



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2015년07월29일
 (11) 등록번호 10-1538971
 (24) 등록일자 2015년07월17일

- | | |
|--|--|
| (51) 국제특허분류(Int. Cl.)
G06F 17/21 (2006.01) G06F 17/28 (2006.01)
(21) 출원번호 10-2013-0045751
(22) 출원일자 2013년04월24일
심사청구일자 2013년04월24일
(65) 공개번호 10-2014-0127113
(43) 공개일자 2014년11월03일
(56) 선행기술조사문헌
Conroy, J. M. and O'leary, D. P. Text summarization via hidden markov models”, In Proceedings of SIGIR '01, New York, NY, USA. 2001. pages 406~407. | (73) 특허권자
한국과학기술원
대전광역시 유성구 대학로 291(구성동)
(72) 발명자
서효원
대전광역시 유성구 대학로 291, 한국과학기술원 (구성동)
윤종석
서울특별시 은평구 진흥로 265, 204-1206호 (불광동, 북한산삼성래미안아파트)
(74) 대리인
박영우 |
|--|--|

전체 청구항 수 : 총 10 항

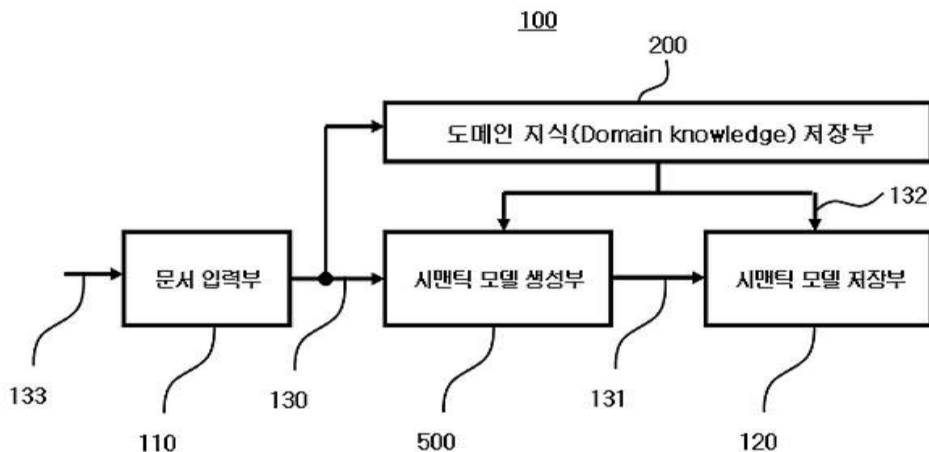
심사관 : 이복현

(54) 발명의 명칭 **온톨로지 자동 번역 장치, 온톨로지 기반 문서 요약 장치, 온톨로지 기반 문서 요약 방법, 온톨로지 기반 문서 추천 장치 및 온톨로지 기반 문서 추천 방법**

(57) 요약

온톨로지 생성 장치는 문서 입력부, 도메인 지식 저장부, 시맨틱 모델 생성부 및 시맨틱 모델 저장부를 포함한다. 문서 입력부는 문서를 입력 받고 입력된 문서를 출력한다. 도메인 지식 저장부에는 입력된 문서와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 저장된다. 시맨틱 모델 생성부는 입력된 문서를 전달 받는다. 시맨틱 모델 생성부는 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 도메인 지식 저장부 내에 미리 선언된 제2 객체인 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 입력된 문서의 시맨틱 모델을 생성한다. 시맨틱 모델 저장부는 입력된 문서의 시맨틱 모델을 저장한다.

대표도 - 도1



명세서

청구범위

청구항 1

삭제

청구항 2

삭제

청구항 3

삭제

청구항 4

삭제

청구항 5

시맨틱 모델을 저장하는 시맨틱 모델 저장부;

상기 시맨틱 모델과 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부; 및

상기 시맨틱 모델 저장부에 저장된 적어도 하나의 시맨틱 모델을 전달받아 번역 문서를 생성하는 시맨틱 모델 번역부를 포함하고,

상기 시맨틱 모델 저장부는

상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스; 및

상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 시맨틱 모델을 저장하고,

상기 시맨틱 모델 번역부는

상기 도메인 지식 저장부에 포함된 규칙을 이용하여 상기 번역 문서의 내용을 생성하는 규칙 기반 번역부; 및

상기 적어도 하나의 시맨틱 모델 간의 유사도(Similarity)를 이용하여 상기 번역 문서의 문단을 생성하는 시맨틱 모델 집합화 처리부를 포함하는 온톨로지 자동 번역 장치.

청구항 6

삭제

청구항 7

삭제

청구항 8

제5 항에 있어서,

상기 규칙 기반 번역부는 상기 적어도 하나의 시맨틱 모델 내부 관계들 간에 중복된 컨셉이 존재하는 경우 축약된 번역 문서를 생성하는 것을 특징으로 하는 온톨로지 자동 번역 장치.

청구항 9

제5 항에 있어서,

상기 규칙 기반 번역부는 상기 적어도 하나의 시맨틱 모델 내부에 적어도 둘의 동일한 구조를 가지는 관계들이 존재하는 경우 상기 동일한 구조를 가지는 관계들 중 하나의 관계의 번역만을 생성하는 것을 특징으로 하는 온톨로지 자동 번역 장치.

청구항 10

문서를 입력 받고 입력된 문서를 출력하는 문서 입력부;
 상기 입력된 문서와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부;
 상기 입력된 문서를 전달받아 적어도 하나의 상기 입력된 문서의 시맨틱 모델을 생성하는 시맨틱 모델 생성부;
 상기 입력된 문서의 시맨틱 모델을 저장하는 시맨틱 모델 저장부; 및
 상기 시맨틱 모델 저장부에 저장된 상기 입력된 문서의 시맨틱 모델을 전달받아 번역 문서를 생성하는 시맨틱 모델 번역부를 포함하고,
 상기 시맨틱 모델 생성부는
 상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스; 및
 상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 입력된 문서의 시맨틱 모델을 생성하고,
 상기 시맨틱 모델 번역부는
 상기 도메인 지식 저장부에 포함된 규칙을 이용하여 상기 번역 문서의 내용을 생성하는 규칙 기반 번역부; 및
 상기 입력된 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 번역 문서의 문단을 생성하는 시맨틱 모델 집합화 처리부를 포함하는 온톨로지 기반 문서 요약 장치.

청구항 11

삭제

청구항 12

삭제

청구항 13

제10 항에 있어서,
 상기 규칙 기반 번역부는 상기 입력된 문서의 시맨틱 모델 내부 관계들 간에 중복된 컨셉이 존재하는 경우 축약된 번역 문서를 생성하는 것을 특징으로 하는 온톨로지 기반 문서 요약 장치.

청구항 14

제10 항에 있어서,
 상기 규칙 기반 번역부는 상기 입력된 문서의 시맨틱 모델 내부에 적어도 둘의 동일한 구조를 가지는 관계들이 존재하는 경우 상기 동일한 구조를 가지는 관계들 중 하나의 관계의 번역만을 생성하는 것을 특징으로 하는 온톨로지 기반 문서 요약 장치.

청구항 15

문서를 입력받는 단계;
 상기 문서와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부 내에 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 문서의 시맨틱 모델을 생성하는 단계;
 상기 문서의 시맨틱 모델을 저장하는 단계;

상기 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 문서의 시맨틱 모델의 번역 문서의 문단을 생성하는 시맨틱 모델 집합화하는 단계;

상기 도메인 지식 저장부에 포함된 규칙에 기반하여 상기 번역 문서의 문단 내용을 생성하는 단계를 포함하는 온톨로지 기반 문서 요약 방법.

청구항 16

문의(Query)를 입력 받고 입력된 문의를 출력하는 문의 입력부;

상기 입력된 문의와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부;

상기 입력된 문의를 전달받아 상기 입력된 문의를 시맨틱 모델을 생성하는 시맨틱 모델 생성부;

추천 대상 문서들의 시맨틱 모델이 저장된 시맨틱 모델 저장부;

상기 입력된 문의를 시맨틱 모델과 유사도가 높은 순서로 적어도 하나의 시맨틱 모델을 상기 시맨틱 모델 저장부에서 추천 문서의 시맨틱 모델로 검색하는 유사 문서 검색부; 및

상기 추천 문서의 시맨틱 모델을 전달받아 번역 문서를 생성하는 시맨틱 모델 번역부를 포함하고

상기 시맨틱 모델 생성부는

상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스; 및

상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 입력된 문의를 시맨틱 모델을 생성하고,

상기 시맨틱 모델 번역부는

상기 도메인 지식 저장부에 포함된 규칙을 이용하여 상기 번역 문서의 내용을 생성하는 규칙 기반 번역부; 및

상기 추천 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 번역 문서의 문단을 생성하는 시맨틱 모델 집합화 처리부를 포함하는 온톨로지 기반 문서 추천 장치.

청구항 17

제16 항에 있어서,

상기 도메인 지식 저장부를 이용하여 상기 입력된 문의를 시맨틱 모델을 확장하는 문의 확장부를 더 포함하는 것을 특징으로 하는 온톨로지 기반 문서 추천 장치.

청구항 18

삭제

청구항 19

문의를 입력받는 단계;

상기 문의와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부 내에 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 상기 문의를 시맨틱 모델을 생성하는 단계;

상기 문의를 시맨틱 모델을 저장하는 단계;

상기 도메인 지식 저장부를 이용하여 상기 문의를 시맨틱 모델의 확장된 시맨틱 모델을 생성하는 단계;

상기 확장된 시맨틱 모델과 유사도가 높은 순서로 적어도 하나의 시맨틱 모델을 시맨틱 모델 저장부에서 추천 문서의 시맨틱 모델로 검색하는 단계;

상기 추천 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 추천 문서의 시맨틱 모델의 번역 문서의 문단을 생성하는 시맨틱 모델 집합화하는 단계; 및

상기 도메인 지식 저장부에 포함된 규칙에 기반하여 상기 추천 문서의 시맨틱 모델의 번역 문서의 내용을 생성

하는 단계를 포함하는 온톨로지 기반 문서 추천 방법.

발명의 설명

기술 분야

[0001] 본 발명은 온톨로지 기반 문서의 처리에 관한 것으로서, 더욱 상세하게는 문서의 온톨로지 자동 생성 장치, 온톨로지 자동 번역 장치, 온톨로지 기반 문서 요약 장치, 온톨로지 기반 문서의 요약 방법, 온톨로지 기반 문서의 추천 장치 및 온톨로지 기반 문서의 추천 방법에 관한 것이다.

배경 기술

[0002] 회사의 시스템 내 제품의 제품 기획, 제품 개발, 제품 제조, 마케팅, 애프터 서비스 과정 등 대부분의 과정에서 다수의 문서들이 쓰여지고 있다. 상기 과정의 구성원의 문서 작성 방법이나 경향에 따라 같은 의미의 문장이라도 다른 방식으로 표현된 문서는 타 과정의 구성원이 이해하기에는 어려움이 존재하는 바, 상기 과정의 구성원이 신속하고 적절한 결정(Decision making)을 하기 위해서는 상기 다수의 문서들은 이해하기 쉽도록 처리되어 있어야 한다. 그러나 의미의 혼동과 문법의 다양성으로 인해 정형화된 문서 작성 방법을 제안하기는 현실적으로 어렵다.

[0003] 상기 과정의 구성원이 타 과정의 구성원의 문서를 빠른 시간 내에 직관적으로 이해하기 위해서는 모든 형태의 작성된 문서에 적용될 수 있는 정형화된 문법적(Syntactic) 및 의미적(Semantic) 검토 과정을 거친 문서의 요약 방법과 문서의 추천 방법이 필요하다.

