



(19)대한민국특허청(KR)
(12) 공개특허공보(A)

(51) 。 Int. Cl.

G06F 17/30 (2006.01)

G06F 17/40 (2006.01)

(11) 공개번호

10-2007-0022448

(43) 공개일자

2007년02월27일

(21) 출원번호 10-2005-0076678

(22) 출원일자 2005년08월22일

심사청구일자 2005년08월22일

(71) 출원인 울산대학교 산학협력단
울산 남구 무거동 산29

(72) 발명자 배재학
울산 남구 무거2동 823-3번지 청구하이츠 1408호

(74) 대리인 장한특허법인

전체 청구항 수 : 총 14 항

(54) 온톨로지 재사용을 위한 범주 재분류 장치 및 방법

(57) 요약

본 발명은 온톨로지 재사용을 위한 범주 재분류 장치 및 방법에 관한 것으로, 기구축된 온톨로지를 목적하는 바에 맞도록 변환하여 재사용할 수 있도록 온톨로지 범주 정보를 단순화하거나 구체화시켜 재분류하는, 온톨로지 재사용을 위한 범주 재분류 장치 및 방법에 관한 것임.

본 발명은 온톨로지를 재사용하기 위하여 범주를 재분류하는 장치로서, 재분류 범주의 대응 선호도를 이용하여 기구축된 온톨로지의 범주표제에 대한 범주표제 대응성을 산출하고, 기저범주를 결정하기 위한 범주표제 재분류 수단, 상기 기구축된 온톨로지의 참조정보를 이용하여 상기 범주표제 재분류 수단에서 결정한 기저범주와 온톨로지 범주간의 반경을 측정하여 거리 선호도를 결정한 후, 상기 재분류 범주의 대응 선호도와 상기 결정한 거리 선호도를 이용하여 기저범주 대응성을 산출하기 위한 참조정보 재분류 수단 및 상기 범주표제 재분류 수단에서 산출한 범주표제 대응성 및 상기 참조정보 재분류 수단에서 산출한 기저범주 대응성을 결합시켜 표준 선호도를 산출하고, 산출한 표준 선호도를 이용하여 상기 기구축된 온톨로지의 범주를 재분류하기 위한 통합 재분류 수단을 포함함.

대표도

도 1

특허청구의 범위

청구항 1.

온톨로지를 재사용하기 위하여 범주를 재분류하는 장치로서,

재분류 범주 간의 중요도를 나타내는 대응 선호도를 이용하여, 기구축된 온톨로지의 범주표제에 대한 범주표제 대응성을 산출하고, 산출한 범주표제 대응성을 이용하여 기저범주를 결정하기 위한 범주표제 재분류 수단;

상기 기구축된 온톨로지의 참조정보를 이용하여 상기 범주표제 재분류 수단에서 결정한 기저범주와 온톨로지 범주간의 반경을 측정하여 거리 선호도를 결정한 후, 상기 재분류 범주의 대응 선호도와 상기 결정한 거리 선호도를 이용하여 기저범주 대응성을 산출하기 위한 참조정보 재분류 수단; 및

상기 범주표제 재분류 수단에서 산출한 범주표제 대응성 및 상기 참조정보 재분류 수단에서 산출한 기저범주 대응성을 결합시켜 표준 선호도를 산출하고, 산출한 표준 선호도를 이용하여 상기 기구축된 온톨로지의 범주를 재분류하기 위한 통합 재분류 수단

을 포함하는 온톨로지 재사용을 위한 범주 재분류 장치.

청구항 2.

제 1 항에 있어서,

상기 기구축된 온톨로지는,

시소러스의 원형을 기계가 사용할 수 있는 형태로 전처리하여 저장하고 있는 어휘사전인

온톨로지 재사용을 위한 범주 재분류 장치.

청구항 3.

제 1 항 또는 제 2 항에 있어서,

상기 범주표제 재분류 수단은,

범주표제를 술어형태로 변환하고, 상기 술어형태로 변환한 범주표제와 재분류 범주 간의 어휘 밀접도를 대응 선호도에 따라 계산하여 재분류 범주에 대한 기저범주를 결정하는

온톨로지 재사용을 위한 범주 재분류 장치.

청구항 4.

제 3 항에 있어서,

상기 범주표제의 대응성은,

범주표제의 계층 가중치와 재분류 범주의 대응 선호도를 곱한 수치인

온톨로지 재사용을 위한 범주 재분류 장치.

청구항 5.

제 1 항 또는 제 2 항에 있어서,

상기 기저범주 대응성은,

참조 정보의 거리 선호도와 재분류 범주의 대응 선호도를 곱한 수치인

온톨로지 재사용을 위한 범주 재분류 장치.

청구항 6.

