서버 통신이 가능하다면 그 어떠한 단말기도 가능하며, 노트북 컴퓨터, 이동통신 단말기, PDA 등 여하한 통신 컴퓨팅 장치를 모두 포함하는 넓은 개념이다. 한편, 단말기(910)는 터치 스크린을 구비한 형태로 제작되는 것이

바람직하나 반드시 이에 한정되는 것은 아니다.

- [335] 단말기(910)는 제 6 클라우드 컴퓨팅 장치(3900)로 수식이 표현된 수식데이터를 입력하고 제 6 클라우드 컴퓨팅 장치(3900)는 통해 클라우드 컴퓨팅 방식으로 수식데이터를 파싱하여 수식데이터의 시멘틱 정보를 추출하여 단말기(910)로 제공할 수 있다. 즉, 단말기(910)는 제 6 클라우드 컴퓨팅 장치(3900)로부터 클라우드 컴퓨팅 방식으로 데이터의 입출력을 위하여 제 6 클라우드 컴퓨팅 장치(3900)와의 입출력 인터페이스를 제공하는 별도의 입출력 인터페이스부를 포함할 수 있으며, 입출력 인터페이스부를 통해 제 6 클라우드 컴퓨팅 장치(3900)에 저장된 저장 매체에 대한 데이터 읽기 및 쓰기가 수행되도록 하는 인터페이스 제어부를 포함할 수 있다. 이에 대해 좀 더 구체적으로 설명하면, 단말기(910)는 입출력 인터페이스부를 통해 수식이 표현된 수식데이터를 제 6 클라우드 컴퓨팅 장치(3900)로 입력할 수 있고, 제 6 클라우드 컴퓨팅 장치(3900)는 수식이 표현된 수식데이터를 입력받고 수신한 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하고 구조화된 오퍼레이터에 대하여 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하므로, 실질적으로 단말기(910)에서는 어떠한 어플리케이션의 탑재 없이도 수식데이터를 파싱하여 시멘틱 정보를 추출할 수 있다.
- [336] 통신망(920)은 인터넷망, 인트라넷망, 이동통신망, 위성 통신망 등 다양한 유무선 통신 기술을 이용하여 인터넷 프로토콜로 데이터를 송수신할 수 있는 망을 말하며, 단말기(910)와 제 6 클라우드 컴퓨팅 장치(3900) 간에 데이터를 중계하는 기능을 수행한다.
- [337] 제 6 클라우드 컴퓨팅 장치(3900)는 자연어 및 수식 처리 장치(100)를 기반으로 구현될 수 있으며, 클라우드 컴퓨팅으로 단말기(910)에게 수식데이터의 파싱 시멘틱 정보를 제공할 수 있도록, 단말기(910)로 하여금 제 6 클라우드 컴퓨팅 장치(3900)에 저장된 저장 매체에 대한 데이터의 읽기 및 쓰기가 수행되도록 하되, 수식데이터가 입력되면, 수신한 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하고 구조화된 오퍼레이터에 대하여 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하여 컴퓨터로 읽을 수 있는 기록매체에 저장하며, 해당 기록매체의 데이터를 단말기(910)로 전송함으로써, 단말기(910)에서 어플리케이션의 탑재 없이 수식데이터를 파싱할 수 있도록 하는 클라우드 컴퓨팅을 제공할 수 있다. 즉, 제 6 클라우드 컴퓨팅 장치(3900)는 수식데이터의 시멘틱 정보를 추출하는 제 6 시멘틱 정보 생성부(3910)와 단말기(910)로 하여금 제 6 시멘틱 정보 생성부(3910)에 의해 저장 매체에 저장된 데이터의 읽기 및 쓰기가 수행되도록 하는 제 6 클라우드 컴퓨팅부(3920)를 구비할 수 있다.
- [338] 도 40은 제 6 실시예에 따른 수학적 시멘틱 정보 생성 방법을 설명하기 위한

순서도이다.

- [339] 제 6 실시예에 따른 수학적 시멘틱 정보 생성 방법은, 수식이 표현된 수식데이터를 입력받는 정보입력단계(S4010), 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하는 수식데이터 구조화단계(S4020), 구조화된 오퍼레이터에 대하여 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하는 오퍼레이터 파싱단계(S4030) 및 파싱 시멘틱 정보 및 수식데이터를 결합하여 시멘틱 정보 결합 수식데이터를 생성하는 시멘틱 정보 결합단계(S4040)를 포함할 수 있다.
- [340] 여기서, 정보입력단계(S4010), 수식데이터 구조화단계(S4020), 오퍼레이터 파싱단계(S4030) 및 시멘틱 정보 결합단계(S4040)는 각각 제 6 정보 입력부(3310), 제 6 수식데이터 구조화부(3320), 제 6 오퍼레이터 파싱부(3330) 및 제 6 시멘틱 정보 결합부(3340)의 동작에 대응되므로 상세한 설명은 생략한다.
- [341] 이상의 설명은 본 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 실시예들은 본 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

산업상 이용가능성

- [342] 이상에서 설명한 바와 같이 본 실시예는 자연어 및 수식을 처리하는 다양한 분야에 적용되어, 사용자가 자연어와 수식을 입력할 수 있도록 별도의 입력 툴을 제공하고, 자연어와 수식이 조합된 데이터의 분석 내용에 기초하여 자연어 및 수식을 재조합한 데이터로 관리할 수 있도록 하는 구조화하고, 시멘틱 정보에 기초하여 사용자 쿼리를 구조화한 정보가 같이 색인화되도록 하며, 의미론적인 시멘틱 정보를 자동으로 추출하고, 자연어와 수학적식이 포함된 복합 문장을 자동적으로 논리적인 관계를 갖도록 표현하며, 추가적으로 시멘틱 정보를 생성할 수 있는 효과를 발생하는 유용한 발명이다.

[343]

[344] CROSS-REFERENCE TO RELATED APPLICATION

- [345] 본 특허출원은 2010. 12. 02.에 한국에 출원한 특허출원번호 제10-2010-0122025호, 2010. 12. 23.에 한국에 출원한 특허출원번호 제10-2010-0133761호, 2010. 12. 22.에 한국에 출원한 특허출원번호 제10-2010-0132141호, 2010. 12. 30.에 한국에 출원한 특허출원번호 제10-2010-0138531호, 2011. 01. 06.에 한국에 출원한 특허출원번호

제10-2011-0001282호, 2011. 02. 21.에 한국에 출원한 특허출원번호
제10-2011-0014968호에 대해 미국 특허법 119(a)조(35 U.S.C § 119(a))에 따라
우선권을 주장하면, 그 모든 내용은 참고문헌으로 본 특허출원에 병합된다.
아울러, 본 특허출원은 미국 이외에 국가에 대해서도 위와 동일한 이유로
우선권을 주장하면 그 모든 내용은 참고문헌으로 본 특허출원에 병합된다.

[346]

청구범위

[청구항 1]

자연어 및 수식을 입력받는 자연어 및 수식 입력부;
 상기 자연어 및 상기 수식의 조합으로 이루어진 조합
 데이터로부터 수학적식에 대한 파싱 시멘틱 정보를 생성하는 정보
 생성부;
 상기 조합 데이터에서 논리적 조건을 이용한 동작 정보를
 추출하는 동작 정보 추출부;
 상기 조합 데이터를 분석하여 특정 의미에 따라 구분한 후
 재조합하는 자연어 및 수식 구조화부;
 상기 동작 정보를 구조화하는 동작 구조화부; 및
 상기 조합 데이터를 색인화하는 자연어 및 수식 색인화부
 를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 2]

제 1 항에 있어서,
 상기 자연어 및 수식 입력부는,
 상기 자연어를 입력받기 위한 텍스트 입력 툴을 제공하는 제 1
 자연어 입력 처리기;
 상기 수식을 입력받기 위한 수식 입력 툴을 제공하는 제 1 수식
 입력 처리기;
 입력된 상기 자연어 및 상기 수식을 취합한 취합 데이터를
 전달하는 제 1 정보 처리부;
 상기 취합 데이터를 수신하며, 상기 자연어와 상기 수식을
 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라
 구분하는 시멘틱 정보를 생성하는 제 1 파싱부; 및
 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중
 적어도 하나 이상의 정보를 재조합하여 저장하는 제 1 데이터
 관리부
 를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 3]

제 1 항에 있어서,
 상기 자연어 및 수식 구조화부는,
 상기 조합 데이터를 입력받는 제 2 정보 입력부;
 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하는
 제 2 분리부;
 분리된 상기 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여
 특정 의미에 따라 구분하는 자연어 제 2 처리부;
 분리된 상기 수식을 구성하고 있는 각각의 제 2 정보를 분석하여
 특정 의미에 따라 구분하는 제 2 수식 처리부; 및
 상기 제 1 정보, 상기 제 2 정보, 상기 자연어 및 상기 수식 중

적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 2 데이터 관리부

를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 4]

제 1 항에 있어서,

상기 자연어 및 수식 색인화부는,

상기 조합 데이터를 입력받는 제 3 정보 입력부;

상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하고,

분리된 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성

정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를

생성하는 제 3 시멘틱 파서부;

상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중

적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는

제 3 데이터 관리부;

입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하는 제 3

쿼리 파서부; 및

상기 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 상기

시멘틱 색인 정보에 상기 키워드 정보를 매칭한 쿼리 색인 정보를

생성하는 제 3 색인부

를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 5]

제 1 항에 있어서,

상기 동작 정보 추출부는,

상기 조합 데이터를 입력받는 제 4 정보 입력부;

상기 조합 데이터에서 상기 자연어 및 상기 수식을 분리하는 제 4

분리부;

분리된 상기 자연어를 토큰화하여 자연어토큰을 생성하는 제 4

자연어처리부;

분리된 상기 수식을 파싱하고 시멘틱의미를 추출하여

수학식토큰을 생성하는 제 4 수식처리부;

자연어와 수학식의 논리적 조건 및 상기 논리적 조건에 대응되는

상기 동작 정보를 결합한 룰을 저장하는 제 4 룰저장부; 및

상기 생성된 자연어토큰과 수학식토큰을 상기 저장된 룰의 논리적

조건과 비교하여 상기 저장된 룰로부터 상기 조합 데이터에 대한

상기 동작 정보를 추출하는 제 4 동작추출부

를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 6]

제 1 항에 있어서,

상기 동작 구조화부는,

상기 조합 데이터를 입력받는 제 5 정보 입력부;

상기 조합 데이터의 문장구성을 분석하여 상기 수식 및 상기

자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하는 제 5 문장분석부;

자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는 상기 동작 정보를 추출하는 제 5 동작추출부; 및

상기 추출된 동작 정보를 상기 수식토큰에 대하여 구조화하는 제 5 동작실행부

를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 7]

제 1 항에 있어서,

상기 정보 생성부는,

상기 수식이 표현된 수식데이터를 입력받는 제 6 정보 입력부;

상기 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여

구조화하는 제 6 수식데이터 구조화부; 및

구조화된 상기 오퍼레이터에 대하여 상기 오퍼레이터의

시멘틱의미를 추출하고 추출된 시멘틱 의미 및 상기 오퍼레이터와

연관된 파라미터를 결합하여 상기 파싱 시멘틱 정보를 생성하는

제 6 오퍼레이터 파싱부

를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 8]

자연어를 입력받기 위한 텍스트 입력 툴을 제공하는 제 1 자연어 입력 처리기;

수식을 입력받기 위한 수식 입력 툴을 제공하는 제 1 수식 입력 처리기;

입력된 상기 자연어 및 상기 수식을 취합한 취합 데이터를 전달하는 제 1 정보 처리부;

상기 취합 데이터를 수신하며, 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 1 파싱부; 및

상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 저장하는 제 1 데이터 관리부

를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.

[청구항 9]

자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 2 정보 입력부;

상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하는 제 2 분리부;

분리된 상기 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하는 제 2 자연어 처리부;

분리된 상기 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하는 제 2 수식 처리부; 및

- 상기 제 1 정보, 상기 제 2 정보, 상기 자연어 및 상기 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 2 데이터 관리부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.
- [청구항 10] 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 3 정보 입력부; 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하고, 분리된 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 3 시멘틱 파서부; 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 3 데이터 관리부; 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하는 제 3 쿼리 파서부; 및 상기 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 상기 시멘틱 색인 정보에 상기 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 제 3 색인부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.
- [청구항 11] 자연어와 수식이 포함된 복합문장을 수신하는 제 4 정보 입력부; 상기 복합문장에서 상기 자연어 및 상기 수식을 분리하는 제 4 분리부; 분리된 상기 자연어를 토큰화하여 자연어토큰을 생성하는 제 4 자연어처리부; 분리된 상기 수식을 파싱하고 시멘틱의미를 추출하여 수학적토큰을 생성하는 제 4 수식처리부; 자연어와 수학적의 논리적 조건 및 상기 논리적 조건에 대응되는 동작 정보를 결합한 룰을 저장하는 제 4 룰저장부; 및 상기 생성된 자연어토큰과 수학적토큰을 상기 저장된 룰의 논리적 조건과 비교하여 상기 저장된 룰로부터 상기 복합문장의 동작 정보를 추출하는 제 4 동작추출부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.
- [청구항 12] 자연어와 수식이 포함된 복합문장을 수신하는 제 5 정보 입력부; 상기 복합문장의 문장구성을 분석하여 수식데이터 및 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하는 제 5 문장분석부; 자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는 동작 정보를 추출하는 제 5 동작추출부; 및

- 상기 추출된 동작 정보를 상기 수식토큰에 대하여 구조화하는 제 5 동작실행부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.
- [청구항 13] 수식이 표현된 수식데이터를 입력받는 제 6 정보 입력부; 상기 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하는 제 6 수식데이터 구조화부; 및 구조화된 상기 오퍼레이터에 대하여 상기 오퍼레이터의 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 상기 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하는 제 6 오퍼레이터 파싱부를 포함하는 것을 특징으로 하는 자연어 및 수식 처리 장치.
- [청구항 14] 자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서, 자연어 및 수식을 입력받는 자연어 및 수식 입력 과정; 상기 자연어 및 상기 수식의 조합으로 이루어진 조합 데이터로부터 수학적식에 대한 파싱 시멘틱 정보를 생성하는 정보 생성 과정; 상기 조합 데이터에서 논리적 조건을 이용한 동작 정보를 추출하는 동작 정보 추출 과정; 상기 조합 데이터를 분석하여 특정 의미에 따라 구분한 후 재조합하는 자연어 및 수식 구조화 과정; 상기 동작 정보를 구조화하는 동작 구조화 과정; 및 상기 조합 데이터를 색인화하는 자연어 및 수식 색인화 과정을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법.
- [청구항 15] 제 14 항에 있어서, 상기 자연어 및 수식 입력 과정은, 상기 자연어를 입력받기 위한 텍스트 입력 툴을 제공하는 제 1 자연어 입력 처리 과정; 상기 수식을 입력받기 위한 수식 입력 툴을 제공하는 제 1 수식 입력 처리 과정; 입력된 상기 자연어 및 상기 수식을 취합한 취합 데이터를 전달하는 제 1 정보 처리 과정; 상기 취합 데이터를 수신하며, 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 1 파싱 과정; 및 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 저장하는 제 1 데이터 관리 과정

[청구항 16]

를 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법.
 제 14 항에 있어서,
 상기 자연어 및 수식 구조화 과정은,
 상기 조합 데이터를 입력받는 제 2 정보 입력 과정;
 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하는 제 2 분리 과정;
 분리된 상기 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하는 자연어 제 2 처리 과정;
 분리된 상기 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하는 제 2 수식 처리 과정; 및
 상기 제 1 정보, 상기 제 2 정보, 상기 자연어 및 상기 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 2 데이터 관리 과정
 를 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법.