[0004] 문서의 요약 방법은 추출적 방법(Extractive method)과 추상적 방법(Abstractive method)으로 분류된다. 상기 추출적 방법은 문서 내의 키워드 및 위치 정보를 이용해서 단어, 절, 문장을 추출한다. 상기 추출적 방법은 구현에 어려움이 적은 장점을 가지나, 요약의 정도(압축률)가 낮고, 모순된 내용을 추출할 수 있는 바, 수사적 구조(Rhetorical structure)에 의한 가이드 라인(Guide line)을 제공받거나 검토 전략(Revision strategy)에 의한 수정이 필요한 단점을 가진다. 추상적 방법의 결과는 입력 받은 문서에 존재하는 내용이 아닌, 유의어 사전(Thesaurus) 또는 자료집(Corpus)과 같은 지식 기반의 배경 지식을 이용하여 상기 문서를 요약한다. 상기 추상적인 방법은 일반화의 정도가 높고, 높은 요약의 정도(압축률)를 가지며, 의미를 중심으로 하여 문서량을 감소시키는 장점을 가지나, 상기 문서의 도메인(Domain)에 한정된 요약만을 제공하며, 일반화 정도의 추출 오류로 인해 독자(Reader)를 잘못 파악할 수 있는 단점을 가진다.

[0005] 온톨로지는 구조화 되어 있지 않은 문서를 분석하고, 분석된 정보를 구조화 하는 최적의 포맷(Format)인 바, 온톨로지 기반 문서의 요약 및 추천 방법은 기존의 문서 요약 및 추천 방법의 문제를 해결 할 수 있다.

선행기술문헌

[1] Svore, K., Vanderwende, L., and Burges, C. (2007). "Enhancing single-document summarization by combining RankNet and third-party sources", In Proceedings of the EMNLP-CoNLL, pages 448~457. [7, 8]

[2] Conroy, J. M. and O'leary, D. P. (2001). "Text summarization via hidden markov models", In Proceedings of SIGIR '01, pages 406~407, New York, NY, USA.

발명의 내용

해결하려는 과제

[0006] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 목적은 문서의 다양한 구조에 불문하여 문서의 온톨로지를 자동으로 생성하는 장치를 제공하는 것이다.

[0007] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 목적은 온톨로지를 자동으로 자연어로 번역하는 장치를 제공하는 것이다.

[0008] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 목적은 온톨로지를 기반으로 자동으로 문서를 요약하는 장치

를 제공하는 것이다.

- [0009] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 목적은 온톨로지를 기반으로 자동으로 문서를 요약하는 방법을 제공하는 것이다.
- [0010] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 목적은 입력된 문의에 대하여 온톨로지를 기반으로 자동으로 유사 문서를 추천하는 장치를 제공하는 것이다.
- [0011] 상기와 같은 문제점을 해결하기 위한 본 발명의 일 목적은 입력된 문의에 대하여 온톨로지를 기반으로 하여 자동으로 유사 문서를 추천하는 방법을 제공하는 것이다.

과제의 해결 수단

- [0012] 상기 본 발명의 일 목적을 달성하기 위한 본 발명의 일 실시예에 따른 온톨로지(Ontology) 생성 장치는 문서 입력부, 도메인 지식 저장부, 시맨틱 모델 생성부 및 시맨틱 모델 저장부를 포함한다. 상기 문서 입력부는 문서를 입력 받고 입력된 문서를 출력한다. 상기 도메인 지식 저장부에는 상기 입력된 문서와 관련된 도메인 지식이 온톨로지 개발 101(Ontology development 101) 방법에 의해 저장된다. 상기 시맨틱 모델 생성부는 상기 입력된 문서를 전달 받는다. 상기 시맨틱 모델 생성부는 상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉(Concept)의 적어도 하나의 제1 인스턴스(Instance) 및 상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계(Relationship)의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 입력된 문서의 시맨틱 모델을 생성한다. 상기 시맨틱 모델 저장부는 상기 입력된 문서의 시맨틱 모델을 저장한다.
- [0013] 일 실시예에서, 상기 도메인 지식 저장부는 동의어 및 유의어 데이터 베이스(Lexicon database(DB)), 참조 모델 DB(Reference model DB) 및 규칙 DB(Rules DB)를 포함한다.
- [0014] 일 실시예에서, 상기 시맨틱 모델 생성부는 문법 분석부(Syntax analysis part) 및 의미 분석부(Semantic analysis part)를 포함한다. 상기 문법 분석부는 상기 입력된 문서를 토큰(token)으로 분할하는 토큰 생성부, 상기 토큰 내 단어의 축약어를 제거하는 단어 매칭부(Word matching), 상기 토큰에 분석한 품사를 태깅(tagging)하는 품사 분석부 및 명사절을 탐지하는 절(Phrase) 탐지부를 포함한다. 상기 의미 분석부는 상기 토큰의 컨셉을 결정하는 컨셉 인식부(Concept recognizing) 및 상기 컨셉 간 관계를 분석하는 컨셉 접합부(Joining)를 포함한다.
- [0015] 일 실시예에서, 상기 컨셉 인식부는 상기 토큰에 대한 컨셉이 복수 개 존재하는 경우 복수 개의 컨셉 중 T-스코어(T-score)가 가장 높은 컨셉을 상기 토큰의 컨셉으로 결정한다.
- [0016] 상기 본 발명의 일 목적을 달성하기 위한 본 발명의 일 실시예에 따른 온톨로지 자동 번역 장치는 도메인 지식 저장부, 시맨틱 모델 번역부 및 시맨틱 모델 저장부를 포함한다. 상기 도메인 지식 저장부에는 상기 시맨틱 모델과 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장된다. 상기 시맨틱 모델 번역부는 상기 시맨틱 모델 저장부에 저장된 적어도 하나의 시맨틱 모델을 전달 받아 번역 문서를 생성한다. 상기 시맨틱 모델 저장부는 상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 시맨틱 모델을 저장한다.
- [0017] 일 실시예에서, 상기 시맨틱 모델 번역부는 상기 도메인 지식 저장부에 포함된 규칙을 이용하여 상기 번역 문서의 내용을 생성하는 규칙 기반 번역부를 포함한다.
- [0018] 일 실시예에서, 상기 시맨틱 모델 번역부는 선택적으로 상기 적어도 하나의 시맨틱 모델 간의 유사도(Similarity)를 이용하여 상기 번역 문서의 문단을 생성하는 시맨틱 모델 집합화 처리부를 더 포함한다.
- [0019] 일 실시예에서, 상기 규칙 기반 번역부는 상기 적어도 하나의 시맨틱 모델 내부 관계들 간에 중복된 컨셉이 존재하는 경우 축약된 번역 문서를 생성한다.
- [0020] 일 실시예에서, 상기 규칙 기반 번역부는 상기 적어도 하나의 시맨틱 모델 내부에 적어도 둘의 동일한 구조를 가지는 관계들이 존재하는 경우 상기 동일한 구조를 가지는 관계들 중 하나의 관계의 번역만을 생성한다.
- [0021] 상기 본 발명의 일 목적을 달성하기 위한 본 발명의 일 실시예에 따른 온톨로지 기반 문서 요약 장치는 문서 입력부, 도메인 지식 저장부, 시맨틱 모델 생성부, 시맨틱 모델 저장부 및 시맨틱 모델 번역부를 포함한다. 상기 문서 입력부는 문서를 입력 받고 입력된 문서를 출력한다. 상기 도메인 지식 저장부에는 상기 입력된 문서와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장된다. 상기 시맨틱 모델 생성부는 상기 입력된 문

서를 전달받는다. 상기 시맨틱 모델 생성부는 상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 입력된 문서의 시맨틱 모델을 생성한다. 상기 시맨틱 모델 저장부는 상기 입력된 문서의 시맨틱 모델을 저장한다. 상기 시맨틱 모델 번역부는 상기 시맨틱 모델 저장부에 저장된 상기 입력된 문서의 시맨틱 모델을 전달받아 번역 문서를 생성한다.

[0022] 일 실시예에서, 상기 시맨틱 모델 번역부는 상기 도메인 지식 저장부에 포함된 규칙을 이용하여 상기 번역 문서의 내용을 생성하는 규칙 기반 번역부를 포함한다.

[0023] 일 실시예에서, 상기 시맨틱 모델 번역부는 선택적으로 상기 입력된 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 번역 문서의 문단을 생성하는 시맨틱 모델 집합화 처리부를 더 포함한다.

[0024] 일 실시예에서, 상기 규칙 기반 번역부는 상기 입력된 문서의 시맨틱 모델 내부 관계들 간에 중복된 컨셉이 존재하는 경우 축약된 번역 문서를 생성한다.

[0025] 일 실시예에서, 상기 규칙 기반 번역부는 상기 입력된 문서의 시맨틱 모델 내부에 적어도 둘의 동일한 구조를 가지는 관계들이 존재하는 경우 상기 동일한 구조를 가지는 관계들 중 하나의 관계의 번역만을 생성한다.

[0026] 상기 본 발명의 일 목적을 달성하기 위한 본 발명의 일 실시예에 따른 온톨로지 기반 문서 요약 방법은 문서를 입력 받는 단계, 상기 문서와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부 내에 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 문서의 시맨틱 모델을 생성하는 단계, 선택적으로 상기 문서의 시맨틱 모델을 저장하는 단계, 선택적으로 상기 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 문서의 시맨틱 모델의 번역 문서의 문단을 생성하는 시맨틱 모델 집합화하는 단계 및 상기 도메인 지식 저장부에 포함된 규칙에 기반하여 상기 번역 문서의 문단 내용을 생성하는 단계를 포함한다.

[0027] 본 발명의 또 다른 실시예에 따른 온톨로지 기반 문서 추천장치는 문의 (Query) 입력부, 도메인 지식 저장부, 시맨틱 모델 생성부, 시맨틱 모델 저장부, 유사 문서 검색부 및 시맨틱 모델 번역부를 포함한다. 상기 문의 입력부는 문의를 입력 받고 입력된 문의를 출력한다. 상기 도메인 지식 저장부에는 상기 입력된 문의와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장된다. 상기 시맨틱 모델 생성부는 상기 입력된 문의를 전달 받는다. 상기 시맨틱 모델 생성부는 상기 도메인 지식 저장부 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 입력된 문의의 시맨틱 모델을 생성한다. 상기 시맨틱 모델 저장부는 추천 대상 문서들의 시맨틱 모델이 저장한다. 상기 유사 문서 검색부는 상기 입력된 문의의 시맨틱 모델과 유사도가 높은 순서로 적어도 하나의 시맨틱 모델을 상기 시맨틱 모델 저장부에서 추천 문서의 시맨틱 모델로 검색한다. 상기 시맨틱 모델 번역부는 상기 추천 문서의 시맨틱 모델을 전달받아 번역 문서를 생성한다.

[0028] 일 실시예에서, 상기 온톨로지 기반 문서 추천 장치는 상기 도메인 지식 저장부를 이용하여 상기 입력된 문의의 시맨틱 모델을 확장하는 문의 확장부를 더 포함한다.

[0029] 일 실시예에서, 상기 시맨틱 모델 번역부는 규칙 기반 번역부를 포함한다. 상기 시맨틱 모델 번역부는 선택적으로 시맨틱 모델 집합화 처리부를 포함한다. 상기 규칙 기반 번역부는 상기 도메인 지식 저장부에 포함된 규칙을 이용하여 상기 번역 문서의 내용을 생성한다. 상기 시맨틱 모델 집합화 처리부는 상기 추천 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 번역 문서의 문단을 생성한다.

[0030] 상기 본 발명의 일 목적을 달성하기 위한 본 발명의 일 실시예에 따른 온톨로지 기반 문서 추천 방법은 문의를 입력 받는 단계, 상기 문의와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부 내에 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 상기 문의의 시맨틱 모델을 생성하는 단계, 선택적으로 상기 문의의 시맨틱 모델을 저장하는 단계, 선택적으로 상기 도메인 지식 저장부를 이용하여 상기 문의의 시맨틱 모델의 확장된 시맨틱 모델을 생성하는 단계, 상기 확장된 시맨틱 모델과 유사도가 높은 순서로 적어도 하나의 시맨틱 모델을 상기 시맨틱 모델 저장부에서 추천 문서의 시맨틱 모델로 검색하는 단계, 선택적으로 상기 추천 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 추천 문서의 시맨틱 모델의 번역 문서의 문단을 생성하는 시맨틱 모델 집합화하는 단계 및 상기 도메인 지식 저장부에 포함된 규칙에 기반하여 상기 추천 문서의 시맨틱 모델의 번역 문서의 내용을 생성하는 단계를 포함한다.