제 1 항 또는 제 2 항에 있어서,

상기 통합 재분류 수단은,

상기 범주표제 재분류 수단에서 산출한 범주표제 대응성의 표준 선호도와 상기 참조정보 재분류 수단에서 산출한 기저범주 대응성의 표준 선호도의 평균을 범주 재분류를 위한 표준 선호도로서 이용하는

온톨로지 재사용을 위한 범주 재분류 장치.

청구항 7.

제 1 항 또는 제 2 항에 있어서,

상기 통합 재분류 수단은,

상기 범주표제 재분류 수단에서 산출한 범주표제 대응성의 표준 선호도와 상기 참조정보 재분류 수단에서 산출한 기저범주 대응성의 표준 선호도에 각각 가중치를 부여하여 산출한 가중 산출 평균을 범주 재분류를 위한 표준 선호도로서 이용하는

온톨로지 재사용을 위한 범주 재분류 장치.

청구항 8.

온톨로지를 재사용하기 위하여 범주를 재분류하는 방법으로서,

재분류 범주 간의 중요도를 나타내는 대응 선호도를 이용하여, 기구축된 온톨로지의 범주표제에 대한 범주표제 대응성을 산출하는 범주표제 대응성 산출 단계;

상기 산출한 범주표제 대응성을 이용하여 기저범주를 결정하는 기저범주 결정 단계;

상기 기구축된 온톨로지의 참조정보를 이용하여 상기 결정한 기저범주와 온톨로지 범주간의 반경을 측정하여 거리 선호도를 결정하는 거리 선호도 결정 단계;

상기 재분류 범주의 대응 선호도와 상기 결정한 거리 선호도를 이용하여 기저범주 대응성을 산출하는 기저범주 대응성 산출 단계;

상기 산출한 범주표제 대응성 및 기저범주 대응성을 결합시켜 표준 선호도를 산출하는 표준 선호도 산출 단계; 및

상기 산출한 표준 선호도를 이용하여 상기 기구축된 온톨로지의 범주를 재분류하는 범주 재분류 단계

를 포함하는 온톨로지 재사용을 위한 범주 재분류 방법.

청구항 9.

제 8 항에 있어서,

상기 기구축된 온톨로지는,

시소러스의 원형을 기계가 사용할 수 있는 형태로 전처리하여 저장하고 있는 어휘사전인

온톨로지 재사용을 위한 범주 재분류 방법.

청구항 10.

제 8 항 또는 제 9 항에 있어서,

상기 범주표제 대응성 산출 단계는,

상기 기구축된 온톨로지의 범주표제를 술어형태로 변환하고, 상기 술어형태로 변환한 범주표제와 재분류 범주 간의 어휘 밀접도를 대응 선호도에 따라 계산하여 재분류 범주에 대한 기저범주를 결정하는

온톨로지 재사용을 위한 범주 재분류 방법.

청구항 11.

제 10 항에 있어서,

상기 범주표제의 대응성은,

범주표제의 계층 가중치와 재분류 범주의 대응 선호도를 곱한 수치인

온톨로지 재사용을 위한 범주 재분류 방법.

청구항 12.

제 8 항 또는 제 9 항에 있어서,

상기 기저범주 대응성은,

참조 정보의 거리 선호도와 재분류 범주의 대응 선호도를 곱한 수치인

온톨로지 재사용을 위한 범주 재분류 방법.

청구항 13.

제 8 항 또는 제 9 항에 있어서,

상기 표준 선호도 산출 단계는,

상기 산출한 범주표제 대응성의 표준 선호도와 상기 산출한 기저범주 대응성의 표준 선호도의 평균을 온톨로지의 범주를 재분류하기 위한 표준 선호도로서 산출하는

온톨로지 재사용을 위한 범주 재분류 방법.

청구항 14.

제 8 항 또는 제 9 항에 있어서,

상기 표준 선호도 산출 단계는,

상기 산출한 범주표제 대응성의 표준 선호도와 상기 산출한 기저범주 대응성의 표준 선호도에 각각 가중치를 부여한 가중 산술 평균을 온톨로지의 범주를 재분류하기 위한 표준 선호도로서 산출하는

온톨로지 재사용을 위한 범주 재분류 장치.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 온톨로지 재사용을 위한 범주 재분류 장치 및 방법에 관한 것으로, 더욱 상세하게는 기구축된 온톨로지를 목적 하는 바에 맞도록 변환하여 재사용할 수 있도록 온톨로지 범주 정보를 단순화하거나 구체화시켜 재분류하는, 온톨로지 재 사용을 위한 범주 재분류 장치 및 방법에 관한 것이다.