[청구항 17]

제 14 항에 있어서,
 상기 자연어 및 수식 색인화 과정은,
 상기 조합 데이터를 입력받는 제 3 정보 입력 과정;
 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하고, 분리된 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 3 시멘틱 파서 과정;
 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 3 데이터 관리 과정;
 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하는 제 3 쿼리 파서 과정; 및
 상기 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 상기 시멘틱 색인 정보에 상기 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 제 3 색인 과정
 를 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법.

[청구항 18]

제 14 항에 있어서,
 상기 동작 정보 추출부 과정은,
 상기 조합 데이터를 입력받는 제 4 정보 입력 과정;
 상기 조합 데이터에서 상기 자연어 및 상기 수식을 분리하는 제 4 분리 과정;
 분리된 상기 자연어를 토큰화하여 자연어토큰을 생성하는 제 4 자연어처리 과정;
 분리된 상기 수식을 파싱하고 시멘틱의미를 추출하여

수학식토큰을 생성하는 제 4 수식처리 과정;
 자연어와 수학식의 논리적 조건 및 상기 논리적 조건에 대응되는
 상기 동작 정보를 결합한 룰을 저장하는 제 4 룰저장 과정; 및
 상기 생성된 자연어토큰과 수학식토큰을 상기 저장된 룰의 논리적
 조건과 비교하여 상기 저장된 룰로부터 상기 조합 데이터의 상기
 동작 정보를 추출하는 제 4 동작추출 과정
 을 포함하는 것을 특징으로 하는 자연어 및 수학식 처리 방법.

[청구항 19]

제 14 항에 있어서,
 상기 동작 구조화부 과정은,
 상기 조합 데이터를 입력받는 제 5 정보 입력 과정;
 상기 조합 데이터의 문장구성을 분석하여 상기 수식 및 상기
 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하는 제
 5 문장분석 과정;
 자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는
 상기 동작 정보를 추출하는 제 5 동작추출 과정; 및
 상기 추출된 동작 정보를 상기 수식토큰에 대하여 구조화하는 제 5
 동작실행 과정
 을 포함하는 것을 특징으로 하는 자연어 및 수학식 처리 방법.

[청구항 20]

제 14 항에 있어서,
 상기 정보 생성부 과정은,
 상기 수식이 표현된 수식데이터를 입력받는 제 6 정보 입력 과정;
 상기 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여
 구조화하는 제 6 수식데이터 구조화 과정; 및
 구조화된 상기 오퍼레이터에 대하여 상기 오퍼레이터의
 시멘틱의미를 추출하고 추출된 시멘틱 의미 및 상기 오퍼레이터와
 연관된 파라미터를 결합하여 상기 파싱 시멘틱 정보를 생성하는
 제 6 오퍼레이터 파싱 과정
 을 포함하는 것을 특징으로 하는 자연어 및 수학식 처리 방법.

[청구항 21]

자연어 및 수식 처리 장치가 자연어 및 수학을 처리하는 방법에
 있어서,
 자연어를 입력받기 위한 텍스트 입력 툴을 제공하는 제 1 자연어
 입력 처리 과정;
 수학을 입력받기 위한 수식 입력 툴을 제공하는 제 1 수식 입력
 처리 과정;
 입력된 상기 자연어 및 상기 수학을 취합한 취합 데이터를
 전달하는 제 1 정보 처리 과정;
 상기 취합 데이터를 수신하며, 상기 자연어와 상기 수학을
 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라

- 구분하는 시멘틱 정보를 생성하는 제 1 파싱 과정; 및
 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 저장하는 제 1 데이터 관리 과정
- [청구항 22] 을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법.
 자연어 및 수식 처리 장치가 자연어 및 수학을 처리하는 방법에 있어서,
 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 2 정보 입력 과정;
 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하는 제 2 분리 과정;
 분리된 상기 자연어를 구성하고 있는 각각의 제 1 정보를 분석하여 특정 의미에 따라 구분하는 제 2 자연어 처리 과정;
 분리된 상기 수식을 구성하고 있는 각각의 제 2 정보를 분석하여 특정 의미에 따라 구분하는 제 2 수식 처리 과정; 및
 상기 제 1 정보, 상기 제 2 정보, 상기 자연어 및 상기 수식 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 2 데이터 관리 과정
- [청구항 23] 을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법.
 자연어 및 수식 처리 장치가 자연어 및 수학을 처리하는 방법에 있어서,
 자연어 및 수식의 조합으로 이루어진 조합 데이터를 입력받는 제 3 정보 입력 과정;
 상기 조합 데이터에서 상기 자연어 및 상기 수식을 각각 분리하고, 분리된 상기 자연어와 상기 수식을 구성하고 있는 각각의 구성 정보를 분석하여 특정 의미에 따라 구분하는 시멘틱 정보를 생성하는 제 3 시멘틱 파서 과정;
 상기 구성 정보, 상기 자연어, 상기 수식 및 상기 시멘틱 정보 중 적어도 하나 이상의 정보를 재조합하여 재조합 데이터로 저장하는 제 3 데이터 관리 과정;
 입력된 사용자 쿼리에 포함된 키워드를 추출하여 구조화하는 제 3 쿼리 파서 과정; 및
 상기 시멘틱 정보를 색인화한 시멘틱 색인 정보를 생성하고, 상기 시멘틱 색인 정보에 상기 키워드 정보를 매칭한 쿼리 색인 정보를 생성하는 제 3 색인 과정
- [청구항 24] 을 포함하는 것을 특징으로 하는 자연어 및 수학적 처리 방법.
 자연어 및 수식 처리 장치가 자연어 및 수학을 처리하는 방법에 있어서,

자연어와 수식이 포함된 복합문장을 수신하는 제 4 정보 입력 과정;

상기 복합문장에서 상기 자연어 및 상기 수식을 분리하는 제 4 분리 과정;

분리된 상기 자연어를 토큰화하여 자연어토큰을 생성하는 제 4 자연어처리 과정;

분리된 상기 수식을 파싱하고 시멘틱의미를 추출하여 수학적토큰을 생성하는 제 4 수식처리 과정;

자연어와 수학적식의 논리적 조건 및 상기 논리적 조건에 대응되는 동작 정보를 결합한 룰을 저장하는 제 4 룰저장 과정; 및

상기 생성된 자연어토큰과 수학적토큰을 상기 저장된 룰의 논리적 조건과 비교하여 상기 저장된 룰로부터 상기 복합문장의 동작 정보를 추출하는 제 4 동작추출 과정

을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법.

[청구항 25]

자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서,

자연어와 수식이 포함된 복합문장을 수신하는 제 5 정보 입력 과정;

상기 복합문장의 문장구성을 분석하여 수식데이터 및 자연어를 각각 토큰화하여 수식토큰 및 자연어토큰을 생성하는 제 5 문장분석 과정;

자연어토큰 규칙을 참조하여 상기 자연어토큰의 의미에 대응되는 동작 정보를 추출하는 제 5 동작추출 과정; 및

상기 추출된 동작 정보를 상기 수식토큰에 대하여 구조화하는 제 5 동작실행 과정

을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법.

[청구항 26]

자연어 및 수식 처리 장치가 자연어 및 수학적식을 처리하는 방법에 있어서,

수식이 표현된 수식데이터를 입력받는 제 6 정보 입력 과정;

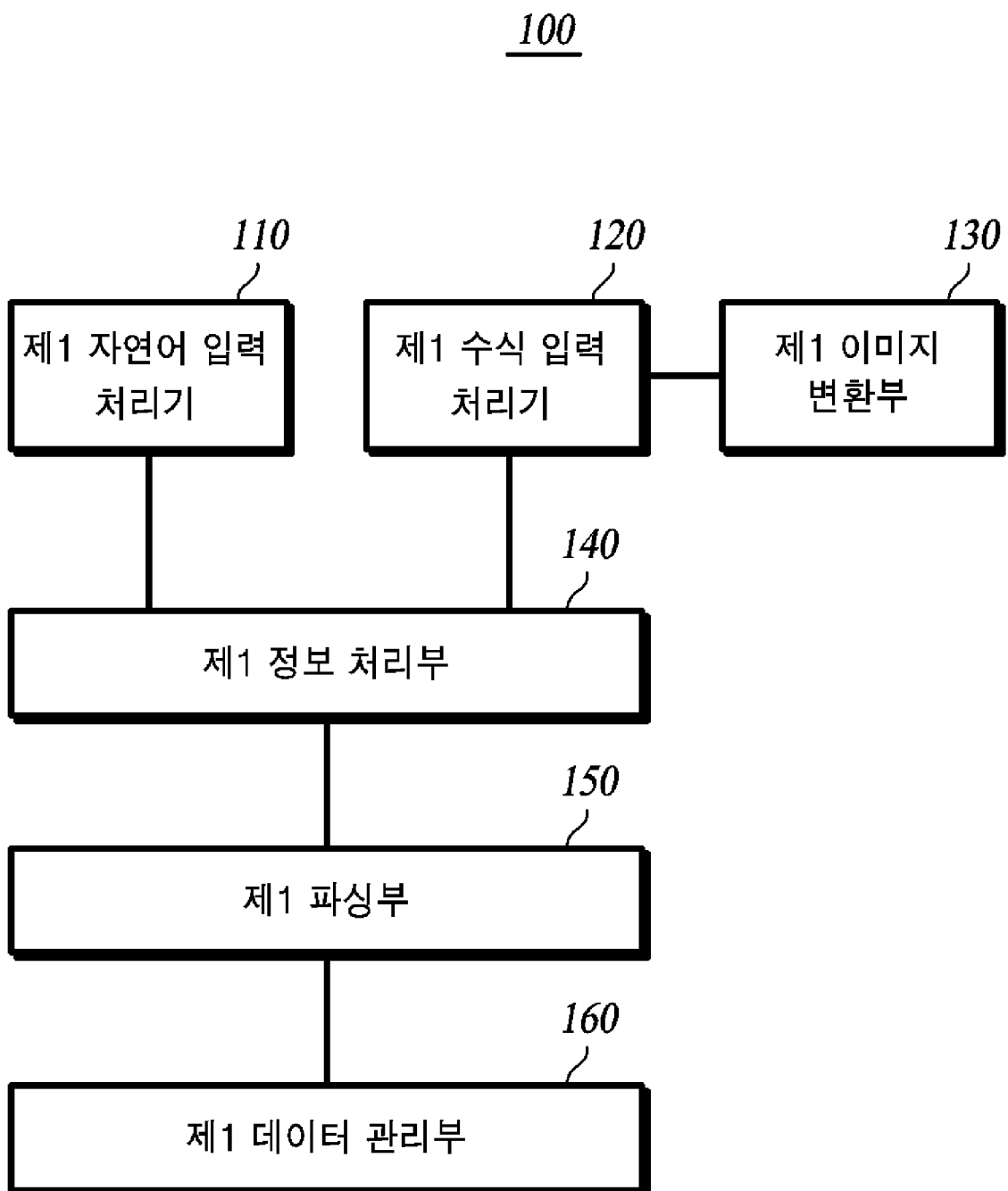
상기 수식데이터로부터 오퍼레이터 및 파라미터를 추출하여 구조화하는 제 6 수식데이터 구조화 과정; 및

구조화된 상기 오퍼레이터에 대하여 상기 오퍼레이터의

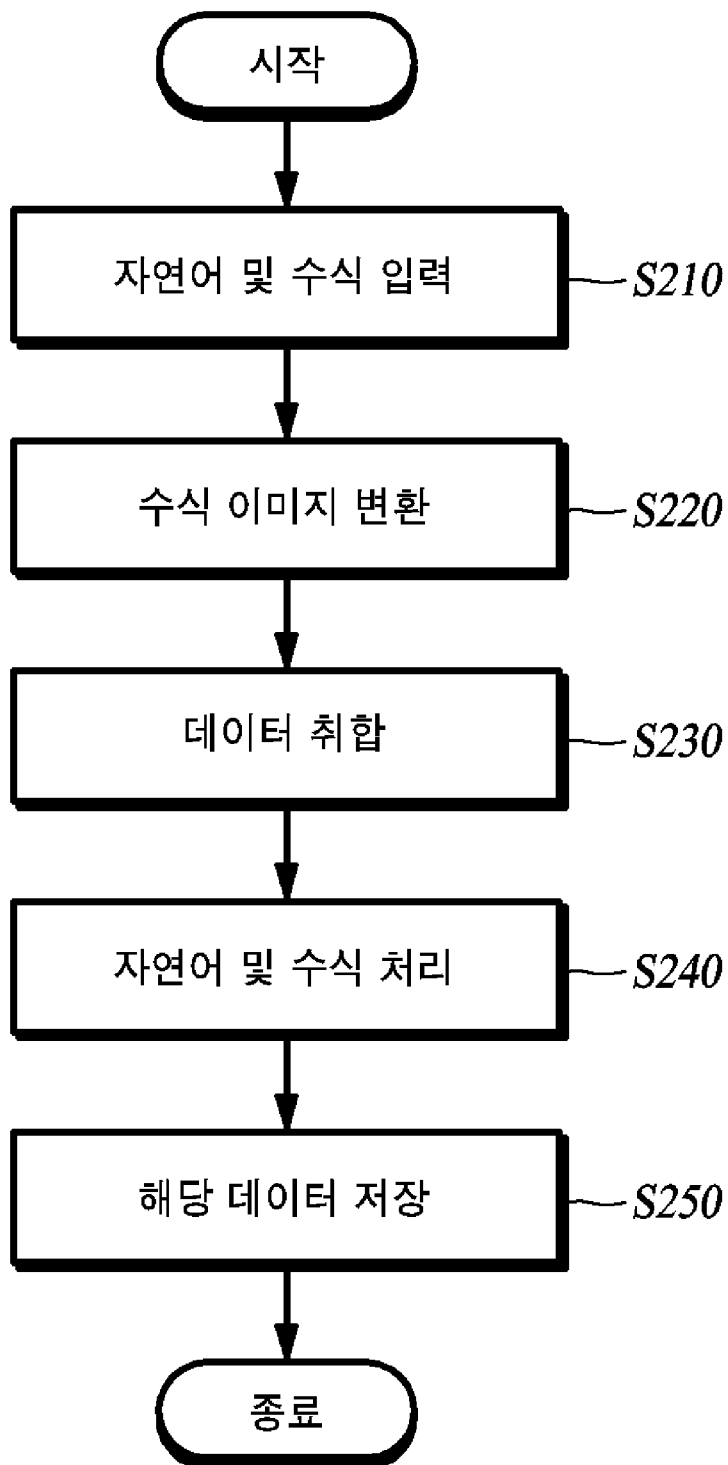
시멘틱의미를 추출하고 추출된 시멘틱 의미 및 상기 오퍼레이터와 연관된 파라미터를 결합하여 파싱 시멘틱 정보를 생성하는 제 6 오퍼레이터 파싱 과정

을 포함하는 것을 특징으로 하는 자연어 및 수학적식 처리 방법.