발명의 효과

- [0031] 본 발명의 실시예들에 따른 문서의 온톨로지 자동 생성 장치는 입력된 문서의 다양한 구조에 불문하여, 문법상 및 의미상 분석 및 정형화된 온톨로지(시맨틱 모델)를 생성한다.
- [0032] 본 발명의 실시예들에 따른 온톨로지 자동 번역 장치는 상기 정형화된 온톨로지(시맨틱 모델)를 기초로 인간이 사용하는 자연어로 구성된 번역문을 자동적으로 생성한다.
- [0033] 본 발명의 실시예들에 따른 온톨로지 기반 문서 요약 장치 및 방법은 입력된 문서의 요약된 문서를 기초로 인간이 사용하는 자연어로 구성된 번역문을 자동적으로 생성한다.
- [0034] 본 발명의 실시예들에 따른 온톨로지 기반 문서 추천 장치 및 방법은 입력된 문 의와 가장 유사도가 높은 문서 들을 찾고, 입력된 문 의와 가장 유사도가 높은 문서들을 기초로 인간이 사용하는 자연어로 구성된 번역문을 자동적으로 생성한다.

도면의 간단한 설명

- [0035] 도 1은 본 발명의 일 실시예에 따른 온톨로지 자동 생성 장치의 일 예를 나타내는 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 온톨로지 자동 생성 장치의 도메인 지식 저장부의 일 예를 나타내는 블록도이다.
- 도 3은 본 발명의 일 실시예에 따른 참조 모델(Reference model)의 일 예를 나타내는 구조도이다.
- 도 4는 본 발명의 일 실시예에 따른 장치(Device) 컨셉의 분류 체계(Taxonomy)를 나타내는 구조도이다.
- 도 5는 본 발명의 일 실시예에 따른 시맨틱 모델 생성부의 일 예를 나타내는 블록도이다.
- 도 6은 본 발명의 일 실시예에 따른 컨셉 인식부의 동작을 나타내는 순서도이다.
- 도 7은 본 발명의 일 실시예에 따른 컨셉 접합부에서 컨셉 간의 관계를 추출하는 방법을 도시한 표이다.
- 도 8은 본 발명의 일 실시예에 따른 온톨로지 자동 번역 장치의 일 예를 나타내는 블록도이다.
- 도 9는 본 발명의 일 실시예에 따른 시맨틱 모델 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- 도 10은 본 발명의 일 실시예에 따른 문서 전체의 그래프 구조의 일 예를 도시한 것이다.
- 도 11은 본 발명의 일 실시예에 따른 규칙 기반 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- 도 12는 본 발명의 일 실시예에 따른 번역 규칙의 일 예를 나타낸 표이다.
- 도 13은 본 발명의 일 실시예에 따른 축약된 번역문을 생성하는 규칙 기반 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- 도 14는 본 발명의 일 실시예에 따른 번역 문장의 축약 규칙의 일 예를 나타낸 표이다.
- 도 15는 본 발명의 일 실시예에 따른 중복문이 제거되고 축약된 번역문을 생성하는 규칙 기반 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- 도 16은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 요약 장치의 일 예를 나타낸 블록도이다.
- 도 17은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 요약 방법의 일 예를 나타낸 순서도이다.
- 도 18은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 추천 장치의 일 예를 나타낸 블록도이다.
- 도 19는 본 발명의 일 실시예에 따른 문의 확장부의 동작의 일 예를 나타낸 순서도이다.
- 도 20은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 추천 방법의 일 예를 나타낸 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0036] 본문에 개시되어 있는 본 발명의 실시예들에 대해서, 특정한 구조적 내지 기능적 설명들은 단지 본 발명의 실시예를 설명하기 위한 목적으로 예시된 것으로, 본 발명의 실시예들은 다양한 형태로 실시될 수 있으며 본문에 설명된 실시예들에 한정되는 것으로 해석되어서는 아니 된다.

- [0037] 본 발명은 다양한 변경을 가할 수 있고 여러 가지 형태를 가질 수 있는 바, 특정 실시예들을 도면에 예시하고 본문에 상세하게 설명하고자 한다. 그러나 이는 본 발명을 특정한 개시 형태에 대해 한정하려는 것이 아니며, 본 발명의 사상 및 기술 범위에 포함되는 모든 변경, 균등물 내지 대체물을 포함하는 것으로 이해되어야 한다. 각 도면을 설명하면서 유사한 참조부호를 구성요소에 대해 사용하였다.
- [0038] 제1, 제2 등의 용어는 다양한 구성요소들을 설명하는데 사용될 수 있지만, 상기 구성요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 사용된다. 예를 들어, 본 발명의 권리 범위로부터 이탈되지 않은 채 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소도 제1 구성요소로 명명될 수 있다.
- [0039] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다거나 "접속되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소가 다른 구성요소에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 중간에 다른 구성요소가 존재하지 않는 것으로 이해되어야 할 것이다. 구성요소들 간의 관계를 설명하는 다른 표현들, 즉 "~사이에"와 "바로 ~사이에" 또는 "~에 이웃하는"과 "~에 직접 이웃하는" 등도 마찬가지로 해석되어야 한다.
- [0040] 본 출원에서 사용한 용어는 단지 특정한 실시예를 설명하기 위해 사용된 것으로, 본 발명을 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 출원에서, "포함하다" 또는 "가지다" 등의 용어는 실시(說示)된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0041] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가지고 있다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥 상 가지는 의미와 일치하는 의미를 가지는 것으로 해석되어야 하며, 본 출원에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0042] 이하, 첨부한 도면들을 참조하여, 본 발명의 바람직한 실시예를 보다 상세하게 설명하고자 한다. 도면상의 동일한 구성요소에 대해서는 동일한 참조부호를 사용하고 동일한 구성요소에 대해서 중복된 설명은 생략한다.
- [0043] 도 1은 본 발명의 일 실시예에 따른 온톨로지 자동 생성 장치의 일 예를 나타내는 블록도이다.
- [0044] 도 1을 참조하면, 온톨로지 자동 생성 장치(100)는 문서 입력부(110), 도메인 지식(Domain knowledge) 저장부(200), 시맨틱 모델 생성부(500) 및 시맨틱 모델 저장부(120)를 포함한다.
- [0045] 문서 입력부(110)는 온톨로지 생성의 기초가 되는 문서(133)를 입력 받아 입력된 문서(130)를 시맨틱 모델 생성부(500)와 도메인 지식 저장부(200)에 전달한다.
- [0046] 도메인 지식 저장부(200)에는 개발 대상인 온톨로지의 도메인과 범위를 결정하는 제1 단계, 선택적으로 기존의 온톨로지를 재사용하는 제2 단계, 상기 개발 대상인 온톨로지의 중요 단어들(131)을 열거하는 제3 단계 및 상기 개발 대상인 온톨로지의 클래스(Class)와 구조(Hierarchy)를 정의하는 제4 단계로 구성되는 통상의 기술자에게 널리 알려져 있는 온톨로지 개발 101(Ontology development 101) 방법에 의해서 입력된 문서(130)와 관련된 도메인 지식(132)이 미리 저장되어 있으며, 도메인 지식(132)은 시맨틱 모델 생성부(500) 및 시맨틱 모델 저장부(120)에 전달된다. 바람직하게는, 도메인 지식 저장부(200)의 도메인 지식(132)은 상기 개발 대상인 온톨로지의 도메인의 통상의 기술자 또는 전문가에 의해서 생성될 수 있다.
- [0047] 시맨틱 모델 생성부(500)는 입력된 문서(130)를 전달받아 도메인 지식 저장부(200) 내에 미리 선언된 제1 객체이며 상기 클래스의 일종인 컨셉(Concept)의 적어도 하나의 제1 인스턴스(Instance) 및 도메인 지식 저장부(200) 내에 미리 선언된 제2 객체이며 상기 구조의 일종인 상기 컨셉 간의 관계(Relationship)의 적어도 하나의 제2 인스턴스를 포함하는 입력된 문서의 시맨틱 모델(131)을 생성한다.
- [0048] 시맨틱 모델 저장부(120)는 입력된 문서의 시맨틱 모델(131)을 도메인 지식(132)을 이용하여 저장한다.
- [0049] 도 2는 본 발명의 일 실시예에 따른 온톨로지 자동 생성 장치의 도메인 지식 저장부의 일 예를 나타내는 블록도이다.

[0050] 도 2를 참조하면, 온톨로지 자동 생성 장치의 도메인 지식 저장부(200)는 동의어 및 유의어 데이터베이스 (Lexicon database (DB); 211), 참조 모델 DB(Reference model DB; 212) 및 규칙 DB(Rules DB; 213)를 포함한다.

[0051] 도메인 지식 저장부(200)는 입력된 문서(130)를 상기 온톨로지 개발 101 방법을 이용하여 동의어 및 유의어 DB(211), 참조 모델 DB(212) 및 규칙 DB(213)에 저장한다. 도메인 지식 저장부(200)의 출력인 도메인 지식 (220)은 동의어 및 유의어 DB(211)의 출력, 참조 모델 DB(212)의 출력 및 규칙 DB(213)의 출력을 포함한다.

[0052] 동의어 및 유의어 DB(211)는 입력된 문서(130)의 도메인의 통상의 기술자 또는 전문가들이 사용하는 주제, 용어 등의 모음이다. 예를 들면, 동의어 및 유의어 DB(211)는 표 1과 같이 구성된다.

표 1

용어	동의어/유의어	약어	설명
Promotion	Advertisement	Promo	...
Customer	Guest	Cus	...
Immediate	Direct	Imm	...

[0053]

[0054] 참조 모델 DB(212)는 참조 모델과 분류 체계(Taxonomy)를 포함한다. 상기 참조 모델은 도 3에서 도시하고, 상기 분류 체계는 도 4에서 도시한다.

[0055] 규칙 DB(213)는 상기 컨셉 간의 관계를 추출하기 위한 시맨틱 규칙(Semantic Rules)과 추출된 컨셉과 관계를 문장 형태로 재구성하기 위해 필요한 번역 규칙(Translation Rules)을 포함한다. 상기 시맨틱 규칙은 도 7에서 도시하고, 상기 번역 규칙은 도12 및 도 14에서 도시한다.

[0056] 도 3은 본 발명의 일 실시예에 따른 참조 모델(Reference model)의 일 예를 나타내는 구조도이다.

[0057] 도 3을 참조하면, 참조 모델은 사람(Person), 기능(Function), 홍보 기본 정보(Basic information of promotion), 속성(Property) 및 단위(Unit)를 컨셉(Concept)으로 포함한다. 세부적으로, 상기 홍보 기본 정보 컨셉은 상표(Brand), 홍보(Promotion), 장치(Device), 상점(Shop), 시기(Date), 홍보 매체(Promotion media) 및 홍보 장소(Promotion area)를 세부 컨셉을 포함한다.

[0058] 상기 컨셉 또는 세부 컨셉은 타 컨셉 또는 타 세부 컨셉과의 관계를 포함할 수 있다. 도 3의 사람의 컨셉은 기능 컨셉과의 관계 (Has-Function), 단위 컨셉과의 관계(Has-Unit) 및 속성 컨셉과의 관계(Has-Property)를 포함한다. 도 3의 나머지 구조는 위 설명으로 비추어 이해할 수 있는 바, 설명을 생략한다.

[0059] 도 4는 본 발명의 일 실시예에 따른 장치(Device) 컨셉의 분류 체계(Taxonomy)를 나타내는 구조도이다.