컴퓨터와 인터넷 기술이 발전함에 따라, 개인이 접하는 문서의 양은 방대하게 증가하여, 개인이 스스로 처리할 수 있는 수준을 넘어서고 있다. 따라서, 문서 처리를 도와줄 자연어 처리 도구 또는 지식기반의 시스템이 필요하게 되었는데, 여기에 필수적인 구성 요소가 바로 온톨로지다.

온톨로지(Ontology)는 철학의 한 갈래로 존재론 또는 존재학이라고도 한다. 학문의 이름이 의미하는 바와 같이, 존재의 본질을 연구하는 형이상학이다. 인공지능 영역에서의 온톨로지는 인지체에 의하여 개별화된 실체(Entity)들에 대하여 논의나 분류를 가능하게 하는 범주(category) 시스템으로 본다.

온톨로지는 시맨틱 웹에서 정보의 내용을 파악하는 데에도 이용된다. 시맨틱 웹은 정보 형태에 기반한 기존의 웹과는 달리 정보의 내용을 중요시한다.

그런데, 시맨틱 웹에 존재하는 각 에이전트의 온톨로지는 서로 상이한 구조나 형태로 존재한다. 따라서, 시맨틱 웹에서 에이전트들이 검색 또는 교환한 정보를 이해하기 위해서는 검색된 결과를 자신의 온톨로지와 정련(Alignment)시키는 과정이 필요하다.

일반적으로 온톨로지 정련(Alignment) 과정은 온톨로지 간의 상이한 구조나 형태를 단순화시키거나 구체화시키는 재편성 과정이 수반된다. 이후, 온톨로지 병합(Ontologies Merging) 또는 온톨로지 집적(Ontologies Integration)과 같은 온톨로지 재사용 과정이 이루어지게 된다.

온톨로지 병합(Ontologies Merging)은 주제가 같거나 비슷한 온톨로지들을 단일화하는 방법으로서, 원 온톨로지의 각 부분이 재사용된 후에는 서로 구분할 수 없다.

이에 비하여 온톨로지 집적(Ontologies Integration)은 다른 온톨로지들을 수집하거나 수정해서 하나로 만드는 방법으로, 원 온톨로지의 각 부분은 재사용된 후에도 서로 구분할 수 있다.

한편, 최근 지능형 정보 시스템이 요구하는 분야 지식이 매우 다양해지면서 온톨로지의 필요성도 증대되고 있으나, 온톨로지를 개발하는 데에는 상당한 시간과 노력이 투입되기 때문에 구축된 온톨로지 자원이 부족한 실정이다.

따라서, 기구축되어 있는 온톨로지를 새로운 요구에 맞도록 가공하여 재사용할 수 있는 방안이 절실히 요구된다.

발명이 이루고자 하는 기술적 과제

본 발명은 상기와 같은 요구에 부응하기 위하여 창안된 것으로서, 기구축된 온톨로지를 목적하는 바에 맞도록 변환하여 재사용할 수 있도록 온톨로지 범주 정보를 단순화하거나 구체화시켜 재분류하는, 온톨로지 재사용을 위한 범주 재분류 장치 및 방법을 제공하는데 그 목적이 있다.

본 발명의 다른 목적 및 장점들은 하기에 설명될 것이며, 본 발명의 실시예에 의해 알게 될 것이다. 또한, 본 발명의 목적 및 장점들은 특허청구범위에 나타난 수단 및 조합에 의해 실현될 수 있다.

발명의 구성

상기와 같은 목적을 달성하기 위한 본 발명은 온톨로지를 재사용하기 위하여 범주를 재분류하는 장치로서, 재분류 범주의 대응 선호도를 이용하여 기구축된 온톨로지의 범주표제에 대한 범주표제 대응성을 산출하고, 기저범주를 결정하기 위한 범주표제 재분류 수단, 상기 기구축된 온톨로지의 참조정보를 이용하여 상기 범주표제 재분류 수단에서 결정한 기저범주와 온톨로지 범주간의 반경을 측정하여 거리 선호도를 결정한 후, 상기 재분류 범주의 대응 선호도와 상기 결정한 거리 선호도를 이용하여 기저범주 대응성을 산출하기 위한 참조정보 재분류 수단 및 상기 범주표제 재분류 수단에서 산출한 범주표제 대응성 및 상기 참조정보 재분류 수단에서 산출한 기저범주 대응성을 결합시켜 표준 선호도를 산출하고, 산출한 표준 선호도를 이용하여 상기 기구축된 온톨로지의 범주를 재분류하기 위한 통합 재분류 수단을 포함한다.