[Fig. 1]



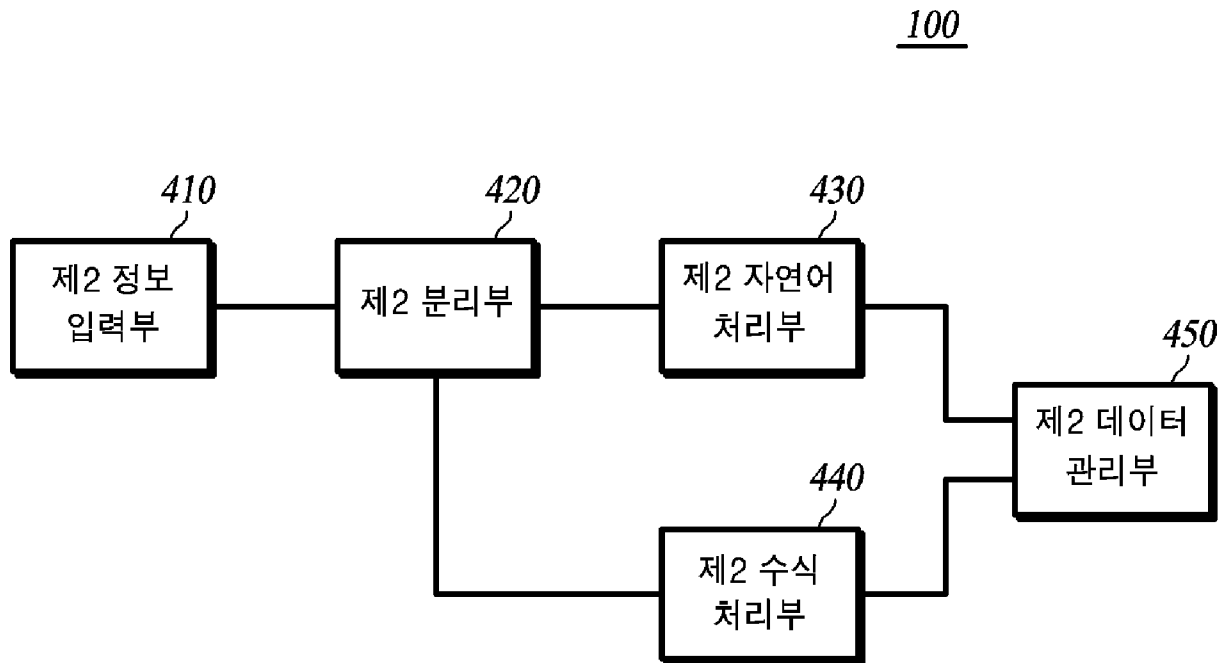
[Fig. 2]



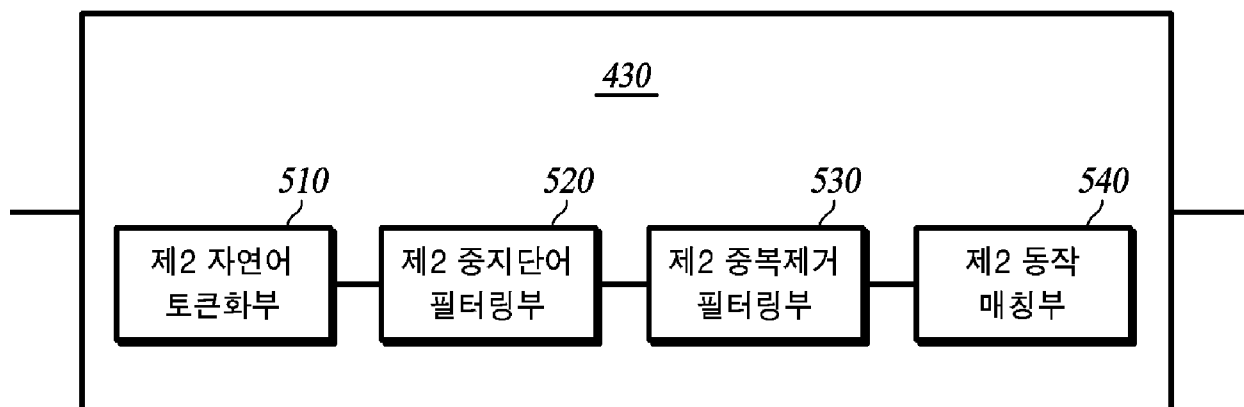
[Fig. 3]

```
<?xml version="1.0" encoding="UTF-8"?>
<MathProblem>
  <MathDescription>
    <Subjects>
      <Subject>
        // 문제 제목
      </Subject>
    </Subjects>
    <Mathbody>
      // MathML 수학식
    </Mathbody>
    <Text>
      // 자연어
    </Text>
  </MathDescription>
</MathProblem>
```

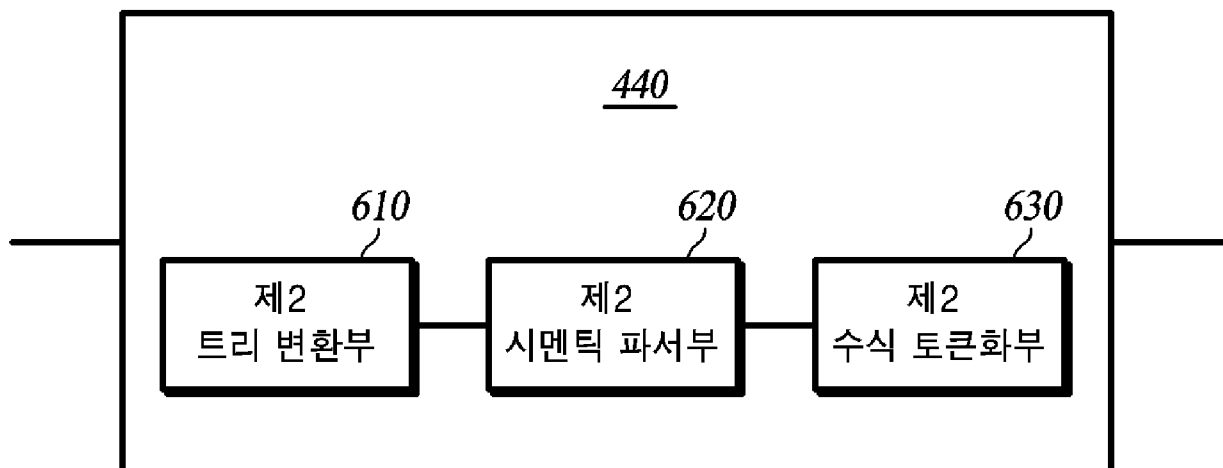
[Fig. 4]



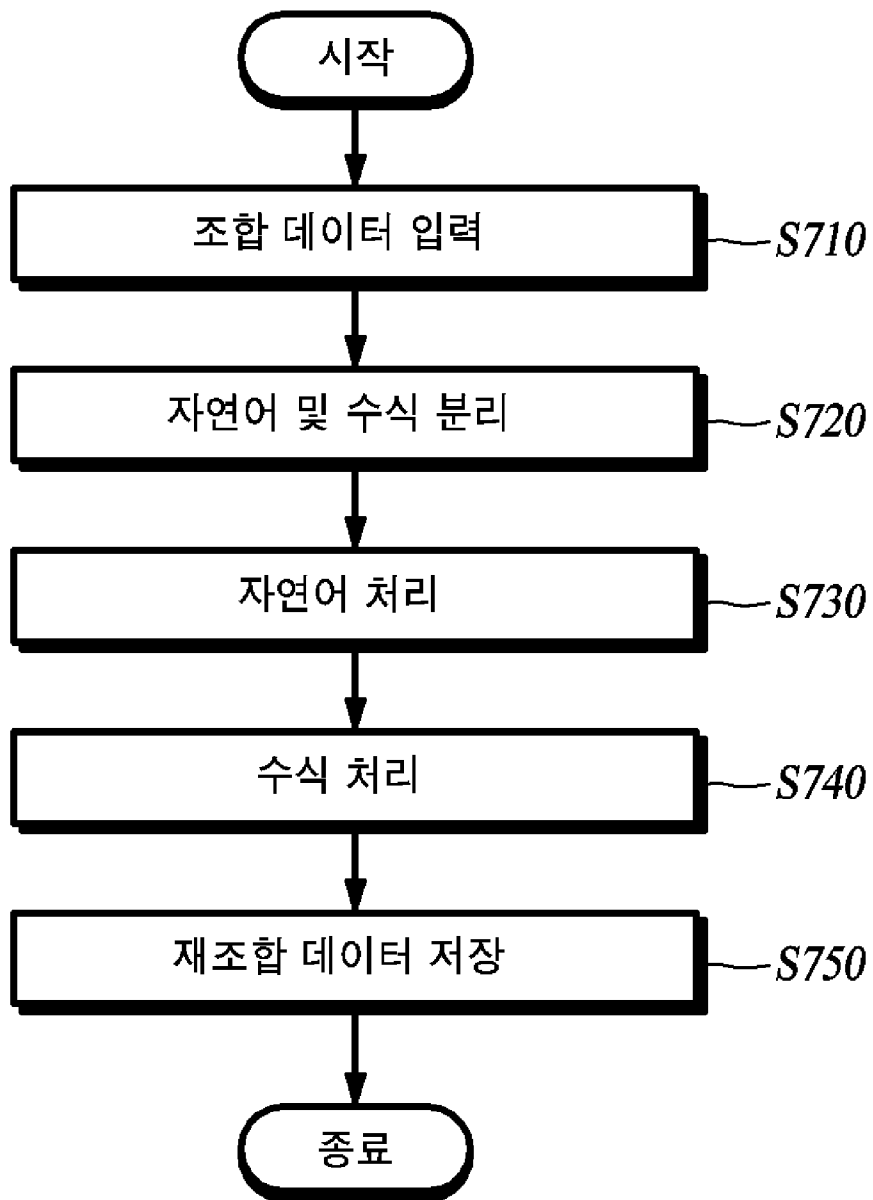
[Fig. 5]



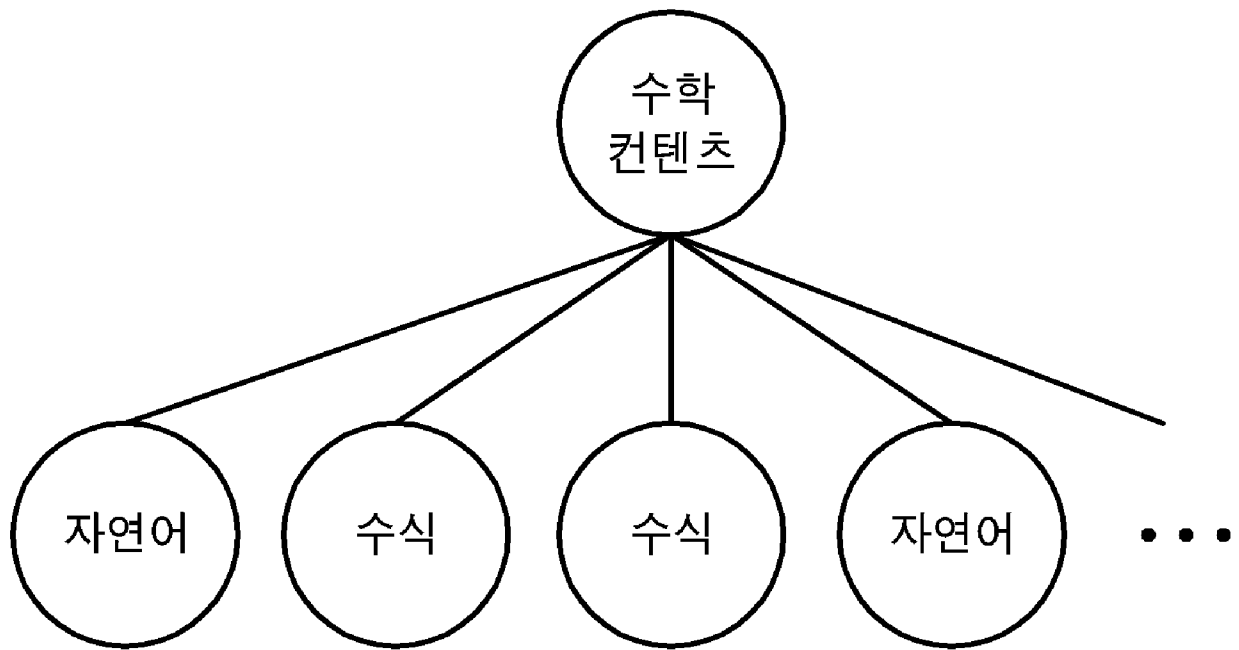
[Fig. 6]



[Fig. 7]



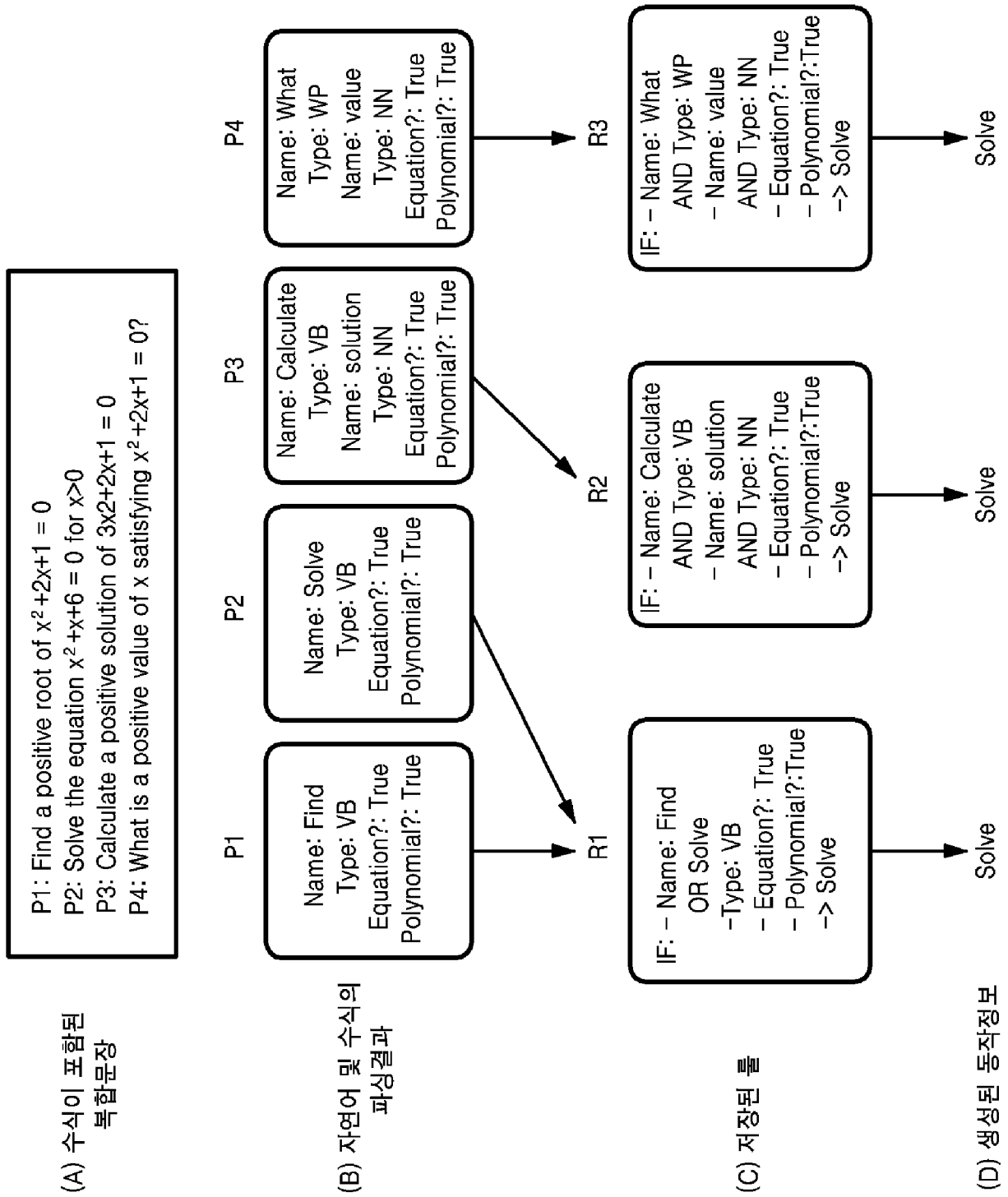
[Fig. 8]