[0060] 도 4를 참조하면, 장치 컨셉의 분류 체계는 장치 컨셉을 최상위 컨셉으로 가진다. 상기 장치 컨셉은 제품(Product) 컨셉과 부품(Component) 컨셉을 하위 컨셉으로 포함한다. 상기 제품 컨셉은 TV 컨셉, 휴대폰 (Cell-phone) 컨셉, 라디오(Radio) 컨셉, 모니터(Monitor) 컨셉 등을 하위 컨셉으로 포함한다. 도 4의 나머지 구조는 위 설명으로 비추어 이해할 수 있는 바, 설명을 생략한다.

[0061] 도 5는 본 발명의 일 실시예에 따른 시맨틱 모델 생성부의 일 예를 나타내는 블록도이다.

[0062] 도 5를 참조하면, 시맨틱 모델 생성부는 문법 분석부(Syntax analysis part; 510) 및 의미 분석부(Semantic analysis part; 520)를 포함한다.

[0063] 문법 분석부(510)는 토큰 생성부(Tokenization; 511), 단어 매칭부(Word matching; 512), 품사 분석부(513) 및 절(Phrase) 탐지부(514)를 포함한다. 의미 분석부(520)는 컨셉 인식부(521) 및 컨셉 접합부(522)를 포함한다.

- [0064] 토큰 생성부(511)는 입력된 문서(531)를 전달받아 음절 단위로 문장을 분할하여 토큰화된 문서(532)를 생성한다. 토큰화된 문서(532)의 각 토큰 뒤에 부가된 "/"는 토큰 분할 기호를 의미한다.
- [0065] 단어 매칭부(512)는 토큰화된 문서(532) 내의 단어의 축약어를 원 단어로 변환한다. 표 1에 나타난 동의어 및 유의어 DB(211)을 이용하여, 토큰화된 문서(532)의 첫 번째 단어인 "Promo"는 단어 매칭된 문서(533)의 첫 번째 단어인 "Promotion"으로 변환한다.
- [0066] 품사 분석부(513)는 단어 매칭된 문서(533)의 토큰의 품사를 분석하여 토큰의 후단에 부가한 품사 분석된 문서(534)를 생성한다. 품사 분석된 문서(534) 내 "("는 품사가 없으므로 "-None-", "Promotion"은 "NN", "available"은 "JJ"를 태깅한다. 품사 분석된 문서(534) 내 다른 토큰에 대한 동작은 위 설명으로 비추어 이해할 수 있는 바, 설명을 생략한다.
- [0067] 절 탐지부(514)는 품사 분석된 문서(534) 내 명사절(Noun phrase)을 탐지한다. 절이 탐지된 문서(535)는 "the/DT end/NN of/IN May/NNP", "Monitor/NNP TV/NNP model/NN" 및 "a/DT 1000euro/-None- shopping/VBG bonus/NN prize/VB"의 명사절을 포함한다.
- [0068] 컨셉 인식부(521)는 절이 탐지된 문서(535)에서 도메인 지식 저장부(200) 내의 참조 모델 DB(212)를 이용하여 컨셉들을 추출한다. 절이 탐지된 문서(535)에서 추출된 컨셉(536)은 "Promotion(), Customer(), LG(), Purchase(), End(), May(), Monitor-TV() 및 Euro()"이다. 절이 탐지된 문서(535) 내의 토큰에 대한 컨셉이 복수 개 존재하는 경우 컨셉의 인식부의 동작은 도 6에서 설명하도록 한다.
- [0069] 컨셉 접합부(522)는 추출된 컨셉(536) 간의 관계(Relationship)(537)를 추출하고, 추출된 컨셉(536)을 노드(Node)로 하고, 추출된 컨셉 간의 관계(537)를 에지(Edge)로 하는 그래프(Graph) 형태를 가지는 적어도 하나의 시맨틱 모델(537)을 생성한다. 추출된 컨셉 간의 관계(537)는 "Promotion()"과 "May()"의 관계인 "Has-DueDate(Promotion(),May())", "May()"와 "End()"의 관계인 "Has-Unit(May(), End())" 및 "Euro()"의 관계인 "Euro(1000)"을 포함한다.
- [0070] 도 6은 본 발명의 일 실시예에 따른 컨셉 인식부의 동작을 나타내는 순서도이다.
- [0071] 도 6을 참조하면, S610 단계는 절이 탐지된 문서(535)에서 단어 및 단어를 읽어내기 위하여 토큰 또는 토큰구를 읽어낸다.
- [0072] S620 단계는 상기 토큰 또는 토큰구의 품사가 동사인지 여부를 판단한다.
- [0073] S621 단계는 S620 단계에서 상기 토큰 또는 토큰구의 품사가 동사가 아닌 경우, 상기 참조 모델의 모든 컨셉과 비교하여 매칭을 수행한다.
- [0074] S622 단계는 S620 단계에서 상기 토큰 또는 토큰구의 품사가 동사인 경우, 상기 참조 모델의 기능(Function) 컨셉과 비교하여 매칭을 수행한다.
- [0075] S630 단계는 S621 단계 또는 S622 단계에서 수행한 매칭의 결과, 상기 토큰에 매칭된 컨셉의 수가 2개 이상인지를 판단한다.
- [0076] S631 단계는 S630 단계에서 상기 토큰에 매칭된 컨셉의 수가 2개 이상이 아닌 경우, 상기 토큰에 매칭된 컨셉의 수가 하나인지를 판단한다. 상기 토큰에 매칭된 컨셉의 수가 1개가 아니라면, 상기 토큰에 매칭되는 컨셉은 존재하지 않는 것이므로 컨셉을 추출하지 않는다.
- [0077] S632 단계는 S631 단계에서 상기 토큰에 매칭된 컨셉의 수가 1개인 경우, 매칭되는 컨셉을 상기 토큰 또는 토큰구의 컨셉으로 추출한다.
- [0078] S633 단계는 S630 단계에서 상기 토큰에 매칭된 컨셉의 수가 2개 이상인 경우, 컨셉 별로 각각의 컨셉과 매칭되는 단어의 개수에 따라 Wm(Weighted score)을 구한다. 절이 탐지된 문서(535) 중 명사절인 "Monitor/NNP TV/NNP model/NN"을 분류 체계(400)를 이용하여 컨셉에 해당 여부를 판단하면 표 2와 같다. Monitor, TV, model의 단어는 Monitor TV 컨셉과 TV 컨셉에 각각 1, 2, 0 번씩 해당이 되므로, W1(Monitor)=1/3, W2(TV)=2/3, W3(model)=0 이다.

표 2

	Monitor	TV	model
Monitor TV	Monitor	TV	
TV		TV	
Weighted score	1/3	2/3	0

[0079]

[0080]

S634 단계는 상기 Weighted score, 수학식 1 및 수학식 2를 이용하여 T 스코어 (T-score)를 구할 수 있다. 수학식 1에서 W_m'은 명사절의 단어와 컨셉의 단어가 매칭이 된 경우(case 1)에는 W_m의 값을 가지고, 명사절의 단어와 컨셉의 단어가 매칭이 되지 않은 경우(case 2)에는 0의 값을 가진다. 수학식 2에서 N은 명사절 내에서 단어의 수를 의미한다.

수학식 1

$$W_m' = \begin{cases} W_m & \dots\dots\dots (case\ 1) \\ 0 & \dots\dots\dots (case\ 2) \end{cases}$$

[0081]

수학식 2

$$T_{score} = \frac{\# \text{ of matching words} \times \sum_{m=1}^N W_m'}{\# \text{ of words in the concept}}$$

[0082]

[0083]

Monitor TV 컨셉의 T스코어는 (2 * (1/3 + 2/3 + 0)) / 2 = 1, TV 컨셉의 T 스코어 는 (1 * (0 + 2/3 + 0)) / 1 = 2/3 이다.

[0084]

S635 단계는 최대 T 스코어를 가지는 컨셉을 상기 토큰 또는 토큰구의 컨셉으로 추출한다. Monitor TV 컨셉의 T 스코어는 1, TV 컨셉의 T 스코어는 2/3인 바 상기 명사구 "Monitor/NNP TV/NNP model/NN"의 컨셉은 최대 T 스코어를 가지는 Monitor TV 컨셉으로 결정된다.

[0085]

S640 단계는 절이 탐지된 문서(535) 내 모든 토큰 또는 토큰구의 컨셉 인식을 수행하였는지를 판단한다. 절이 탐지된 문서(535)의 마지막 토큰 또는 토큰구에 대한 컨셉을 인식하였으면 컨셉 인식부의 동작을 종료한다. 절이 탐지된 문서(535)에 컨셉을 인식할 토큰 또는 토큰구가 존재하면 S610의 단계를 다시 수행한다.

[0086]

S640 단계에서 절이 탐지된 문서(535)의 마지막 토큰 또는 토큰 구의 컨셉 인식을 완료하면, 컨셉 인식부(521)는 절이 탐지된 문서(535)의 모든 컨셉을 추출한 결과를 생성한다.

[0087]

도 7은 본 발명의 일 실시예에 따른 컨셉 접합부에서 컨셉 간의 관계를 추출하는 방법을 도시한 표이다.

[0088]

도 7을 참조하면, 입력된 문서(130) 내에 존재하는 문장은 "Promotion available until the end of May"이고, 상기 문장의 추출된 컨셉(536)은 "Promotion(), End(), May()"이다. "if(Promotion() & by/until & Date()) then (Has-DueDate(Promotion(), Date()))"이 추출된 컨셉(536)에 상응하는 제1 시맨틱 규칙으로 존재한다면,

상기 문장은 상기 제1 시맨틱 규칙의 조건에 부합하기 때문에 "Has-DueDate(Promotion(), May())"의 관계가 생성된다. 도 7의 나머지 문장의 관계를 추출하는 과정은 위 설명에 비추어 이해할 수 있는 바, 설명을 생략한다.