또한, 본 발명은 온톨로지를 재사용하기 위하여 범주를 재분류하는 방법으로서, 재분류 범주 간의 중요도를 나타내는 대응 선호도를 이용하여, 기구축된 온톨로지의 범주표제에 대한 범주표제 대응성을 산출하는 범주표제 대응성 산출 단계, 상기 산출한 범주표제 대응성을 이용하여 기저범주를 결정하는 기저범주 결정 단계, 상기 기구축된 온톨로지의 참조정보를 이용하여 상기 결정한 기저범주와 온톨로지 범주간의 반경을 측정하여 거리 선호도를 결정하는 거리 선호도 결정 단계, 상기 재분류 범주의 대응 선호도와 상기 결정한 거리 선호도를 이용하여 기저범주 대응성을 산출하는 기저범주 대응성 산출 단계, 상기 산출한 범주표제 대응성 및 기저범주 대응성을 결합시켜 표준 선호도를 산출하는 표준 선호도 산출 단계 및 상기 산출한 표준 선호도를 이용하여 상기 기구축된 온톨로지의 범주를 재분류하는 범주 재분류 단계를 포함한다.

이하 첨부된 도면을 참조로 본 발명의 바람직한 실시예를 상세히 설명하기로 한다. 이에 앞서, 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니 되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다.

따라서, 본 명세서에 기재된 실시 예와 도면에 도시된 구성은 본 발명의 가장 바람직한 일 실시 예에 불과할 뿐이고 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형예들이 있을 수 있음을 이해하여야 한다.

도 1은 본 발명의 일실시예에 따른 온톨로지 재사용을 위한 범주 재분류 장치의 구성도이다.

도 1에 도시된 바와 같이, 본 발명의 일실시예에 따른 온톨로지 재사용을 위한 범주 재분류 장치(100)는 재분류 범주 간의 중요도를 나타내는 대응 선호도를 이용하여, 기구축된 어휘사전(시소러스)의 범주표제에 대한 범주표제 대응성을 산출하고, 산출한 범주표제 대응성을 이용하여 기저범주를 결정하기 위한 범주표제 재분류부(11), 어휘사전(시소러스)의 참조정보를 이용하여 범주표제 재분류부(11)에서 결정한 기저범주와 어휘사전(시소러스) 범주간의 반경을 측정하여 거리 선호도를 결정한 후, 대응 선호도와 거리 선호도를 이용하여 기저범주 대응성을 산출하기 위한 참조정보 재분류부(12) 및 범주표

제 재분류부(11)에서 산출한 범주표제 대응성 및 참조정보 재분류부(12)에서 산출한 기저범주 대응성을 결합시켜 표준 선호도를 산출하고, 산출한 표준 선호도를 이용하여 어휘사전(시소러스)의 범주를 재분류하기 위한 통합 재분류부(13)를 포함한다.

본 발명에 따른 범주 재분류 장치(100)는 어휘사전(시소러스)의 범주표제와 참조정보를 활용하여 정보를 범주화하고, 그 범주를 단순화 또는 구체화한다. 이를 위하여, 어휘사전은 시소러스의 원형을 기계가 사용할 수 있는 형태로 전처리하여 저장하고 있다. 여기서, 어휘사전(시소러스)은 색인어 혹은 검색어를 선택하기 위한 색인 어휘의 집합으로서, 의미 분류체계를 갖춘 유의어 사전이다.

범주표제 재분류부(11)는 어휘사전의 범주표제를 이용하여 범주를 단순화한다. 즉, 범주표제 재분류부(11)는 도 4에 도시된 바와 같이, 범주표제를 술어형태로 변환하고(S401), 술어형태로 변환한 범주표제와 재분류 범주 간의 어휘 밀접도를 대응 선호도에 따라 계산하여(S402) 재분류 범주에 대한 기저범주를 결정한다(S403). 여기서, 대응 선호도는 재분류 범주 간의 중요도를 나타낸다.

범주표제 재분류부(11)에서 수행되는 범주 단순화 알고리즘은 다음과 같다.

//범주표제 변형 과정

repeat (모든 H_p 에 대하여)

if H_p 에 W_i 가 포함되어 있다 then

W_i 에 대응하는 C_0 를 T_c 에서 찾는다.

W_i 를 C_0 로 대체한다.

end if

$H_0 = W_i$ 가 C_0 로 치환된 H_p

end repeat

//범주표제와 단순화 범주의 대응 과정

repeat (모든 H_p 에 대하여)

repeat (H_0 의 모든 표제에 대하여)

$H_{co} = H_0$ 내의 C_0 개수

end repeat

$M_{ho} = H_{co} \times H_{wo}$

if (가장 큰 M_{ho} 가 하나 존재한다) then

repeat (M_{ho})

가장 큰 M_{ho} 를 찾는다.

end repeat

가장 큰 M_{ho} 를 갖는 C_r 을 C_o 에 대응시킨다.

else

repeat (동일 값의 M_{ho})

M_{ho} 간의 M_c 를 비교한다.