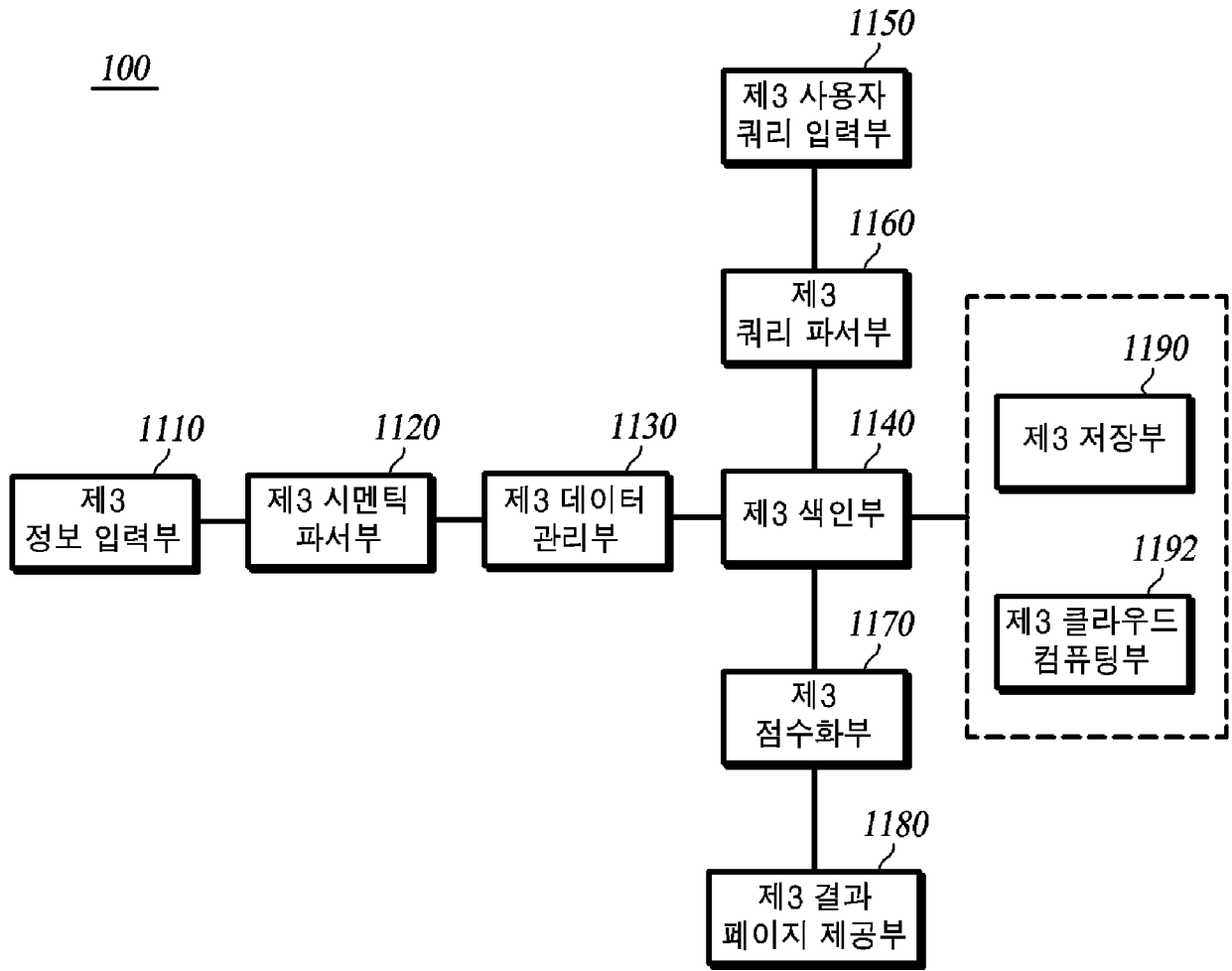
[Fig. 9]



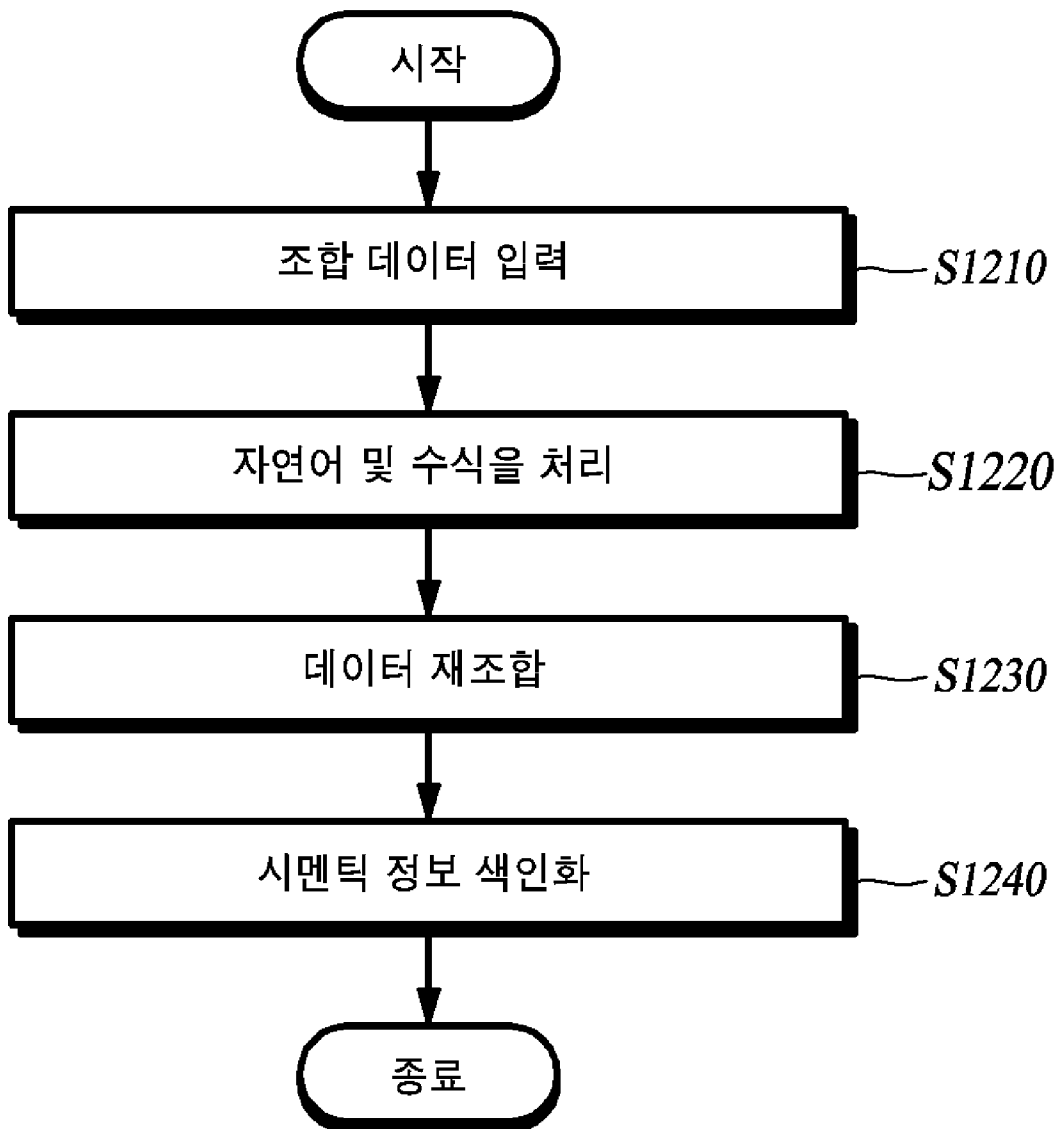
[Fig. 10]



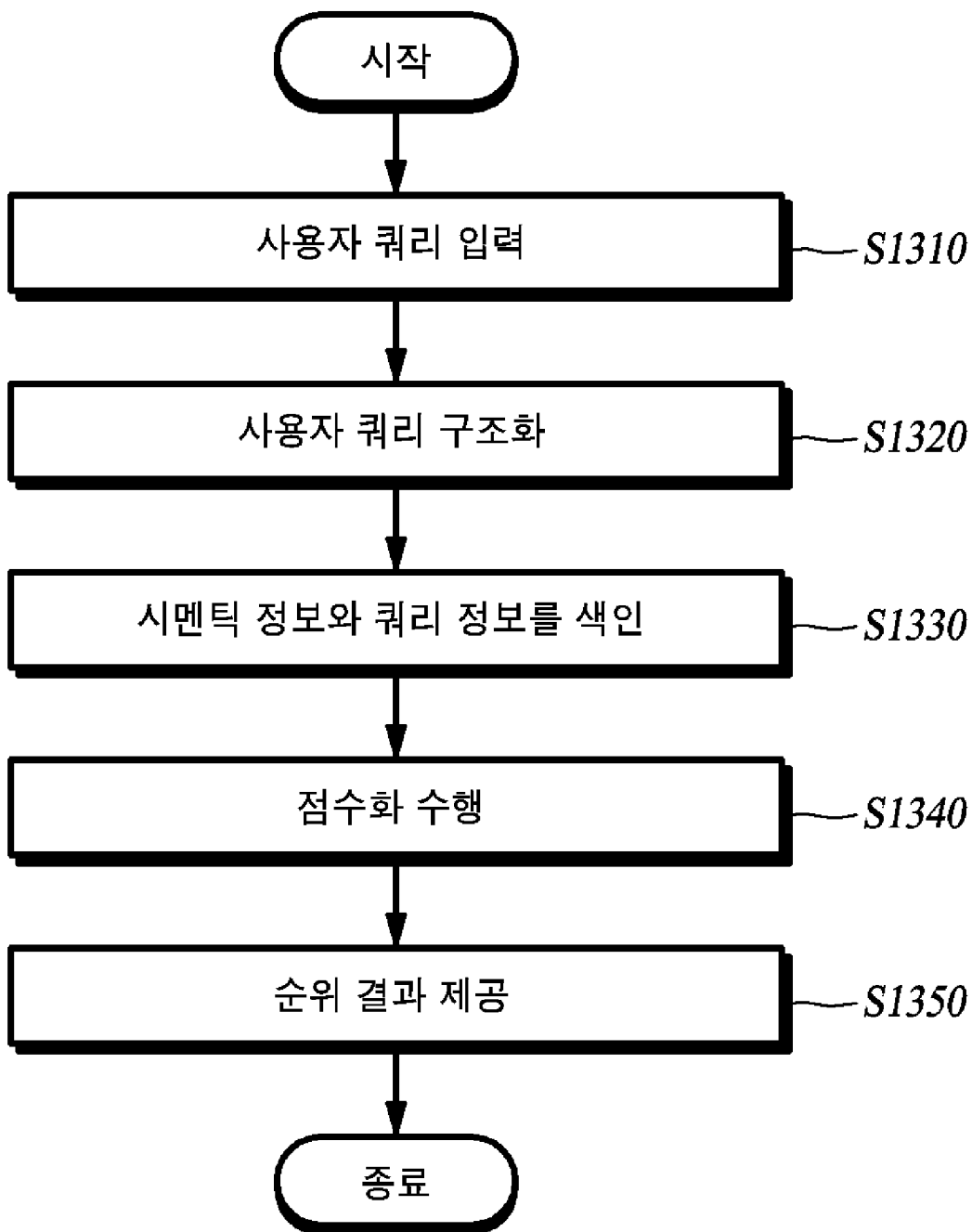
[Fig. 11]



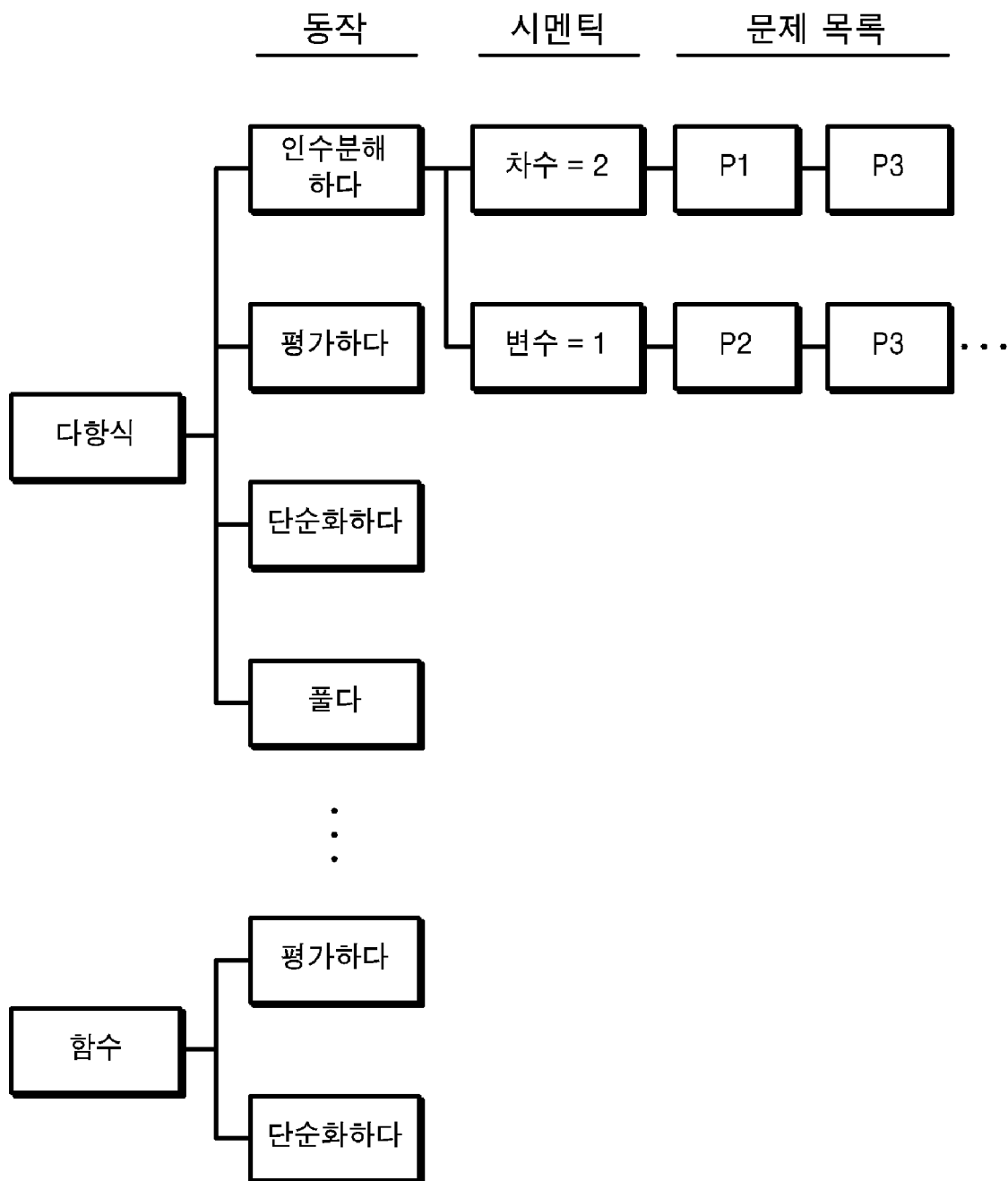
[Fig. 12]