- [0089] 도 8은 본 발명의 일 실시예에 따른 온톨로지 자동 번역 장치의 일 예를 나타내는 블록도이다.
- [0090] 도 8을 참조하면, 온톨로지 자동 번역 장치(800)는 시맨틱 모델 저장부(810), 도메인 지식 저장부(820) 및 시맨틱 모델 번역부(830)를 포함한다. 시맨틱 모델 번역부(830)는 규칙 기반 번역부(832)를 포함하며, 선택적으로 시맨틱 모델 집합화 처리부(831)를 포함할 수 있다.
- [0091] 시맨틱 모델 저장부(810)는 시맨틱 모델 저장부(810)에 저장된 적어도 하나의 시맨틱 모델(840)을 시맨틱 모델 번역부(831)에 전달한다.
- [0092] 도메인 지식 저장부(820)는 도메인 지식(842)을 시맨틱 모델 번역부(830)에 전달한다. 도메인 지식 저장부(820)는 도 1 및 도 2 를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0093] 시맨틱 모델 번역부(830)는 적어도 하나의 시맨틱 모델(840) 및 도메인 지식(842)을 전달 받아 적어도 하나의 시맨틱 모델(840)의 번역 문서(843)를 생성한다.
- [0094] 시맨틱 모델 집합화 처리부(831)는 적어도 하나의 시맨틱 모델(840) 간의 유사도(Similarity)를 이용하여 적어도 하나의 시맨틱 모델(840)의 번역 문서의 문단을 생성한다. 시맨틱 모델 집합화 처리부(831)의 상세한 동작 방법은 도 9 및 도 10 에서 도시한다.
- [0095] 규칙 기반 번역부(832)는 적어도 하나의 시맨틱 모델(840) 또는 시맨틱 모델 집합화 처리부에 의해 문단이 생성된 시맨틱 모델(841)을 전달받아, 도메인 지식 저장부(820) 내부의 참조 모델 DB(212)의 상기 번역 규칙을 이용하여 상기 번역 문서의 내용을 생성한다. 규칙 기반 번역부(832)의 상세한 동작 방법은 도 11 내지 도 15에서 도시한다.
- [0096] 도 9는 본 발명의 일 실시예에 따른 시맨틱 모델 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- [0097] 도 9를 참조하면, S910 단계는 전달받은 적어도 하나의 시맨틱 모델(840) 간의 유사도를 측정한다. 바람직하게는, 상기 유사도는 통상의 기술자에 널리 알려진 유사도 측정 방법인 Aleander Maedche 방법론을 활용하여 구할 수 있다.
- [0098] S920 단계는 적어도 하나의 시맨틱 모델(840) 중 시맨틱 모델을 그래프(Graph)의 노드(Node)로 포함하고, 상기 시맨틱 모델(840) 간의 유사도의 역수를 상기 그래프의 에지(Edge)로 포함하는 제1 그래프를 생성한다. 상기 제1 그래프의 구조는 도 10 에서 상세하게 설명하도록 한다.
- [0099] S930 단계는 상기 제1 그래프의 에지의 가중치인 유사도의 역수 값이 가장 작은 순서대로 상기 에지의 양 말단의 노드에 해당하는 두 시맨틱 모델을 한 문단에 포함(집합화)시킨다. 적어도 하나의 시맨틱 모델(840)의 번역 문서의 문단의 개수에 따라 상기 포함시키는 작업을 반복한다.
- [0100] S930 단계에서 상기 제1 그래프의 집합화를 완료하면, 적어도 하나의 시맨틱 모델(840)의 번역 문서의 문단이 생성된다.
- [0101] 도 10은 본 발명의 일 실시예에 따른 문서 전체의 그래프 구조의 일 예를 도시한 것이다.
- [0102] 도 10을 참조하면, 문서 전체의 그래프 구조(1000)는 시맨틱 모델 번역부에 의해 생성된 시맨틱 모델들을 노드로서 포함하고, 상기 시맨틱 모델 간의 유사도의 역수를 상기 노드를 연결하는 에지(Edge)의 가중치(Weight)로 포함한다.
- [0103] 문서 전체의 그래프(1000)는 삼성 시맨틱 모델(제1 노드; 1010), 필립스 시맨틱 모델(제2 노드; 1011) 및 엘지 시맨틱 모델(제3 노드; 1012)을 포함한다. 삼성 시맨틱 모델(1010)과 필립스 시맨틱 모델(1011) 간의 제1 에지(1020)는 삼성 시맨틱 모델(1010)과 필립스 시맨틱 모델(1011)의 유사도 4의 역수인 1/4의 가중치를 가진다. 삼성 시맨틱 모델(1010)과 엘지 시맨틱 모델(1012) 간의 제2 에지(1021)는 삼성 시맨틱 모델(1010)과 엘지 시맨틱 모델(1012)의 유사도 1의 역수인 1의 가중치를 가진다. 필립스 시맨틱 모델(1011)과 엘지 시맨틱 모델(1012) 간의 제3 에지(1022)는 필립스 시맨틱 모델(1011)과 엘지 시맨틱 모델(1012)의 유사도 2의 역수인 1/2의 가중치를 가진다.
- [0104] 번역 문서의 문단의 개수가 2개인 경우, 도 10의 경우, 두 개의 시맨틱 모델을 하나의 문단으로 집합화 하면 된다. 문서 전체의 그래프(1000) 중 최소의 가중치를 가지는 제1 에지(1020)를 선택하여, 삼성 시맨틱 모델(101

0)과 필립스 시맨틱 모델(1011)을 하나의 문단, 엘지 시맨틱 모델(1012)을 다른 하나의 문단으로 집합화한다.

- [0105] 도 11은 본 발명의 일 실시예에 따른 규칙 기반 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- [0106] 도 11을 참조하면, S1110 단계는 문단이 생성된 시맨틱 모델(841)에서 제1 관계를 선택한다.
- [0107] S1120 단계는 상기 제1 관계를 도메인 지식 저장부(820) 내의 규칙 DB(213)를 이용하여 번역 문장을 생성한다.
- [0108] S1130 단계는 상기 제1 관계가 문단이 생성된 시맨틱 모델(841)의 마지막 관계인지를 판단한다. 상기 제1 관계가 문단이 생성된 시맨틱 모델(841)의 마지막 관계이면 문단이 생성된 시맨틱 모델(841)에서 제2 관계에 대하여 S1110 단계 및 S1120 단계를 반복 수행한다.
- [0109] S1140 단계는 S1130 단계에서 상기 제1 관계가 문단이 생성된 시맨틱 모델(841)의 마지막 관계인 경우, 컨셉 집합부(522)에서 관계를 추출한 순서대로 상기 번역 문장을 나열한다.
- [0110] S1140 단계에서 모든 번역 문장의 나열이 완료되면, 규칙 기반 번역부(832)는 적어도 하나의 시맨틱 모델(840) 또는 문단이 생성된 시맨틱 모델(841)의 번역 문서의 내용을 생성하게 된다.
- [0111] 도 12는 본 발명의 일 실시예에 따른 번역 규칙의 일 예를 나타낸 표이다.
- [0112] 도 12를 참조하면, 컨셉의 경우, "Concept(Instance)"은 "Instance"로 번역, "Unit(Instance)"은 "Instance + Unit()"으로 번역한다. 예를 들면, "Brand(LG)"는 "LG"로 번역, "Euro(100)"은 "100 Euro"로 번역한다.
- [0113] 관계의 경우, "HasFunction(Range, Domain)"은 "Range+Domain"으로 번역, "HasObject(Range, Domain)"의 경우 "Range+Domain"으로 번역한다. 예를 들면, "HasFunction (Customer(), buy())"는 "Customer buy"로 번역, "HasObject (buy(), Product())"는 "buy Product"로 번역한다. 도 12의 나머지 경우의 번역 과정은 위 설명에 비추어 이해할 수 있는 바, 설명을 생략한다.
- [0114] 도 13은 본 발명의 일 실시예에 따른 축약된 번역문을 생성하는 규칙 기반 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- [0115] 도 13을 참조하면, S1310 단계는 문단이 생성된 시맨틱 모델(841)의 모든 관계들을 나열한다.
- [0116] S1320 단계는 도 12의 번역 규칙에 의해 상기 나열한 모든 관계들에 대한 적어도 하나의 번역 문장을 생성한다.
- [0117] S1330 단계는 상기 적어도 하나의 번역 문장 내에 단어의 중복이 있는지를 판단한다.
- [0118] S1331 단계는 상기 적어도 하나의 번역 문장 내에 단어의 중복이 존재하는 경우, 단어의 중복이 존재하는 문장들을 접속사 또는 관계 대명사를 이용해 축약 과정을 수행한 후, S1330 단계를 다시 수행한다. 상기 축약 과정의 규칙은 도 14에서 상세히 설명하도록 한다.
- [0119] S1340 단계는 상기 적어도 하나의 번역 문장 내에 단어의 중복이 더 이상 존재하지 않는 경우, 컨셉 집합부(522)에서 관계를 추출한 순서대로 번역 문장을 나열한다.
- [0120] S1340 단계에서 모든 번역 문장의 나열이 완료되면, 규칙 기반 번역부(832)는 적어도 하나의 시맨틱 모델(840) 또는 문단이 생성된 시맨틱 모델(841)의 축약된 번역 문서의 내용을 생성하게 된다.
- [0121] 도 14는 본 발명의 일 실시예에 따른 번역 문장의 축약 규칙의 일 예를 나타낸 표이다.
- [0122] 도 14를 참조하면, 관계 대명사에 대하여, "S1(제1 주어) + V1(제1 동사) + O1(제1 목적어)"과 "S2(제2 주어) + V2(제2 동사) + O2(제2 목적어)"의 문장에서 O1과 S2의 단어가 동일한 경우, 이를 "S1 + V1 + O1 + 관계대명사 + V2 + O2"의 문장으로 축약하는 규칙이 표현되어 있다. 예를 들면, 제1 문장인 "Customer buy M62."와 제2 문장인 "M62 has lottery promotion."의 경우, "M62"는 제1 문장의 O1이면서, 제2 문장의 S2이기도 한 바, 상기 제1 문장과 상기 제2 문장은 "Customer buy M62 which has lottery promotion."의 문장으로 축약할 수 있다.
- [0123] 접속사에 대하여, "S1 + V1 + O1"과 "S2 + V2 + O2"의 문장에서 V1과 V2가 동일하고, O1과 O2가 동일한 경우, 이를 "S1 and S2 + V1 + O1"의 문장으로 축약하는 규칙이 표현되어 있다. 예를 들면, 제3 문장인 "Phillips has bundle promotion."과 제4 문장인 "Samsung has bundle promotion."의 경우, V1과 V2가 "has"로 동일하고, O1과 O2가 "bundle promotion"으로 동일한 바, 상기 제3 문장과 상기 제4 문장은 "Phillips and Samsung has bundle promotion."의 문장으로 축약할 수 있다. 도 14의 나머지 경우는 위 설명에 비추어 이해할 수 있는 바, 설명을 생략한다.

- [0124] 도 15는 본 발명의 일 실시예에 따른 중복문이 제거되고 축약된 번역문을 생성하는 규칙 기반 번역부의 동작 방법의 일 예를 나타내는 순서도이다.
- [0125] 도 15를 참조하면, S1510단계는 문단이 생성된 시맨틱 모델(841)의 모든 관계들을 나열한다.
- [0126] S1520 단계는 도12의 번역 규칙에 의해 상기 나열한 모든 관계들에 대한 적어도 하나의 번역 문장을 생성한다.
- [0127] S1530 단계는 상기 적어도 하나의 번역 문장 내에 단어의 중복이 존재하는지 판단한다.
- [0128] S1531 단계는 S1530 단계에서 상기 적어도 하나의 번역 문장 내에 단어의 중복이 존재하는 경우, 상기 단어를 포함하는 번역 문장들의 단어가 모두 중복되는지를 판단한다.
- [0129] S1532 단계는 S1531 단계에서 상기 단어를 포함하는 번역 문장들의 단어가 모두 중복되지 않는 경우, 단어의 중복이 존재하는 문장들을 접속사 또는 관계 대명사를 이용해 축약 과정을 수행한 후, S1530 단계를 다시 수행한다. 상기 축약 과정의 규칙은 도 12 를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0130] S1533 단계는 S1531 단계에서 상기 단어를 포함하는 번역 문장들의 단어가 모두 중복되는 경우, 중복된 번역 문장이 존재하는 것인바, 중복 번역 문장은 삭제한다.
- [0131] S1540 단계는 상기 적어도 하나의 번역 문장 내에 단어의 중복이 더 이상 존재하지 않는 경우, 컨셉 집합부(522)에서 관계를 추출한 순서대로 번역 문장을 나열한다.
- [0132] S1540 단계에서 모든 번역 문장의 나열이 완료되면, 규칙 기반 번역부(832)는 적어도 하나의 시맨틱 모델(840) 또는 문단이 생성된 시맨틱 모델(841)의 중복문이 제거되고, 축약된 번역 문서의 내용을 생성하게 된다.
- [0133] 도 16은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 요약 장치의 일 예를 나타낸 블록도이다.
- [0134] 도 16을 참조하면, 온톨로지 기반 문서 요약 장치(1600)는 문서 입력부(1610), 도메인 지식 저장부(1620), 시맨틱 모델 생성부(1630), 시맨틱 모델 저장부(1640) 및 시맨틱 모델 번역부(1650)를 포함한다.
- [0135] 문서 입력부(1610)는 문서(1665)를 입력 받아 입력된 문서(1660)를 시맨틱 모델 생성부(1630) 및 도메인 지식 저장부(1620)에 전달한다.
- [0136] 도메인 지식 저장부(1620)는 입력된 문서(1660)를 전달 받아 도메인 지식(1661)을 생성하여 저장하며, 도메인 지식(1661)은 시맨틱 모델 생성부(1630), 시맨틱 모델 저장부(1640) 및 시맨틱 모델 번역부(1650)에 전달된다. 도메인 지식 저장부(1620)의 구조는 도 2 를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0137] 시맨틱 모델 생성부(1630)는 입력된 문서(1660)를 전달 받아 도메인 지식 저장부(1620) 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 도메인 지식 저장부(1620) 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 입력된 문서의 시맨틱 모델(1662)을 생성한다. 시맨틱 모델 생성부(1630)의 구조는 도 5 를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0138] 시맨틱 모델 저장부(1640)는 입력된 문서의 시맨틱 모델(1662)을 도메인 지식(1661)을 이용하여 저장한다.
- [0139] 시맨틱 모델 번역부(1650)는 시맨틱 모델 저장부(1640)에 저장된 입력된 문서의 시맨틱 모델(1663) 및 도메인 지식(1661)을 전달받아 저장된 입력된 문서의 시맨틱 모델(1663)의 번역 문서(1664)를 생성한다. 시맨틱 모델 번역부(1650)의 구조는 도 8 내지 도 15 를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0140] 온톨로지 기반 문서 요약 장치(1600)는 문서(1665)를 전달받아 도메인 지식(1661) 상에 존재하는 주제, 용어에 관련된 컨셉 만을 추출하고, 상기 컨셉 간의 관계를 추출하여 적어도 하나의 입력된 문서의 시맨틱 모델(1662)을 생성하게 되는 바, 이 과정에서 도메인 지식에 관련된 내용으로 요약된 문서가 생성된다. 저장된 입력된 문서의 시맨틱 모델(1663)을 전달받아 시맨틱 모델 번역부(1650)는 번역문을 생성하는 바, 상기 요약된 문서의 번역문이 생성된다.
- [0141] 도 17은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 요약 방법의 일 예를 나타낸 순서도이다.
- [0142] 도 17을 참조하면, S1710 단계는 문서를 입력 받는다.
- [0143] S1720 단계는 상기 문서와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부 내에 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 적어도 하나의 상기 문서의 시맨틱 모델