M_c 가 큰 C_o 를 찾는다.

end repeat

C_r 을 M_c 가 큰 C_o 에 대응시킨다.

end if

end repeat

여기서, C_o 는 단순화 범주를 나타내고, W_i 는 단순화 범주를 묘사하는 어휘를 나타내며, T_c 는 범주 묘사 어휘와 C_o 를 대응시킨 테이블을 나타낸다. 또한, C_r 은 시소러스의 범주, H_c 는 시소러스의 범주 표제, H_p 는 술어 형태로 변형시킨 범주표제, H_o 는 단순화 범주 형태로 변형된 범주표제를 각각 나타낸다. 또한, M_h 는 어휘사전의 범주표제에 대한 단순화 범주의 범주표제 대응성을 나타내고, H_w 는 범주표제 계층 가중치를 나타내며, M_c 는 대응 선호도를 나타낸다.

일반적으로 범주표제는 하위 계층으로 갈수록 범주 특유의 속성을 보다 명확하게 반영한다. 따라서, 단순화 범주로 변환된 범주표제에서, 범주표제에 대한 계층 가중치(Weight)는 최상위를 1이라 하고, 순차적으로 감소하면서 하위 계층에 대한 가중치를 부여한다. 위의 알고리즘에서 이러한 계층 가중치는 H_w 로 나타내고 있다.

또한, 단순화 범주 사이의 우선순위는 대응 선호도(M_c)로 나타내었다. 대응 선호도는 우선순위가 높을수록 큰 값을 갖고, 우선 순위가 낮을수록 작은 값을 갖는다. 여기서, 계층 가중치(H_w)와 대응 선호도(M_c)를 곱한 수치가 범주표제의 대응성 M_h 이다. 이 때, 단순화 범주를 결정하기 위해서는 어휘사전 범주에 대한 단순화 범주의 대응성을 정규화하는 것이 바람직할 것이다.

예를 들어, 계층 가중치 H_w 에 대한 대응 선호도 행렬을 H 라 할 때, 행렬 H 의 원소 a_{ij} 의 첨자 i 는 단순화 범주 D_o 의 각 범주이고, j 는 계층 가중치 H_w 이다. 즉, 범주표제의 계층 수준이다. 따라서, 원소 a_{ij} 는 단순화 범주의 대응 선호도 M_c 이다. 이 때, 각 단순화 범주의 범주표제 대응성 M_h 는 계층 가중치 H_w 와 대응 선호도 M_c 의 곱에 대한 합이라고 볼 수 있다. 최종적으로 범주표제 H_c 의 계층수준 개수를 1이라 하면, 단순화 범주 대응성 M_h 는 하기의 [수학식 1]과 같이 표현될 수 있다.

$$Mh_i = \sum_{j=1}^I a_{ij}$$

상기의 범주 단순화 알고리즘에 따르면, 단순화 범주 C_o 의 각 범주에 대한 범주표제 H_c 와의 어휘 밀접도를 기정의되어 있는 범주 변환표에 따라 조사한 후, 어휘사전의 범주 C_r 을 밀접도가 가장 높은 OfN(Ontology for Narrative) 범주 C_o 로 단순화시킨다. 이를 위하여, 도 2에 도시된 바와 같이, C_o 를 묘사하는 어휘를 참고로 H_c 의 형태를 변환한다. 그리고, C_o 를 묘사하는 어휘를 범주표제와 어휘사전 본문에서 검색한 후, C_r 과 C_o 의 대응성 M_h 를 구한다. 그리고, M_h 를 서로 비교하여 가장 큰 값을 갖는 단순화 범주를 결정한다.

여기서, 어휘사전의 범주 C_r 을 밀접도가 가장 높은 OfN(Ontology for Narrative) 범주 C_0 로 대응시키는 절차는 다음과 같다. 먼저, C_r 의 계층적 범주표제 H_c 의 OfN 대응성 M_h 를 구한다. 그리고, OfN 기저범주에서 C_r 까지의 참조경로 길이로서 C_r 의 OfN 대응성 M_p 를 계산한다. 그리고, M_h 와 M_p 의 가중평균치를 근거로 C_r 에 대응하는 OfN 범주 C_0 를 결정한다.

이러한 과정을 모든 어휘사전의 범주에 적용하면 어휘 사전의 범주를 단순화시킬 수 있다. 경우에 따라서는 단순화 범주 값이 동일할 수 있는데, 이 때에는 단순화 범주의 대응 선호도 M_c 를 따른다.