[Fig. 13]



[Fig. 14]



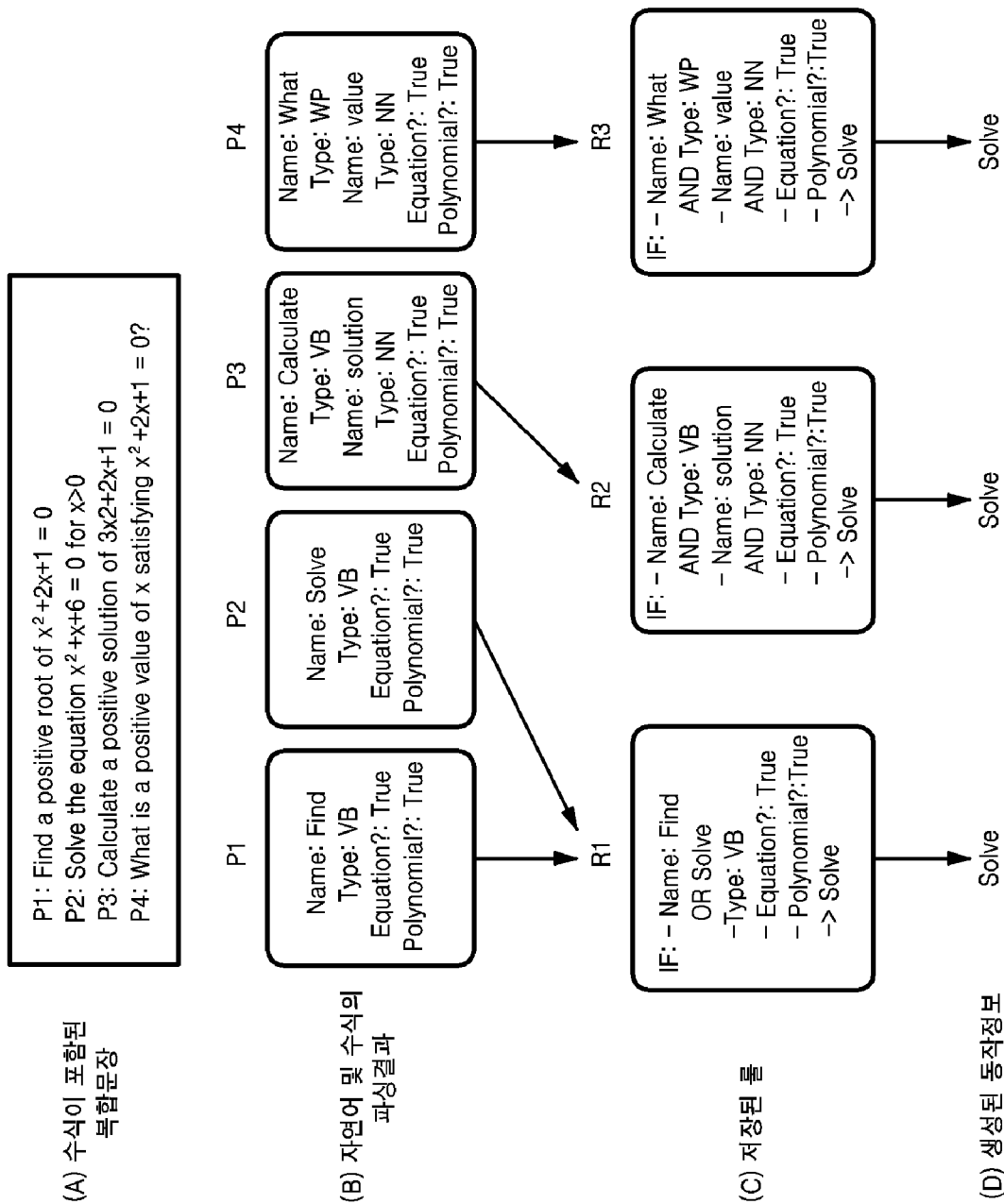
[Fig. 15]

	문제 1	문제 2	문제 3	문제 4	문제 5	문제 6
다항식	1	1	0	0	0	1
함수	1	1	0	1	0	0
인수분해하다	1	1	0	1	1	1
풀다	0	1	0	0	0	0
평가하다	1	0	0	0	0	0
변수 = 1	1	0	1	1	1	1
차수 = 2	1	0	1	1	1	0

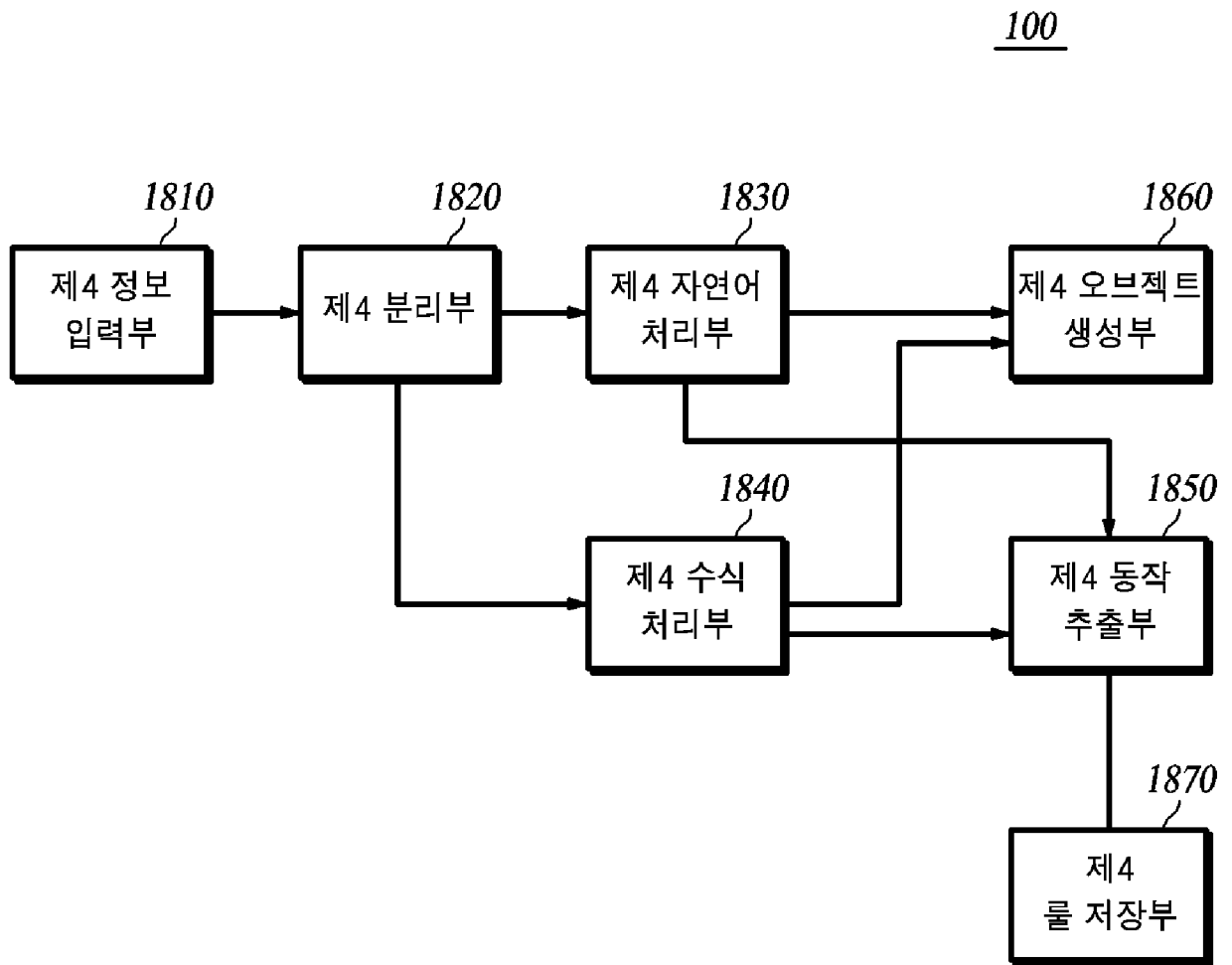
[Fig. 16]



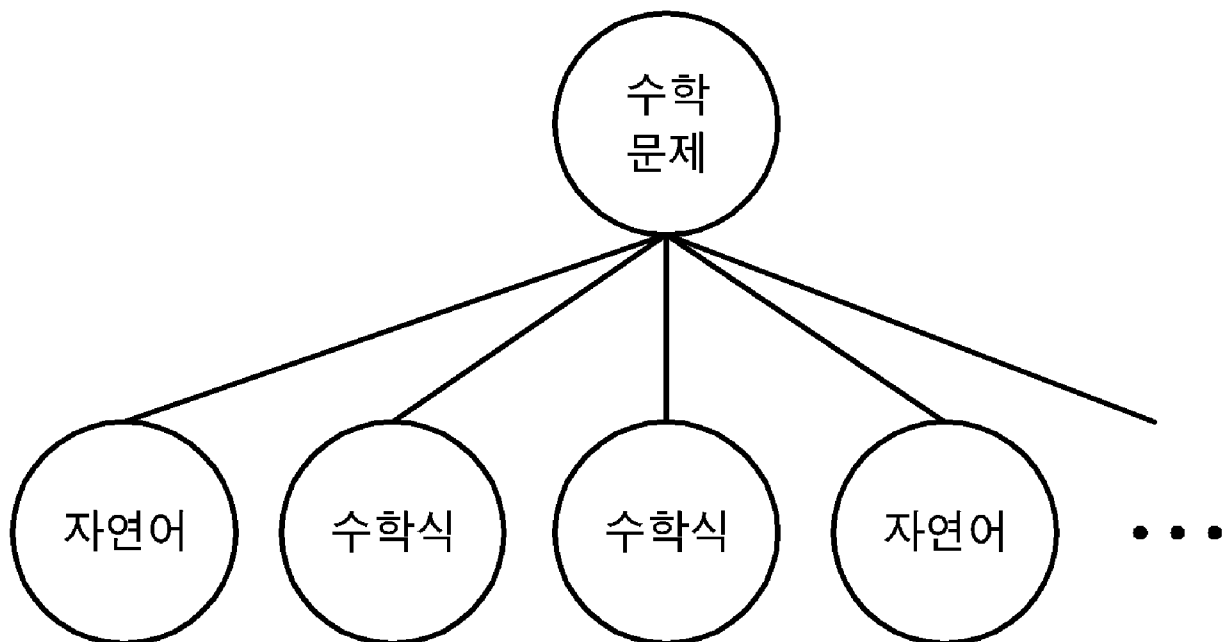
[Fig. 17]



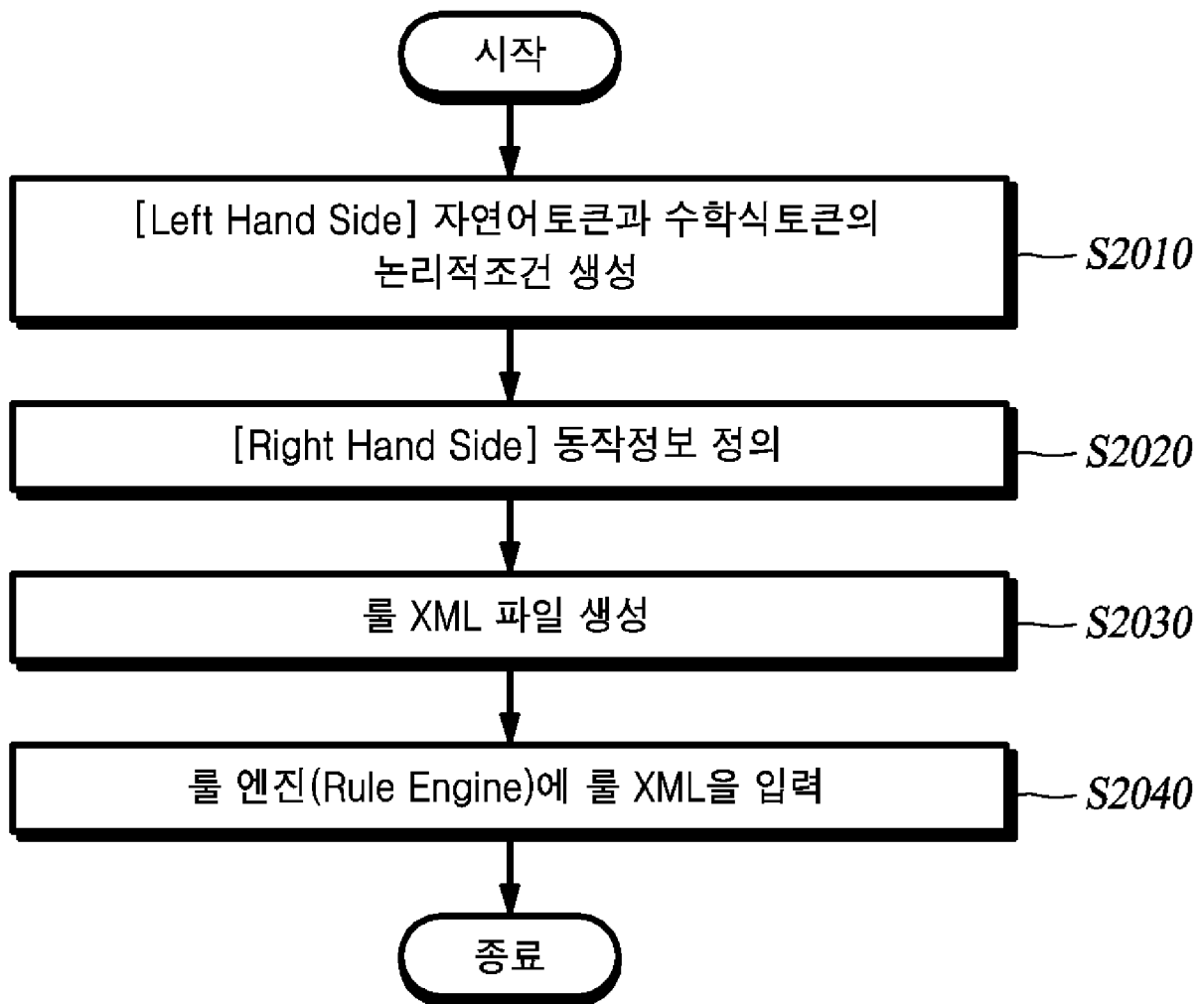
[Fig. 18]