을 생성한다.

- [0144] S1730 단계는 선택적으로 상기 문서의 시맨틱 모델을 저장한다.
- [0145] S1740 단계는 선택적으로 상기 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 문서의 시맨틱 모델의 번역 문서의 문단을 생성(집합화)한다. S1740의 단계는 도 10 을 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0146] S1750 단계는 상기 도메인 제식 저장부에 포함된 규칙에 기반하여 상기 번역 문서의 내용을 생성한다.
- [0147] S1750 단계가 완료되면, 상기 문서의 요약된 번역문이 생성된다.
- [0148] 도 18은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 추천 장치의 일 예를 나타낸 블록도이다.
- [0149] 도 18을 참조하면, 온톨로지 기반 문서 추천 장치(1800)는 문의(Query) 입력부(1810), 도메인 지식 저장부(1820), 문의의 시맨틱 모델 생성부(1830), 시맨틱 모델 저장부(1840), 유사 문서 검색부(1860) 및 시맨틱 모델 번역부(1870)를 포함한다. 온톨로지 기반 문서 추천 장치(1800)는 선택적으로 문의 확장부(Query extension)를 더 포함할 수 있다.
- [0150] 문의 입력부(1810)는 추천 문서를 찾을 수 있는 중심 단어인 문의(1887)를 입력 받아 입력된 문의(1880)를 문의의 시맨틱 모델 생성부(1830) 및 도메인 지식 저장부(1820)에 전달한다.
- [0151] 도메인 지식 저장부(1820)는 입력된 문의(1880)를 전달 받아 도메인 지식(1881)을 생성하고 저장하며, 도메인 지식(1881)은 문의의 시맨틱 모델 생성부(1830), 시맨틱 모델 저장부(1840), 문의 확장부(1850), 유사 문서 검색부(1860) 및 시맨틱 모델 번역부(1870)에 전달된다. 도메인 지식 저장부(1820)의 구조는 도 2 를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0152] 문의의 시맨틱 모델 생성부(1830)는 입력된 문의(1880)를 전달 받아 도메인 지식 저장부(1820) 내에 미리 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 도메인 지식 저장부(1820) 내에 미리 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 입력된 문의의 시맨틱 모델(1882)을 생성한다. 시맨틱 모델 생성부(1830)의 구조는 도 5를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0153] 시맨틱 모델 저장부(1840)는 입력된 문의의 시맨틱 모델(1882)을 도메인 지식(1881)을 이용하여 저장한다.
- [0154] 문의 확장부(1850)는 저장된 입력된 문의의 시맨틱 모델(1883)을 전달 받아 동의어 및 유의어 DB(211)를 이용하여 확장된 문의의 시맨틱 모델(1884)을 생성한다. 문의의 확장의 방법은 도 19에서 상세히 설명하도록 한다.
- [0155] 유사 문서 검색부(1860)는 확장된 문의의 시맨틱 모델(1884)과 유사도가 높은 순서로 시맨틱 모델 저장부(1840)에서 적어도 하나의 추천 문서의 시맨틱 모델(1885)을 검색한다.
- [0156] 시맨틱 모델 번역부(1870)는 추천 문서의 시맨틱 모델(1885)을 전달받아 추천 번역 문서(1886)를 생성한다. 시맨틱 모델 번역부(1870)의 내부 구조는 도 8 내지 도 15를 참조하여 상세히 설명하였으므로 중복된 설명은 생략한다.
- [0157] 온톨로지 기반 문서 추천 장치(1800)는 입력된 문의(18807)와 유사도가 높은 순서로 시맨틱 모델 저장부(1840)에서 검색한 적어도 하나의 추천 문서의 시맨틱 모델(1885)의 번역 문서(1886)를 생성한다.
- [0158] 도 19는 본 발명의 일 실시예에 따른 문의 확장부의 동작의 일 예를 나타낸 순서도이다.
- [0159] 도 19를 참조하면, S1910 단계는 저장된 입력된 문의의 시맨틱 모델(1883)에서 제1 컨셉을 선택한다.
- [0160] S1920 단계는 상기 제1 컨셉의 동의어 또는 유의어가 동의어 및 유의어 DB(211) 내 존재하는지 여부를 판단한다. 상기 제1 컨셉의 동의어 또는 유의어가 동의어 및 유의어 DB(211) 내 존재하지 않는 경우, 상기 제1 컨셉은 확장하지 않는다.
- [0161] S1930 단계는 S1920 단계에서 상기 제1 컨셉의 동의어 또는 유의어가 동의어 및 유의어 DB(211) 내 존재하는 경우, 상기 제1 컨셉을 상기 제1 컨셉의 동의어 또는 유의어로 확장한다.
- [0162] S1940 단계는 상기 제1 컨셉이 저장된 입력된 문의의 시맨틱 모델(1883) 내에서 확장이 검토된 마지막 컨셉인지 여부를 판단한다. 상기 제1 컨셉이 저장된 입력된 문의의 시맨틱 모델(1883) 내에서 확장이 검토된 마지막 컨셉인 경우 동작을 종료하고, 상기 제1 컨셉이 저장된 입력된 문의의 시맨틱 모델(1883) 내에서 확장이 검토된 마지막 컨셉이 아닌 경우 저장된 입력된 문의의 시맨틱 모델(1883) 내 제 2 컨셉에 대해서 S1910 단계를 다시 수

행한다.

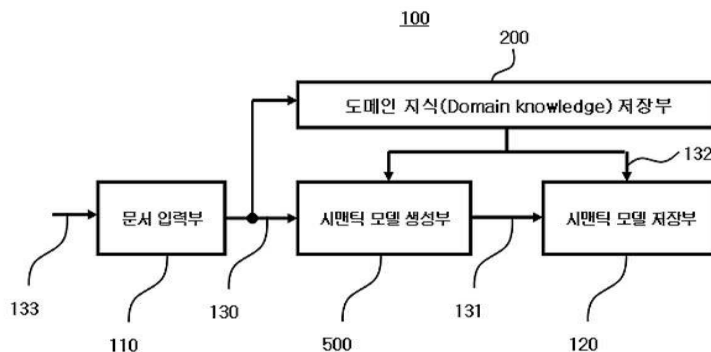
- [0163] S1940 단계가 완료되면, 저장된 입력된 문의의 시맨틱 모델(1883) 내에서 동의어 및 유의어 DB(211)을 이용하여 확장된 컨셉을 포함하는 확장된 문의의 시맨틱 모델(1884)이 생성된다.
- [0164] 도 20은 본 발명의 일 실시예에 따른 온톨로지 기반 문서 추천 방법의 일 예를 나타낸 순서도이다.
- [0165] 도 20을 참조하면, S2010 단계는 문의를 입력 받는다.
- [0166] S2020 단계는 상기 문의와 관련된 도메인 지식이 온톨로지 개발 101 방법에 의해 미리 저장되는 도메인 지식 저장부 내에 선언된 제1 객체인 컨셉의 적어도 하나의 제1 인스턴스 및 상기 도메인 지식 저장부 내에 선언된 제2 객체인 상기 컨셉 간의 관계의 적어도 하나의 제2 인스턴스를 포함하는 상기 문의의 시맨틱 모델을 생성한다.
- [0167] S2030 단계는 선택적으로 상기 문의의 시맨틱 모델을 저장한다.
- [0168] S2040 단계는 상기 도메인 지식 저장부를 이용하여 확장된 문의의 시맨틱 모델을 생성한다.
- [0169] S2050 단계는 상기 확장된 문의의 시맨틱 모델과 유사도가 높은 순서로 시맨틱 모델 저장부에서 적어도 하나의 추천 문서의 시맨틱 모델로 검색한다.
- [0170] S2060 단계는 선택적으로 상기 적어도 하나의 추천 문서의 시맨틱 모델 간의 유사도를 이용하여 상기 적어도 하나의 추천 문서의 시맨틱 모델의 번역 문서의 문단을 생성(집합화)한다.
- [0171] S2070 단계는 상기 도메인 지식 저장부에 포함된 규칙에 기반하여 상기 적어도 하나의 추천 문서의 시맨틱 모델의 번역 문서의 내용을 생성한다.
- [0172] S2070 단계가 완료되면, 입력된 상기 문의와 유사도가 높은 순서로 상기 시맨틱 모델 저장부 내의 상기 적어도 하나의 추천 문서의 번역 문서가 생성된다.