한편, 참조정보 재분류부(12)는 도 5에 도시된 바와 같이, 어휘사전의 참조정보를 이용하여 범주표제 재분류부(11)에서 결정한 기저범주와 어휘사전 범주간의 반경을 측정한 후(S501), 상기 측정한 반경을 이용하여 거리 선호도를 결정한다(S502). 그리고, 범주표제 재분류부(11)에서 계산한 대응 선호도와 상기 결정한 거리 선호도를 이용하여 기저범주 대응성을 계산한다(S503).

이처럼, 참조정보 재분류부(12)는 단순화 범주와의 어휘적 밀접도를 고려하여 결정된 기저범주를 참조 탐색의 출발점으로 이용한다. 초기의 기저범주 반경 값은 0으로 정한다. 이후, 참조정보를 이용하여 기저범주가 참조하는 범주들과 그 기저범주를 참조하는 범주들을 검색한다. 이 과정에서 검색된 범주의 반경 값은 1로 정한다. 같은 방식으로 계속 반경을 확장한다.

기저범주를 이용한 단순화의 경우에는 우선 단순화 대상인 복수의 정보 혹은 범주들에 대하여 기저범주를 결정한다. 그리고, 결정한 각 기저범주의 참조 반경을 확장한다. 각 참조 반경은 집합으로 가정하고 다수의 집합 사이의 공통요소를 취한다. 이 공통요소가 복수의 정보 혹은 범주에 대한 단순화 결과이다. 만약, 단순화 결과가 복수인 경우에는 도 3에 도시된 바와 같이, 순환적 관계를 보이지 않는 범위에서 그 결과를 대상으로 다시 단순화시킨다.

참조정보 재분류부(12)의 동작 과정을 좀 더 상세히 살펴보면, 앞에서 살펴본 범주표제를 이용한 범주 단순화의 경우와 마찬가지로, 기저범주에 대한 거리 선호도를 설정할 수 있다.

거리 선호도는 기저범주와 단순화 범주 C_r 과의 거리 L_p 에 영향을 받는다. 기저범주의 반경은 0이고 이 기저범주와 가까운 C_r 일수록 작은 반경 값을 갖는다. 그러므로, 거리 선호도는 반경이 작을수록 크게 적용되어야 할 것이다. 반경을 7까지 확장하면 모든 어휘사전의 범주를 포함할 수 있으므로 반경 값에 대한 8의 보수가 기저범주의 거리 선호도($L_c = 8 - L_p$)이다. 기저범주에 대한 단순화 범주의 대응성 M_p 는 단순화 범주에 지정한 대응 선호도 M_c 와 거리 선호도 L_c 의 곱이다.

여기서, 참조정보의 거리 선호도를 L_c 이라 하고 대응 선호도를 M_c 라 하면 단순화 범주 C_0 의 각 범주 i 에 대한 기저범주의 대응성 M_p 는 하기의 [수학식 2]로 표현할 수 있다.

$$M_{p_i} = L_{c_i} \times M_{c_i}$$

각 어휘사전의 범주 C_r 에 대하여 각 단순화 범주 C_0 와의 대응성 M_p 를 구한 후, 가장 큰 값을 가지는 범주를 단순화 범주로 결정한다. 이 과정을 모든 어휘사전 범주에 적용하면 어휘사전 범주를 단순화시킬 수 있다.

이처럼, 범주표제 재분류부(11) 및 참조정보 재분류부(12)에서 수행하는 범주표제 재분류 방법과 참조정보 재분류 방법은 각각 장단점을 가지고 있다. 따라서, 한 가지 범주 재분류 방법보다는 각각의 장점을 취하는 것이 바람직하다. 본 발명에서는 범주표제와 기저범주의 대응성을 수치적으로 결합하는 방식을 제안한다.

즉, 통합 재분류부(13)는 범주표제 재분류부(11) 및 참조정보 재분류부(12)에서 산출한 범주표제 대응성과 기저범주 대응성을 결합시키기 위하여 표준 선호도를 산출하여, 기구축된 어휘사전의 범주를 재분류한다.

여기서, 범주표제의 계층 가중치는 범주표제의 계층수준이다. 계층 가중치를 H_w 라 하면 범주표제 대응성 M_h 의 표준 선호도는 M_{h_u} 는 하기의 [수학식 3]에 의하여 구할 수 있다.

$$M_{h_{ui}} = \frac{M_{h_i}}{\text{계층가중치}(H_w)\text{의 합}} = \frac{M_{h_i}}{\text{등차수열 } H_w\text{의 합}}$$

마찬가지로, 기저범주 대응성 Mp 의 표준 선호도는 하기의 [수학식 4]에 의하여 구할 수 있다.

$$Mp_s = \frac{Mp_i}{\text{거리선호도}(Lc)\text{의 합}} = \frac{Mp_i}{\text{등차수열 } Lc\text{의 합}}$$