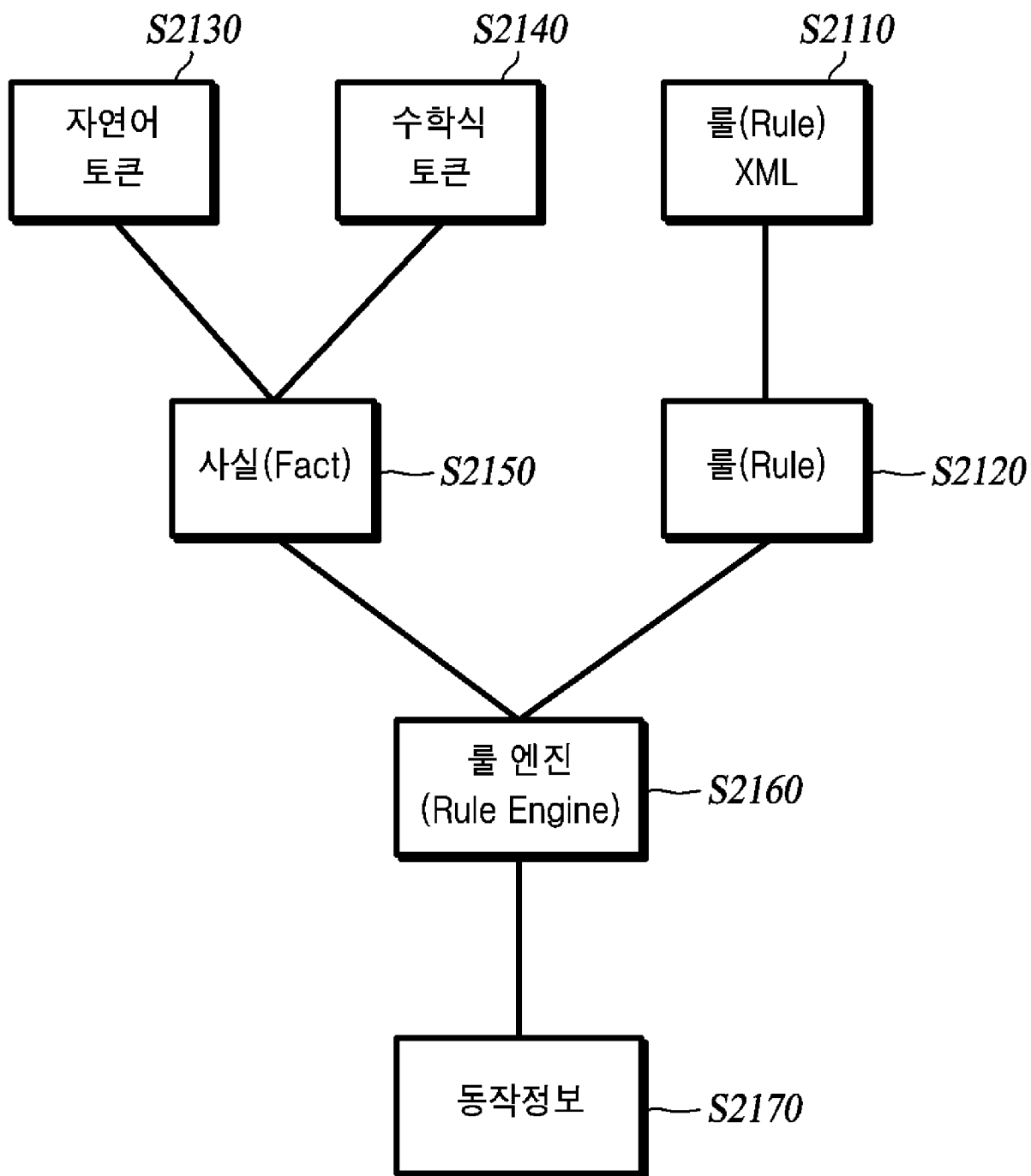
[Fig. 19]



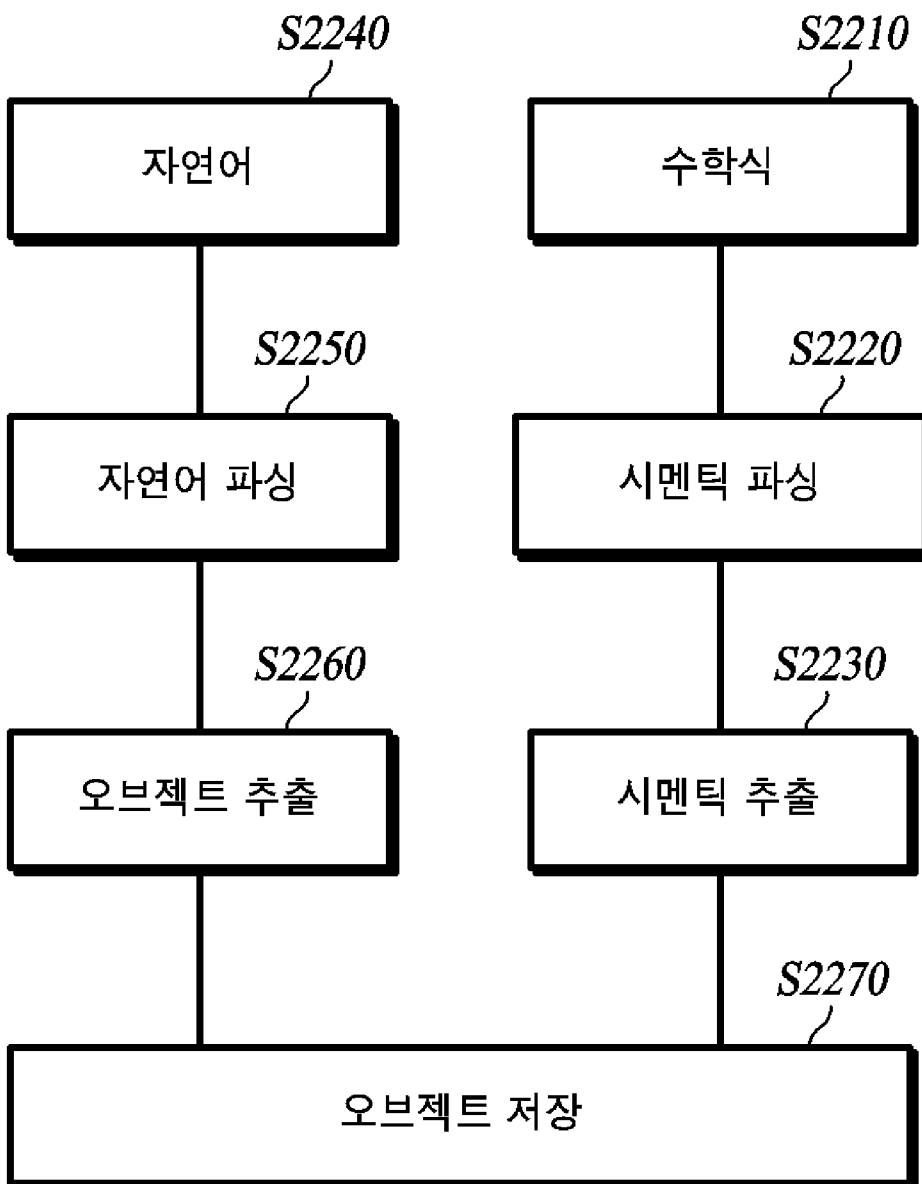
[Fig. 20]



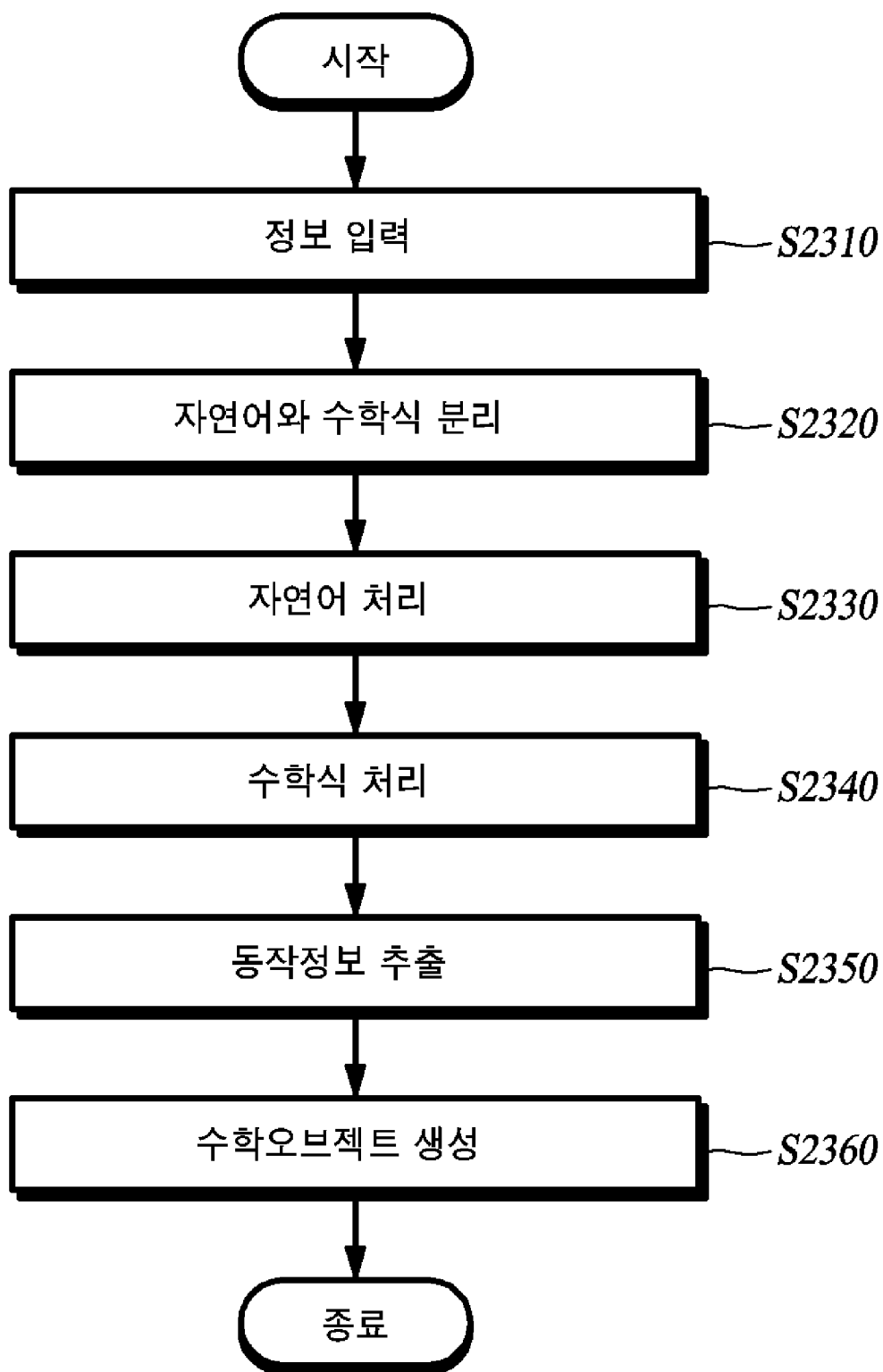
[Fig. 21]



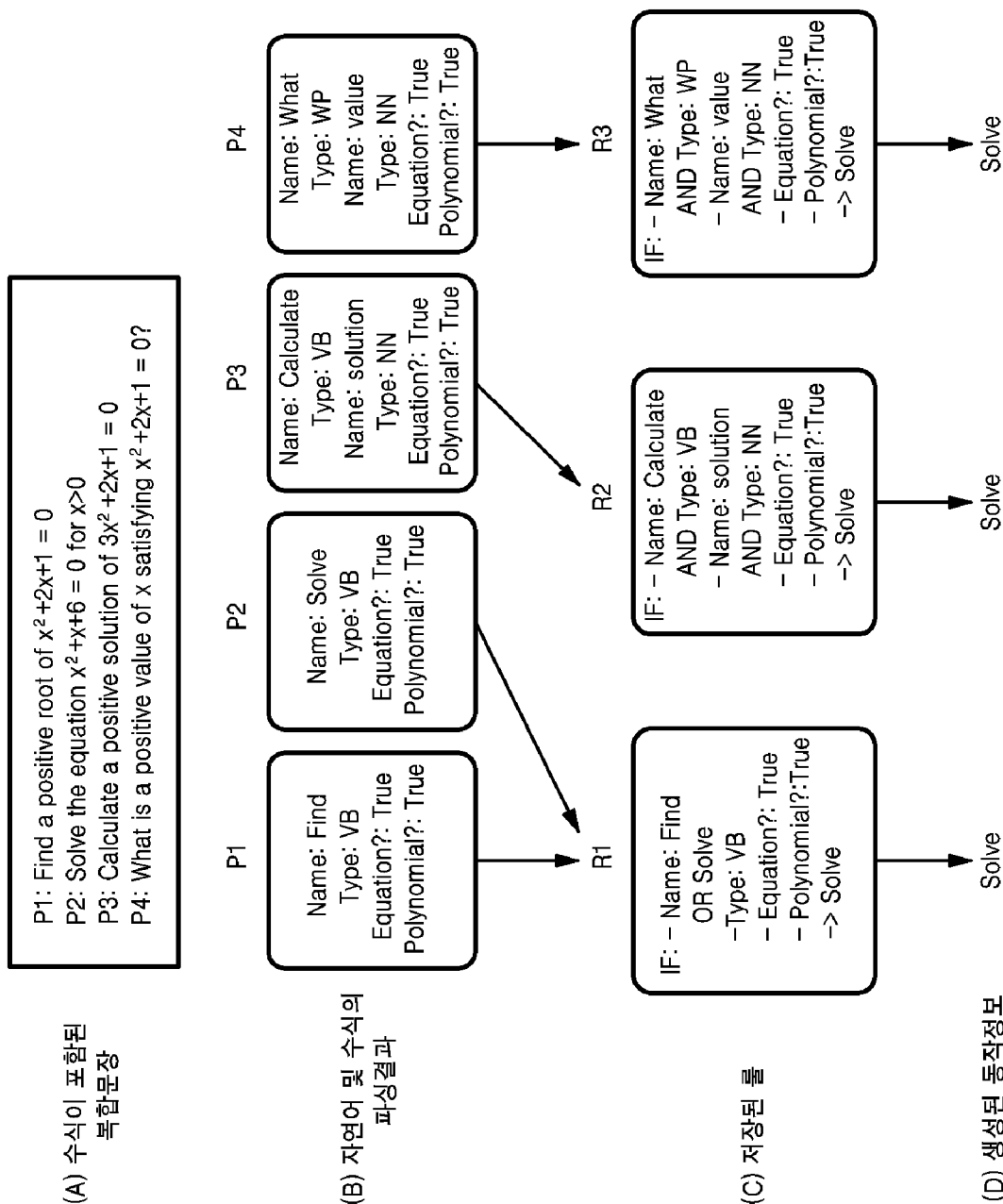
[Fig. 22]