산업상 이용가능성

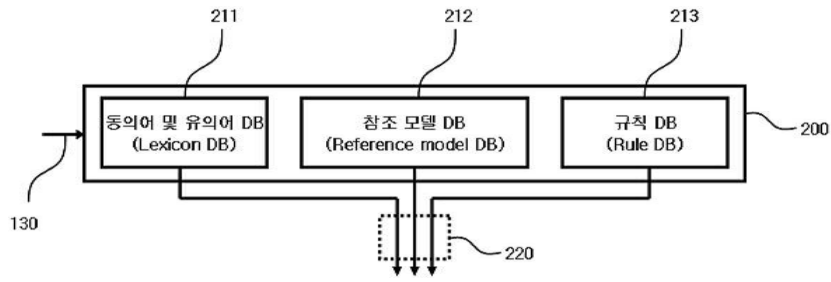
- [0173] 본 발명은 다수의 문서들이 생성되는 분화된 여러 과정을 가지는 회사의 시스템에 유용하게 이용될 수 있다. 특히 회사 시스템 상의 구성원이 자신이 속하지 않은 타 과정의 다수의 문서를 빠르게 이해하고 결정을 내리는데 더욱 유용하게 이용될 수 있다.
- [0174] 상술한 바와 같이, 본 발명의 바람직한 실시예를 참조하여 설명하였지만 해당 기술 분야에서 통상의 지식을 가진 자라면 하기의 특허청구범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

도면

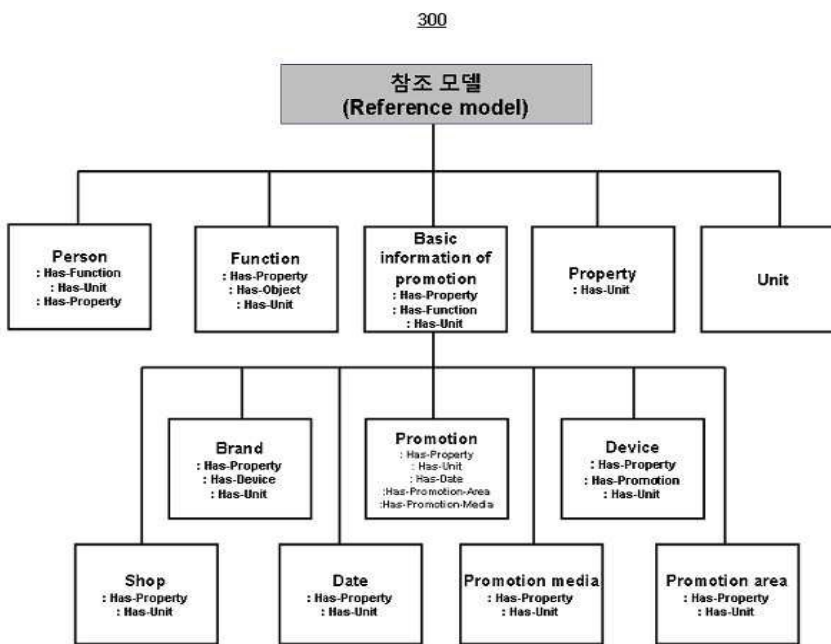
도면1



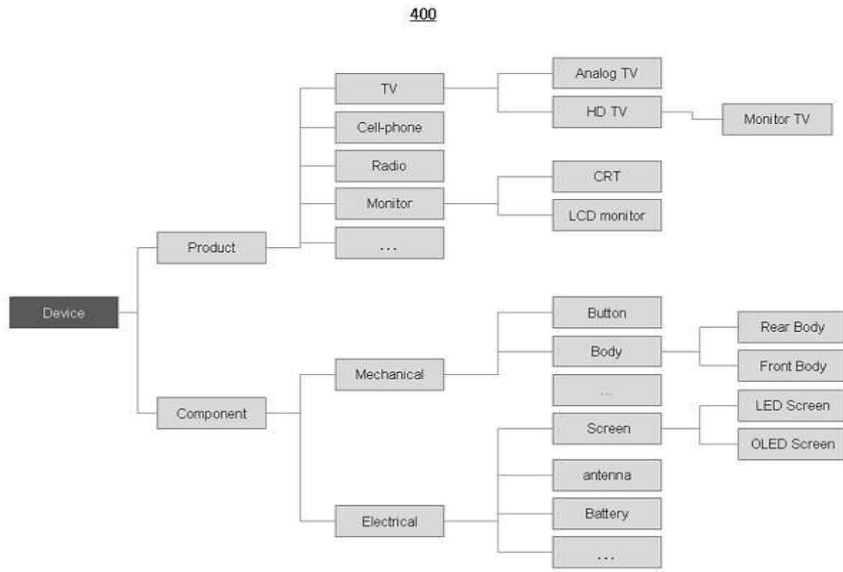
도면2



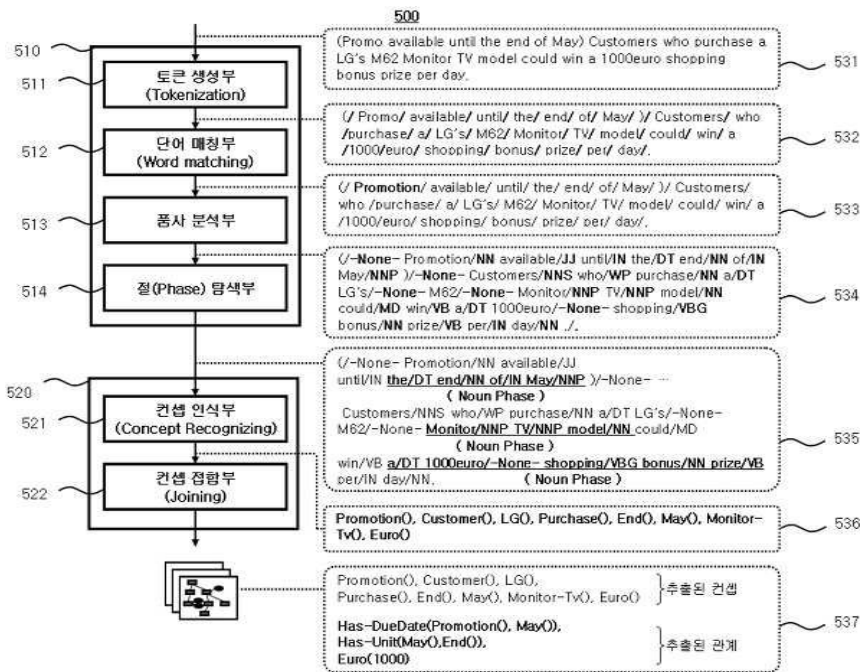
도면3



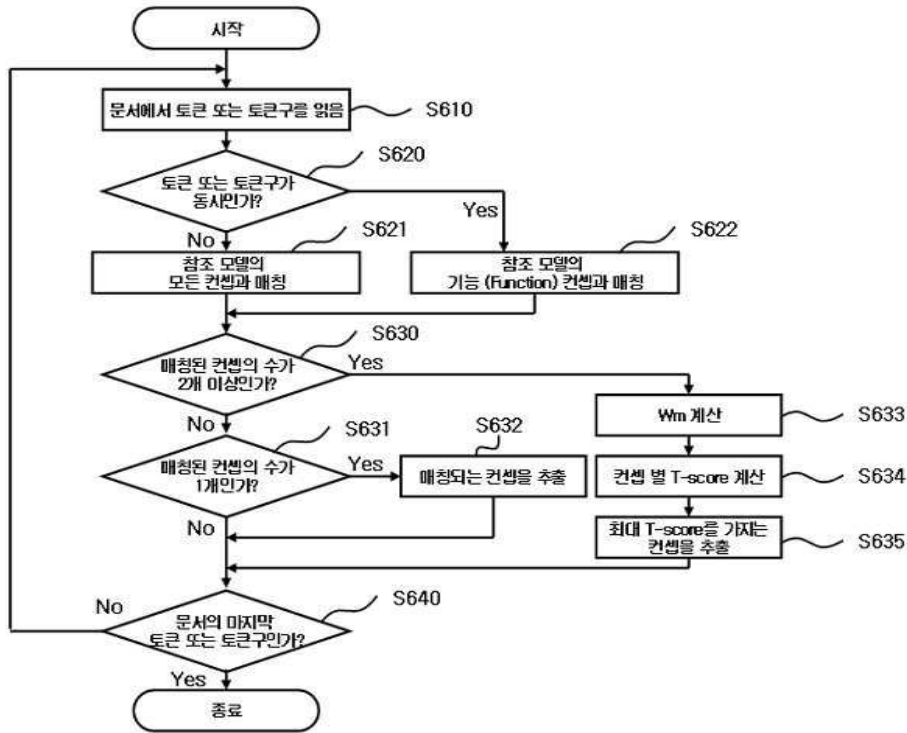
도면4



도면5



도면6

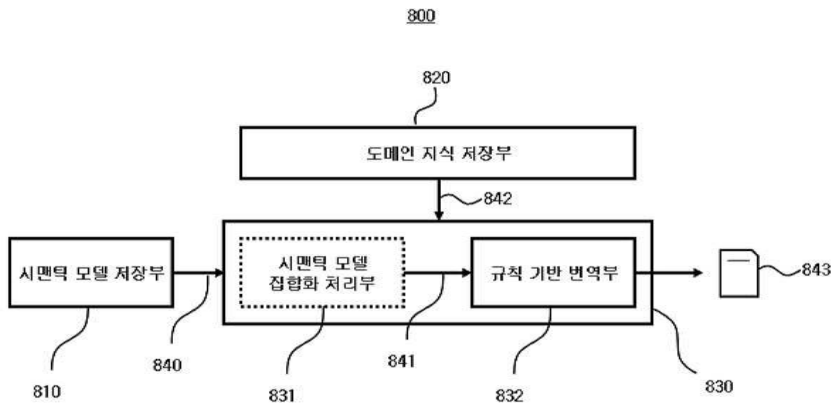


도면7

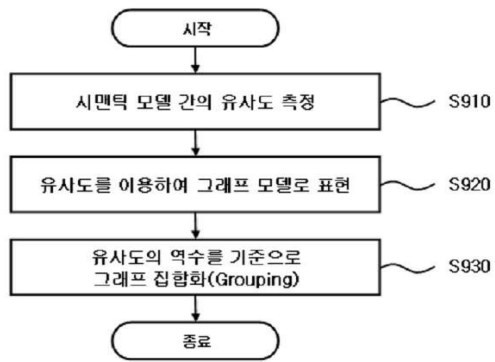
700

문장	추출된 컨셉	시맨틱 규칙	추출된 관계
Promotion available until the end of May	Promotion(), End(), May()	if (Promotion() & by/until & Date()) then (Has-DueDate(Promotion(), Date()))	Has-DueDate (Promotion(), May())
End of May	End(), May()	if (Unit() ^ of ^ Any concept()) then (Has-Unit(Concept(), Unit()))	Has-Unit (May(),End())
1000 Euro	Euro()	if (Numerical Unit ^ Unit()) then (Unit(Numerical Unit))	Euro(1000)
...