최종적으로 범주 재분류에 사용할 값은 범주표제의 표준 선호도와 기저범주의 표준 선호도의 평균으로서, 하기의 [수학식 5]에 의하여 구할 수 있다.

$$S_i = \frac{Mh_s + Mp_s}{2}$$

경우에 따라서는 범주표제의 대응성과 기저범주 대응성 중 어느 한쪽에 더 큰 비중을 줄 수 있을 것이다. 이 경우에는 각 표준 선호도에 범주표제와 기저범주의 중요도를 고려하여 가중 산술 평균을 구한다. 예를 들어, 범주표제 대응성에 가중치 2를 설정하고 기저범주 대응성에 가중치 1을 설정한다면, 범주표제 대응성과 기저범주 대응성의 가중 산술 평균은 하기의 [수학식 6]과 같다.

$$S_i = \frac{Mh_s \times 2 + Mp_s \times 1}{3}$$

여기서, 각 i 에 대한 S_i 의 값을 구한 후 가장 큰 S_i 를 보이는 범주를 해당 어휘사전 범주의 재분류 범주로 결정한다.

실제로, 이야기용 문장을 추상화시키는데 사용할 목적으로 일곱 가지 범주로 구성된 온톨로지를 재구성하는 실험을 해보았다. 실험에서 온톨로지를 재구성하기 위하여 기구축되어 있는 어휘사전(로젯 시소러스)의 범주를 단순화하는 방법을 이용하였다. 이 온톨로지의 범주는 다음과 같은 7가지 유형이 포함되어 있다.

- (1) 등장인물
- (2) 심상
- (3) 사건
- (4) 상태
- (5) 공간
- (6) 시간
- (7) 담화 표지

먼저, 온톨로지를 재분류하기 위하여 로젯 시소러스 범주표제 대응값과 기저범주의 대응값을 산출하였다. 그리고, 산출한 각 범주 값의 가중치 평균을 근거로 재분류 범주를 결정하였다. 각각의 범주 값을 병합하는 과정에서는 범주표제와 기저범주의 중요도를 조절하기 위해 가중치를 적용하였다. 이러한 과정을 통하여 설화문장의 추상화를 위한 로젯 시소러스 단순화 재분류 결과를 얻었다.

또한, 인지 및 감성을 고려한 22가지 흡인요소를 로젯 시소러스의 범주정보를 바탕으로 22가지 범주로 분류하였다. 이 범주를 바탕으로 기저범주를 탐색하여 가까운 반경의 범주를 병합하였다. 그리고, 22가지 범주를 구체화시켜서 207가지 범주들을 찾을 수 있었다. 이 범주들은 22가지 흡인요소와는 다른 세분화된 흡인요소들이다. 이 요소는 개별적이고 독립적인 흡인요소로도 존재할 수도 있으며, 여러 가지 요소가 함께 할 때 더 강한 흡인력으로 작용하기도 한다.

또한, 유식학의 선심소와 게임 흡인요소간의 대응관계를 범주 구체화를 응용하여 밝혀 보았다. 우선 게임의 22가지 흡인요소와 11종의 선심소에 대한 범주 재분류를 시행하였다. 이 범주 집합들의 반경을 각기 달리하여 3가지 경우로 나누어 대응시켜 보았다. 그 결과 흡인요소 범주의 반경이 0, 1, 2이고 선심소 범주의 반경이 0, 1, 2일때 두 범주간의 반경이 적절하게 유지되어 가장 만족할 만한 결과를 얻을 수 있었다.

이와같이, 본 발명에서는 자연어 처리 분야와 지식기반 시스템에서 요구하는 온톨로지 자원의 개발 방안으로서, 범주정보 단순화와 구체화를 통한 범주 재분류에 의하여 기존 자원을 재사용하는 방법을 제안하였다.

본 발명에 따른 범주 단순화 방법은 복수의 정보 또는 범주에 대한 대표성을 발견할 수 있도록 하며, 개별적인 정보 혹은 범주집합을 하나의 범주로 통합할 수 있도록 한다.

또한, 본 발명에 따른 범주 구체화 방법은 하나 혹은 그 이상의 범주를 의미와 구조는 유지시키면서 원하는 목적에 맞게 분할할 수 있도록 한다. 이를 통하여 범주를 구성하는 새로운 요소들을 발견할 수 있다.

또한, 본 발명에 따른 범주 단순화와 범주 구체화를 응용하여 서로 다른 범주집합들 사이의 관계를 밝혀낼 수 있다. 이는 상이한 온톨로지 범주집합 사이의 관계를 기존 온톨로지 범주정보를 활용하여 명시화할 수 있음을 시사하고, 서로 다른 온톨로지를 통합하는데 유용함을 나타낸다.