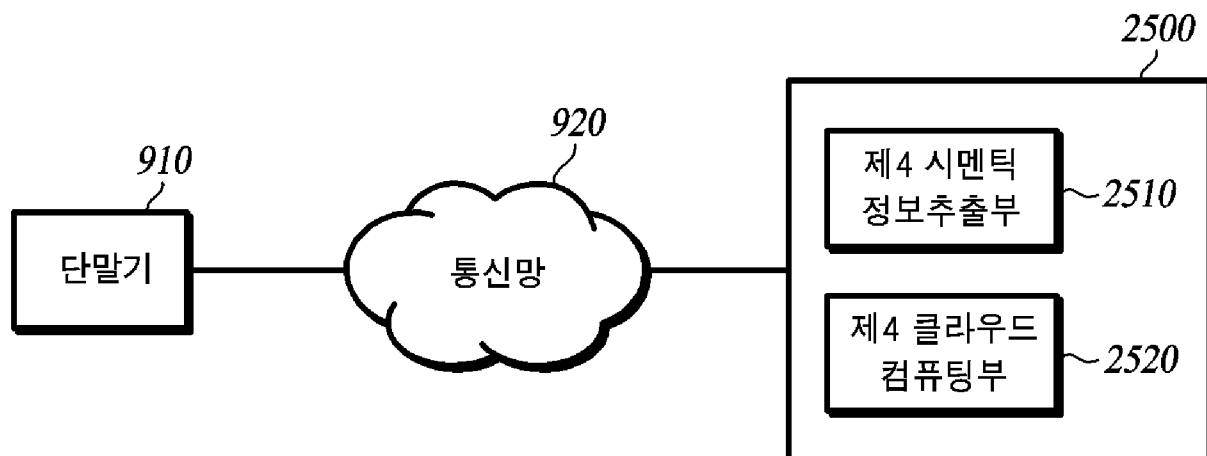
[Fig. 23]



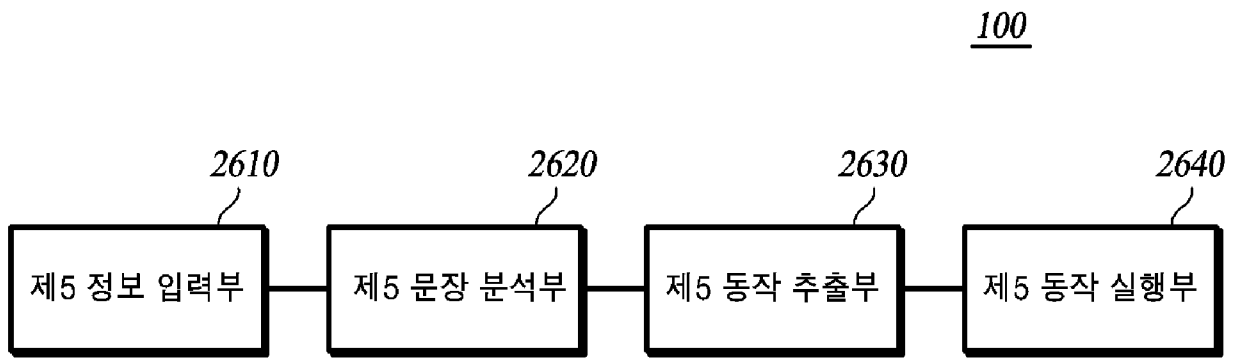
[Fig. 24]



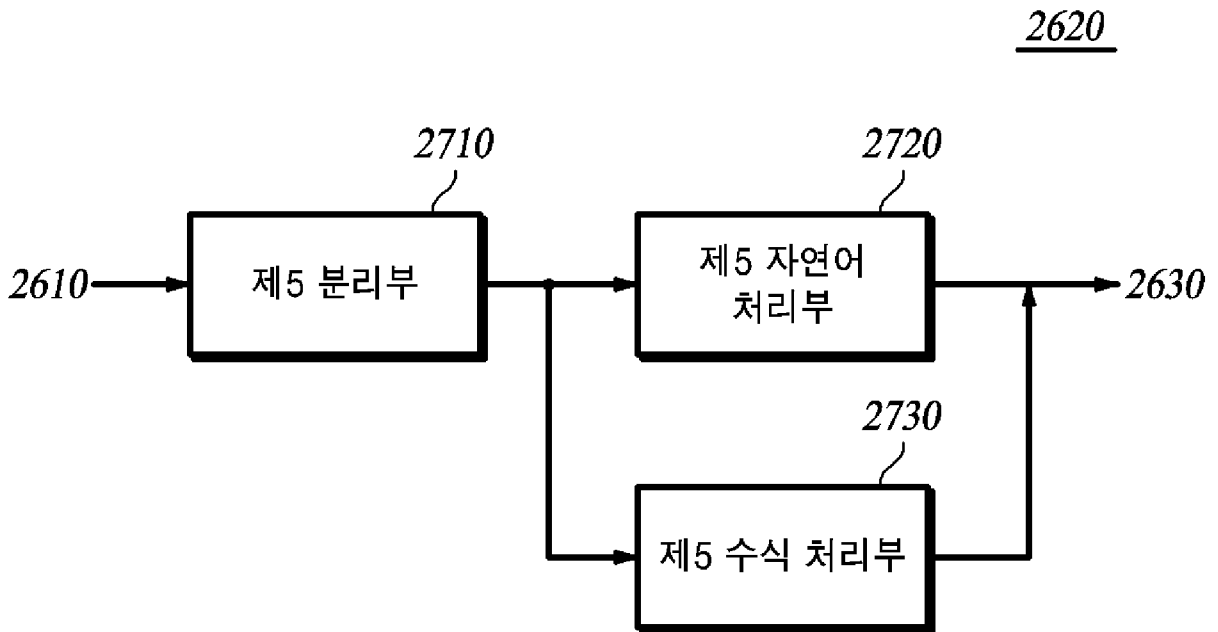
[Fig. 25]



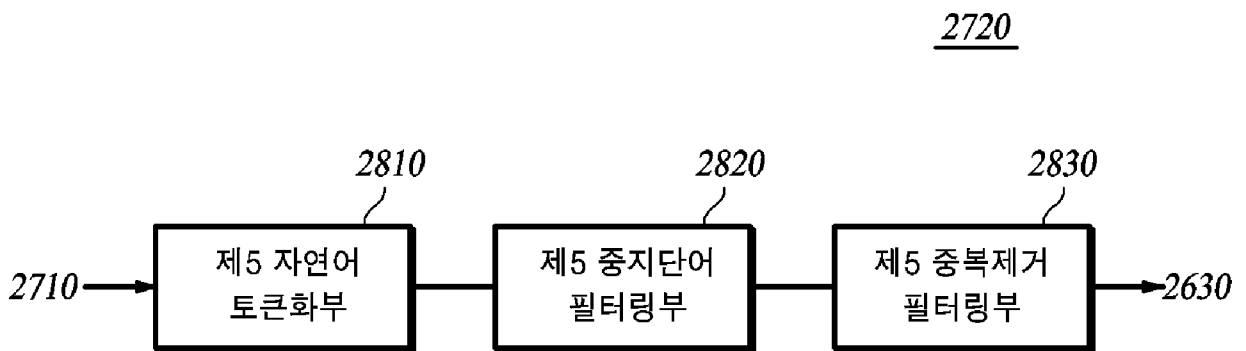
[Fig. 26]



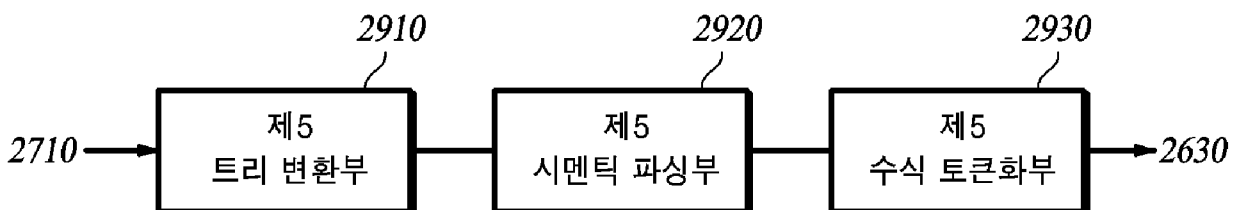
[Fig. 27]



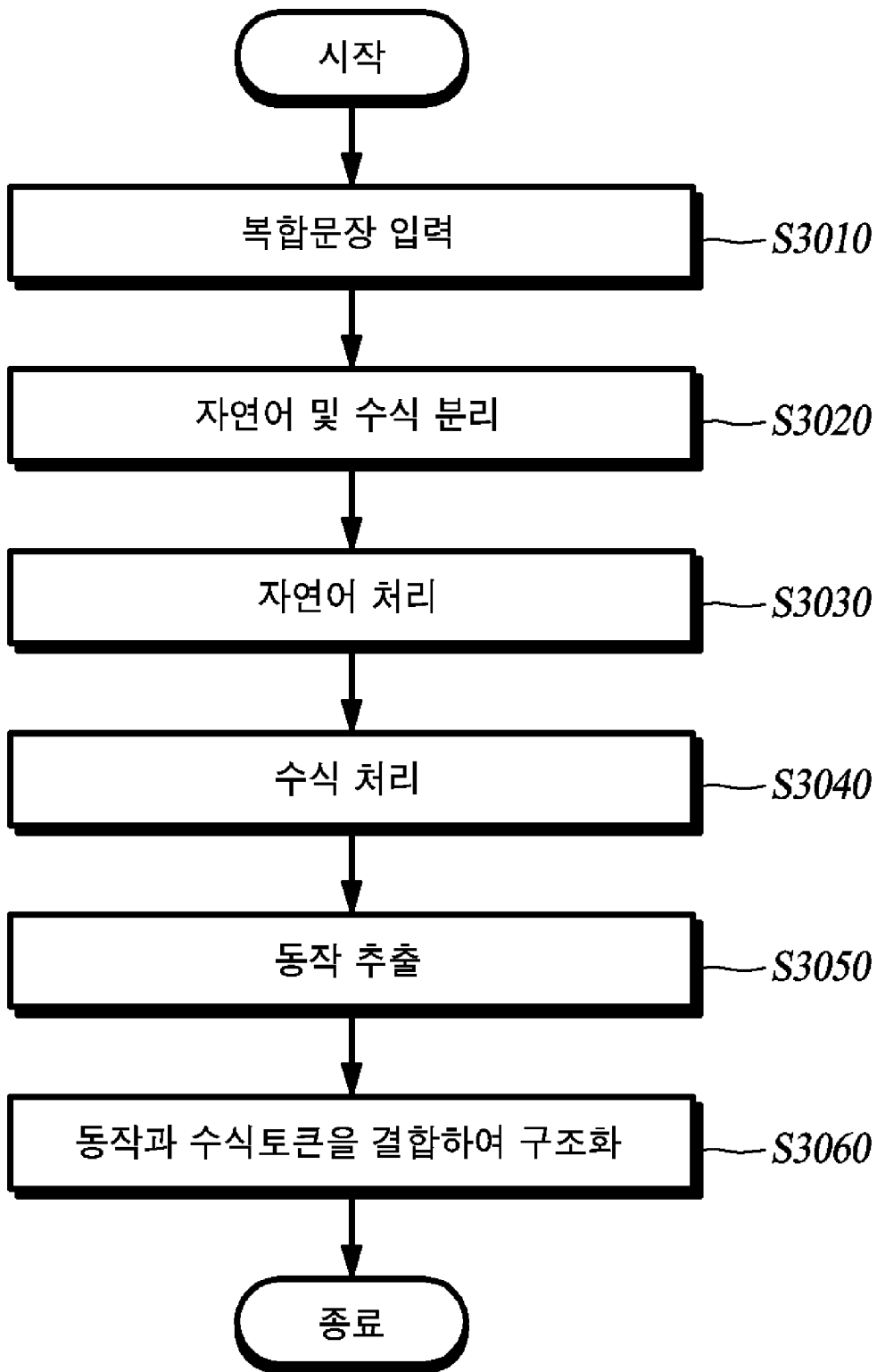
[Fig. 28]



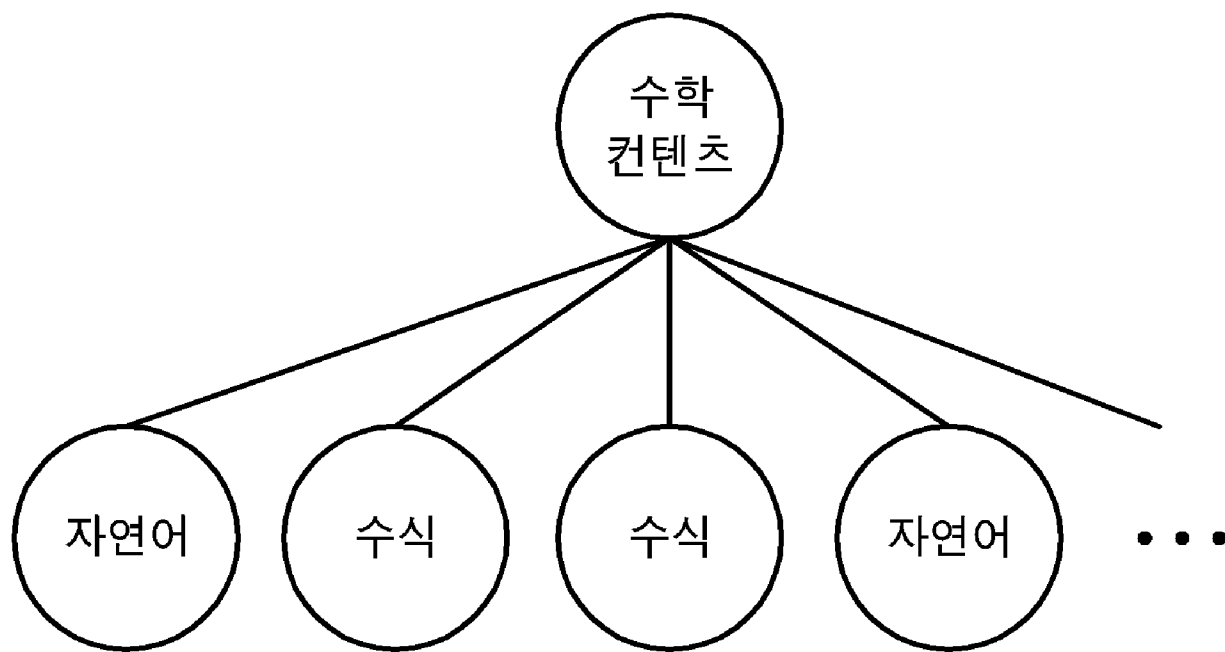
[Fig. 29]



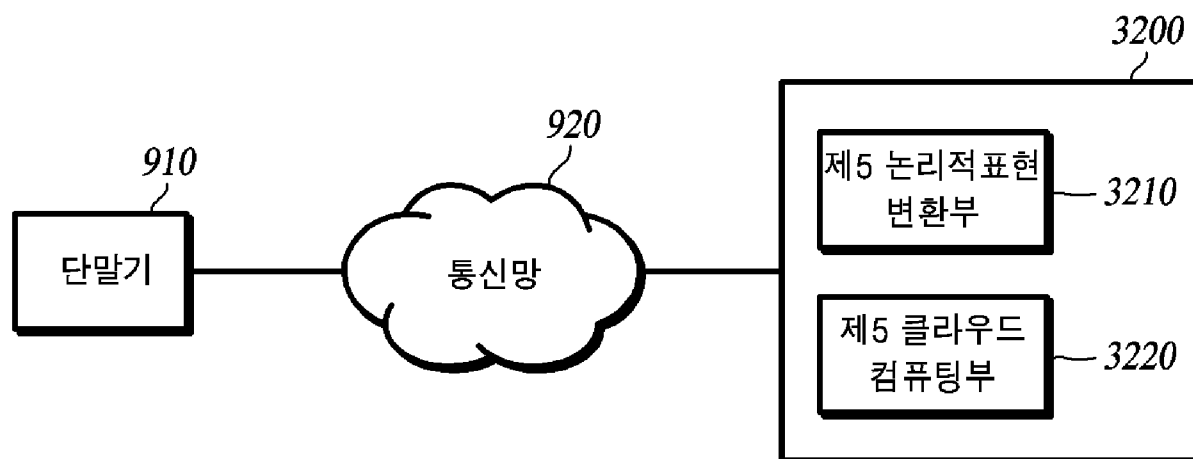
[Fig. 30]