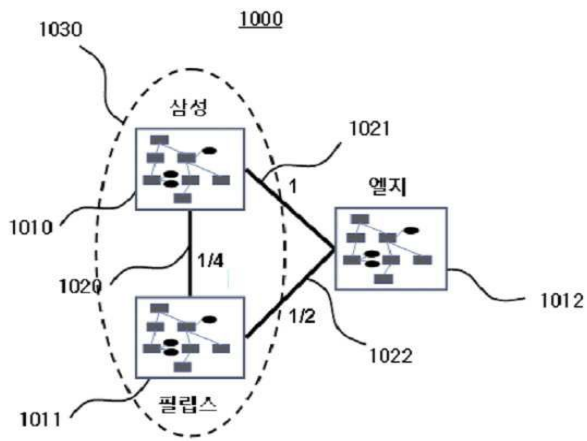
도면8



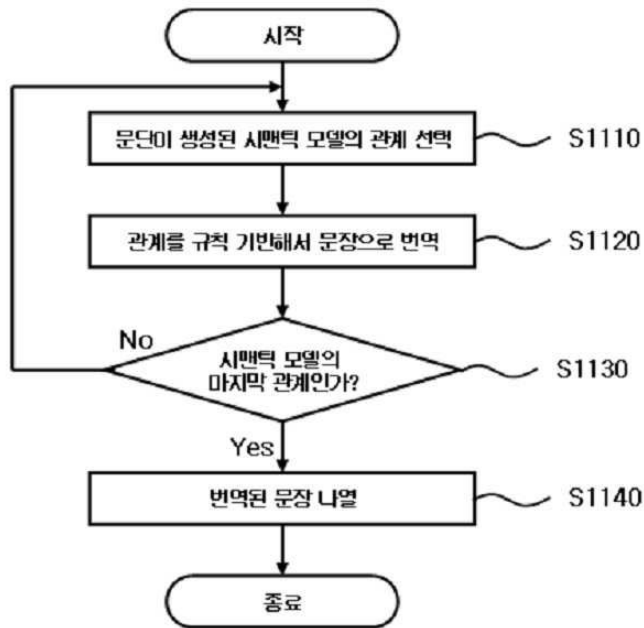
도면9



도면10



도면11

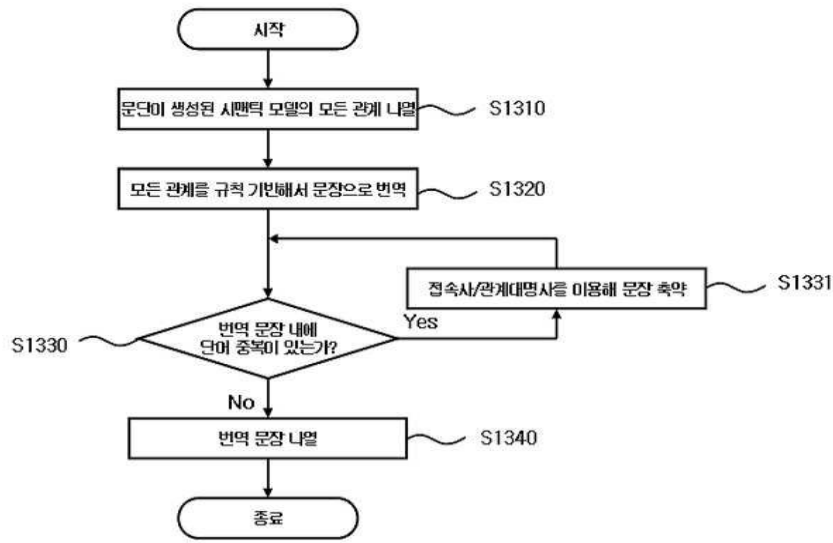


도면12

1200

	경우	규칙	적용 전 예제	적용 후 예제
컨셉	Concept(Instance)	Instance	Brand(LG)	LG
	Unit(Instance)	Instance + Unit()	Euro(100)	100 Euro
	~ Object()	No + Object()	~ Positive()	No positive
	~ Function()	Doesn't + Function()	HasFunction (Customer(), ~buy())	Customer doesn't buy
관계	HasFunction (Range, Domain)	Range + Domain	HasFunction (Customer(), buy())	Customer buy
	HasObject (Range, Domain)	Range + Domain	HasObject (buy(), Product())	Buy Product
	HasUnit (Range, Domain)	Domain + Range	HasUnit (M62, Euro(100))	100 Euro M62
	HasProperty (Range, Domain)	Range + has + Domain	Has-Property (Lottery(), Feedback())	Lottery has Feedback
	HasX (Range, Domain)	Range + has + Domain + X	HasDue-date (Lottery(), End of Max())	Lottery has End of Max Due-date

도면13

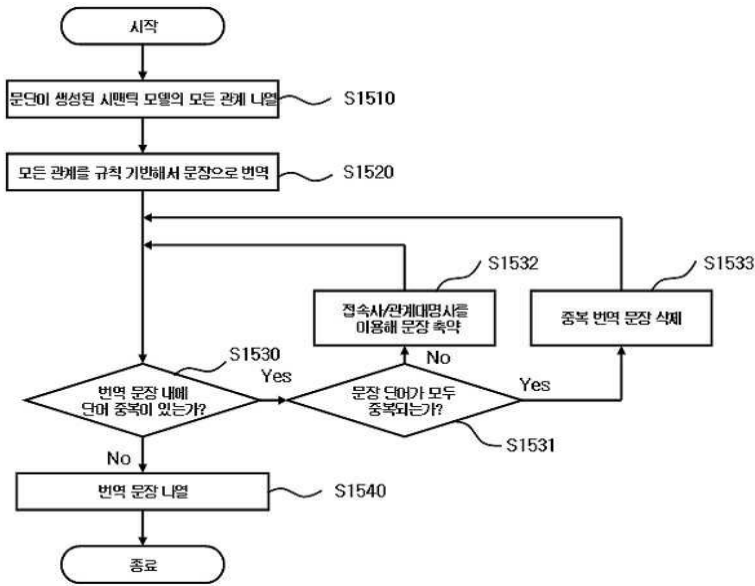


도면14

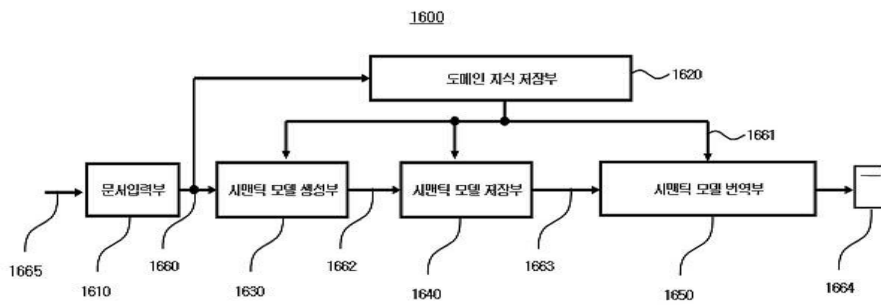
1400

	경우	규칙	적용 전 예제	적용 후 예제
관계대명사	O1 과 S2 과 같을 경우	S1 + V1 + O1 + 관계대명사 + V2 + O2	- Customer buy M62, - M62 has lottery promotion,	- Customer buy M62 which has lottery promotion,
	S1 과 S2 과 같을 경우	S1 + 관계대명사 + V2 + O2 + V1 + O1	- Samsung has bundle promotion - Samsung sell Full HDTV	- Samsung which sell Full HDTV has bundle promotion,
접속사	V1, V2 과 O1, O2 가 같을 경우	S1 and S2 + V1 + O1	- Philips has bundle promotion, - Samsung has bundle promotion	- Philips and Samsung has bundle promotion
	S1, S2 과 V1, V2 가 같을 경우	S1 + V1 + O1 and O2	- Customer buy M62, - Customer buy BDP2000,	- Customer buy M62 and BDP2000,

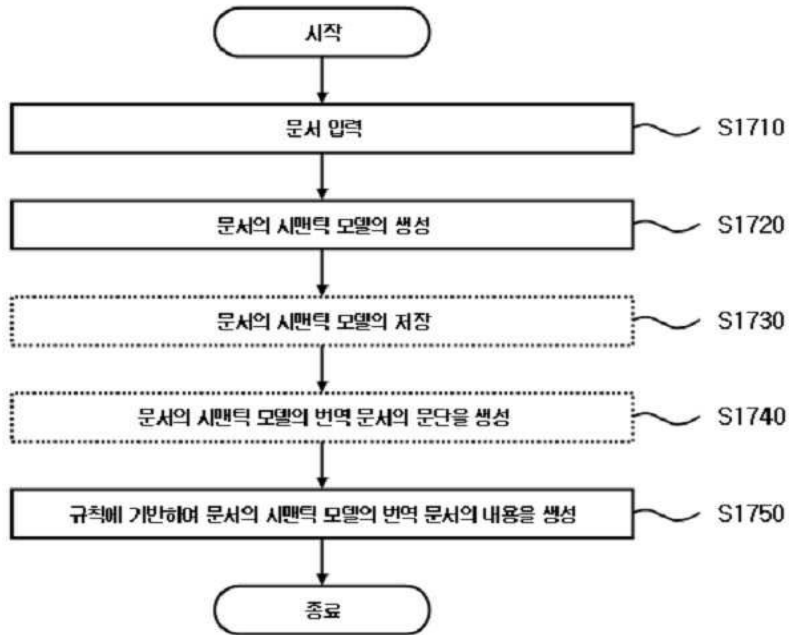
도면15



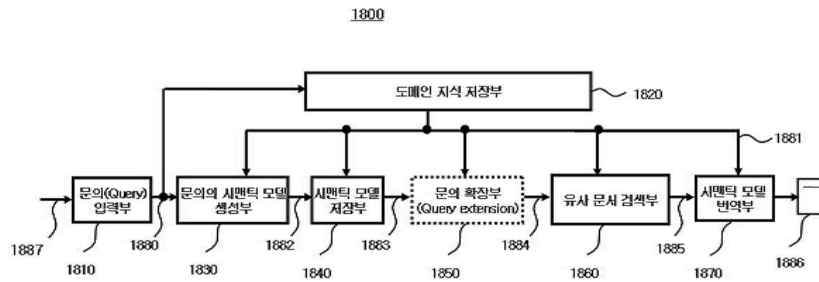
도면16



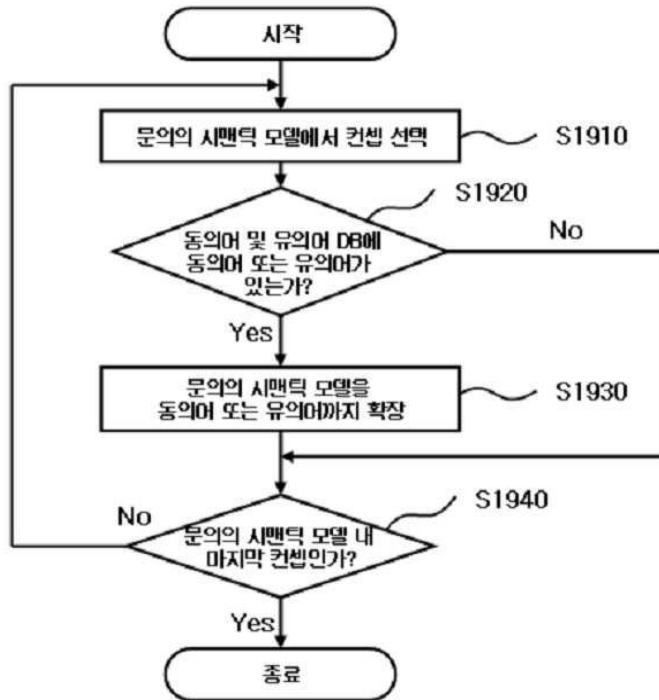
도면17



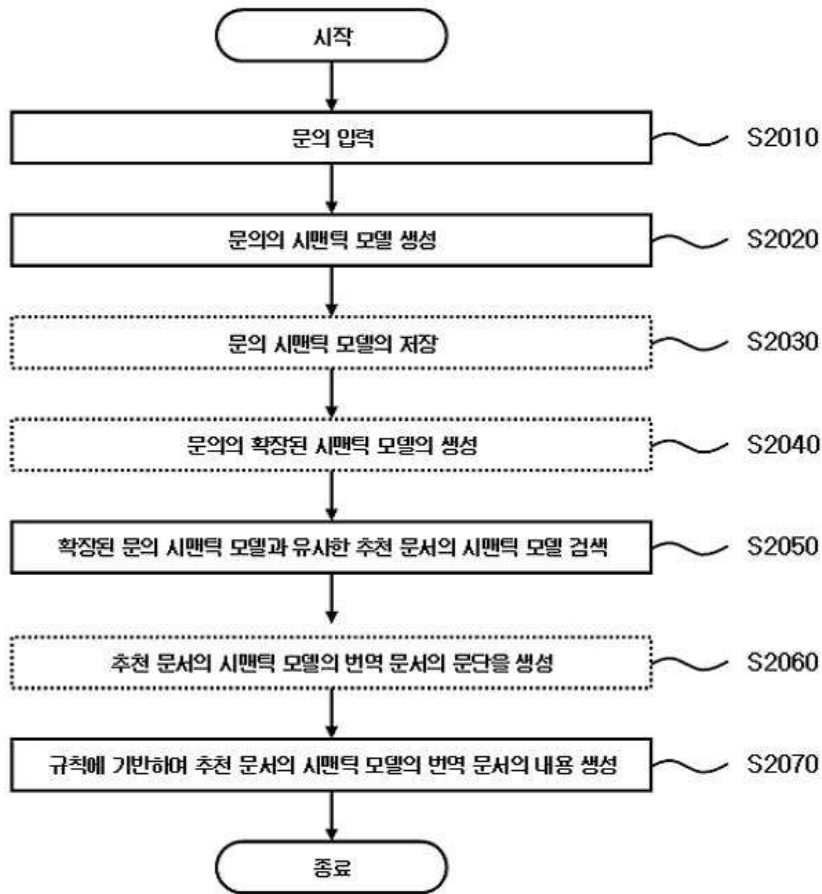
도면18



도면19



도면20



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 14항

【변경전】

온톨로지를 이용한 문서 요약 장치

【변경후】

온톨로지 기반 문서 요약 장치