이상과 같이, 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 이것에 의해 한정되지 않으며 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 본 발명의 기술 사상과 아래에 기재될 특허 청구범위의 균등 범위 내에서 다양한 수정 및 변형이 가능함은 물론이다.

발명의 효과

상술한 바와 같이 본 발명은, 기구축된 온톨로지의 범주를 재분류함으로써, 온톨로지의 재사용성을 높여 시간과 노력을 절약하고, 온톨로지 범주 사이의 관계를 밝힐 수 있는 효과가 있다.

도면의 간단한 설명

도 1은 본 발명의 일실시예에 따른 온톨로지 재사용을 위한 범주 재분류 장치의 구성도,

도 2는 도 1의 범주 재분류 장치의 범주표제 변환 과정에 대한 설명도,

도 3은 도 1의 범주 재분류 장치의 기저범주를 이용한 범주 재분류 과정에 대한 설명도,

도 4는 본 발명의 일실시예에 따른 온톨로지 재사용을 위한 범주 재분류 방법 중 범주표제를 이용한 범주 재분류 과정에 대한 흐름도,

도 5는 본 발명의 일실시예에 따른 온톨로지 재사용을 위한 범주 재분류 방법 중 참조정보를 이용한 범주 재분류 과정에 대한 흐름도이다.

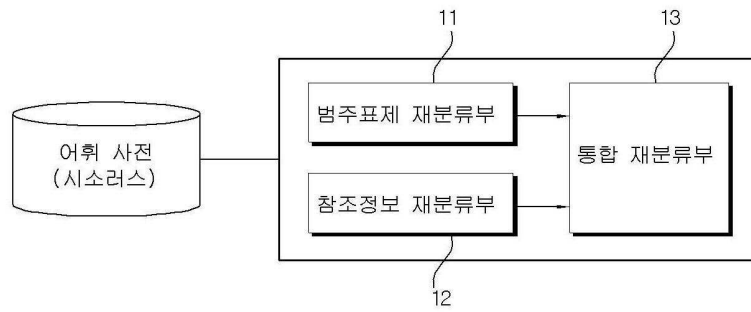
< 도면의 주요부분에 대한 부호의 설명 >

100 : 범주 재분류 장치 11 : 범주표제 재분류부

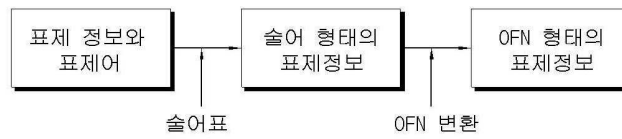
12 : 참조정보 재분류부 13 : 통합 재분류부

도면

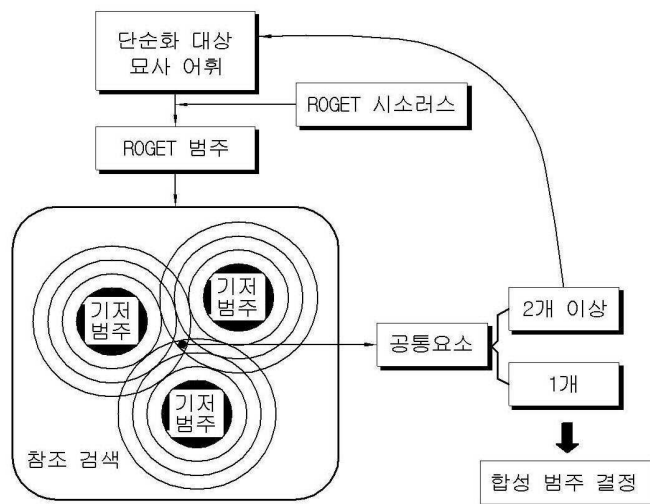
도면1



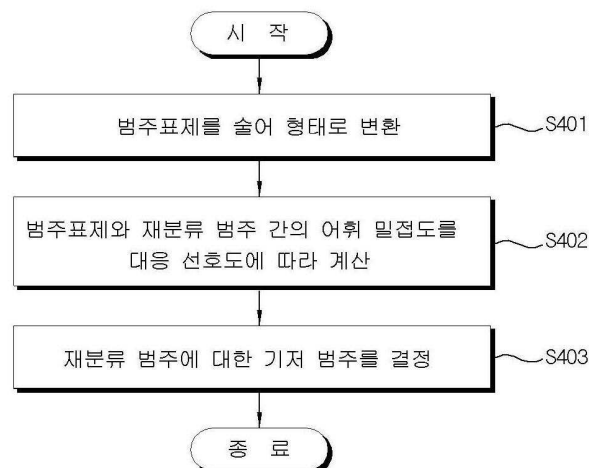
도면2



도면3



도면4



도면5