[Fig. 31]

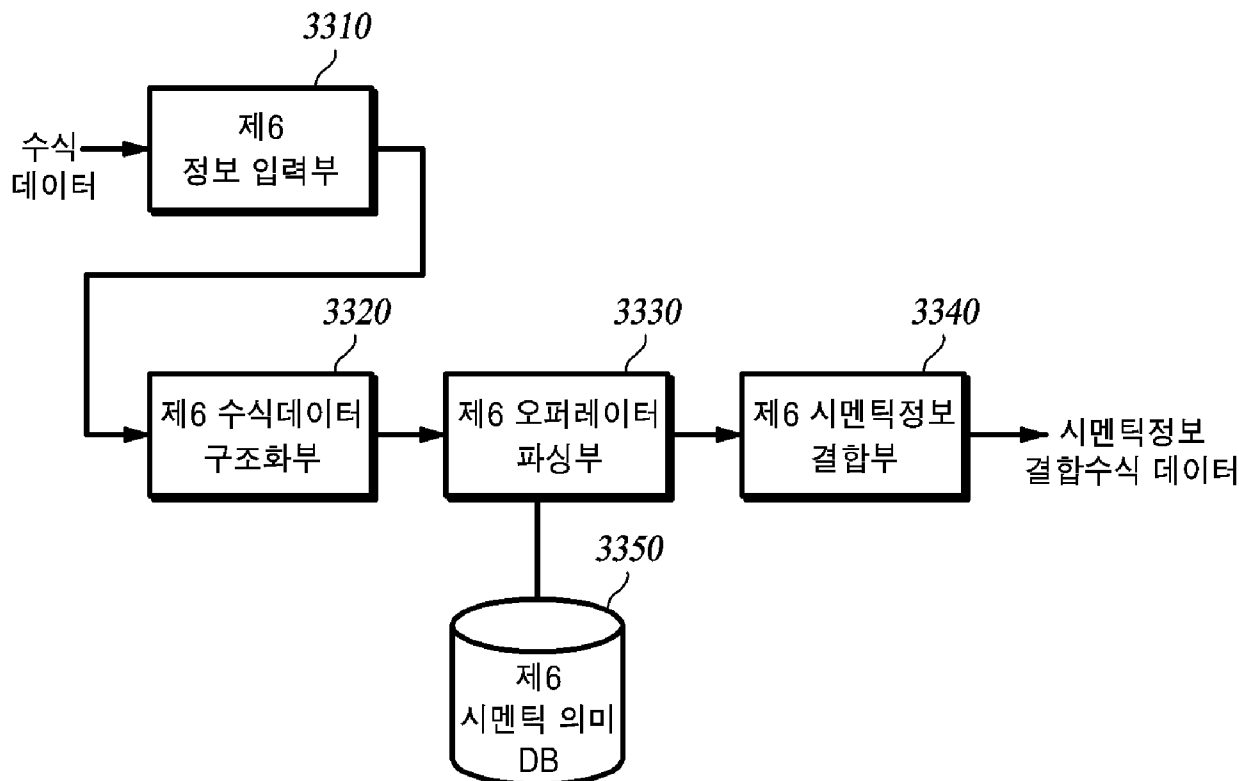


[Fig. 32]

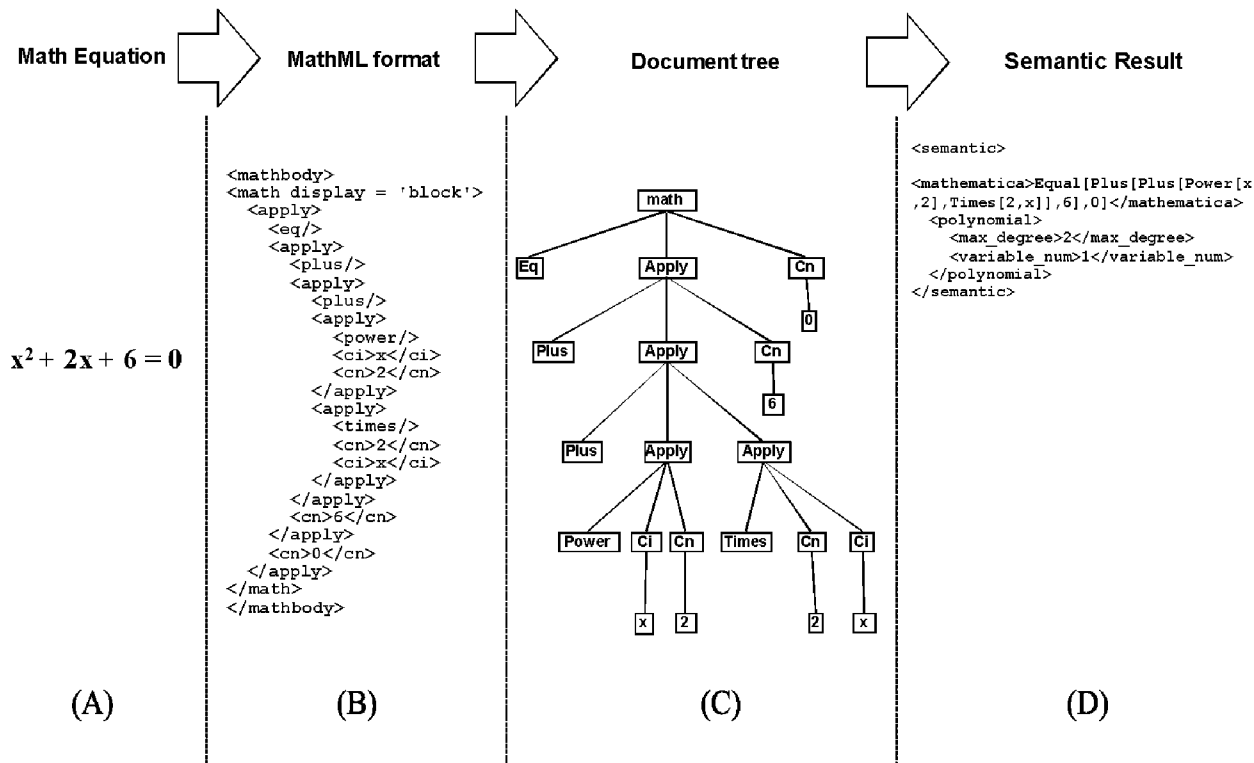


[Fig. 33]

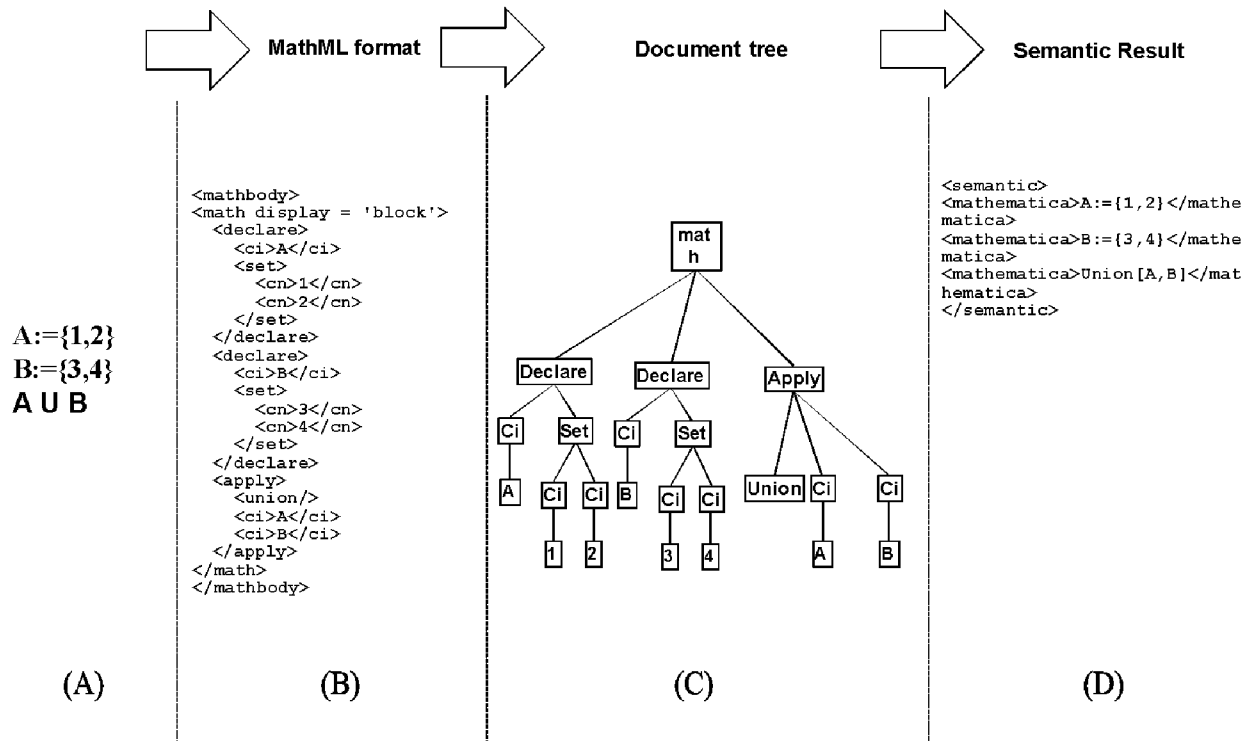
100



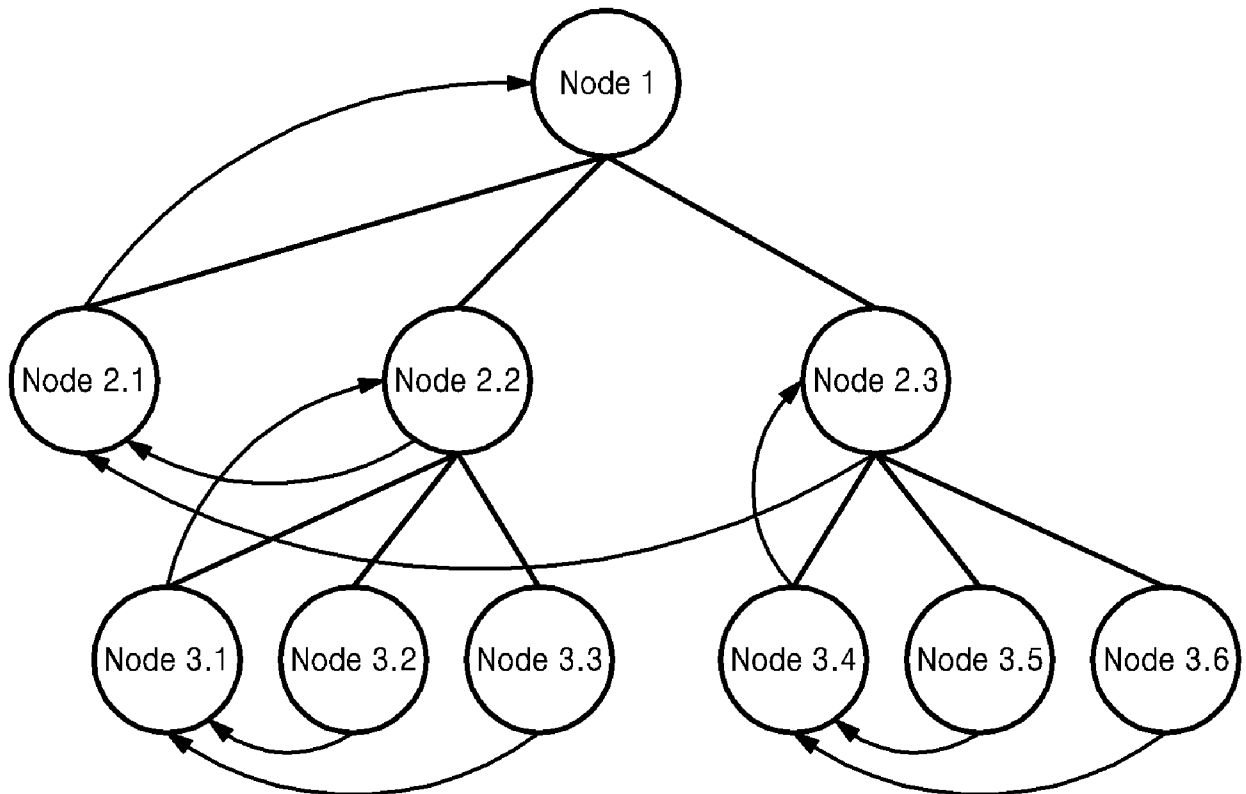
[Fig. 34]



[Fig. 35]



[Fig. 36]



[Fig. 37]

```

<mathbody>
<math display = 'block'>
  <declare>
    <ci>A</ci>
    <set>
      <cn>1</cn>
      <cn>2</cn>
    </set>
  </declare>
  <declare>
    <ci>B</ci>
    <set>
      <cn>3</cn>
      <cn>4</cn>
    </set>
  </declare>
  <apply>
    <union/>
    <ci>A</ci>
    <ci>B</ci>
  </apply>
</math>
<semantic>
<mathematica>A:={1,2}</mathe
matica>
<mathematica>B:={3,4}</mathe
matica>
<mathematica>Union[A,B]</mat
hematica>
</semantic>
</mathbody>

```

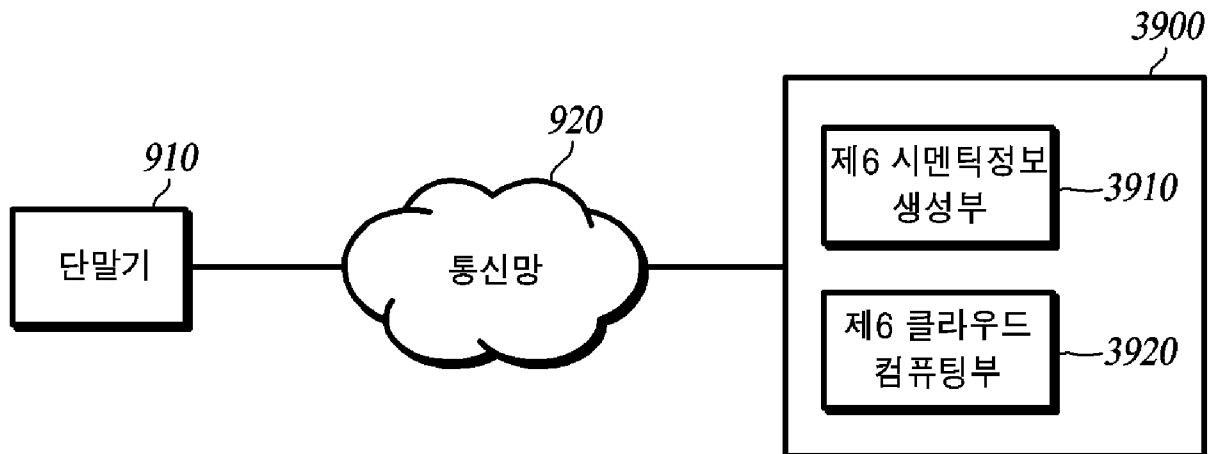
(a)

(b)

[Fig. 38]

계수	변수 세트
계수	(변수 이름, 차수) / 변수 세트

[Fig. 39]



[Fig. 40]

