

동사 정보를 활용한 의미 관계 추출을 위한 패턴 구축

김세종⁰ 이용훈 이종혁
포항공과대학교 전자컴퓨터공학부 컴퓨터공학과
{sejong⁰, yhlee95, jhlee}@postech.ac.kr

Pattern Construction for Semantic Relation Extraction using Verb Information

Se-Jong Kim⁰ Yong-Hun Lee Jong-Hyeok Lee
Dept. of Computer Science and Engineering, Division of Electrical
and Computer Engineering, POSTECH

요 약

온톨로지란 실세계에 존재하는 사물 및 개념, 그리고 용어들 간의 관계들을 컴퓨터가 이해할 수 있는 형태로 표현한 것이다. 온톨로지 구축에 있어서 대용량 코퍼스의 활용은 해당 코퍼스에서 등장하는 용어들과 이들 사이에서 나타나는 문자열을 일종의 패턴으로 취급하여 특정 패턴과 함께 나타나는 용어 쌍들을 해당 패턴이 대표하는 의미 관계로 설정하는 방식을 취한다. 그러나 기존의 방법은 주로 두 용어들 사이에서 나타나는 문자열만을 고려하여 패턴을 추출하기 때문에 해당 문장에 포함된 보다 다양한 문장 정보들을 활용할 수 없다. 본 논문은 이러한 한계점을 감안하여, 용어 쌍 사이에서 나타나는 문자열과 주변 동사 정보를 함께 고려함으로써 패턴의 정교성을 향상시키는 방법을 제안한다. 또한 동사들의 동의어를 활용하여 다양한 용어들을 포괄할 수 있는 일반화된 패턴을 구축한다. 본 방법론은 is-a 관계의 경우 64%, part-of 관계의 경우 83%, made-of 관계의 경우 73%, use 관계의 경우 72%의 정확률을 보였으며 모두 기존 방법보다 향상된 결과를 가져왔다.

1. 서론

온톨로지란 실세계에 존재하는 사물 및 개념, 그리고 용어들 간의 관계들을 컴퓨터가 이해할 수 있는 형태로 표현한 것이다. 온톨로지는 다양한 분야에서 활용될 수 있으며 시맨틱웹이나 지능화된 검색서비스 등이 대표적인 활용 예라고 할 수 있겠다. 온톨로지의 구축 방법은 크게 두 가지로 분류되며 하나는 기존 온톨로지 및 시소러스를 통합하여 큰 규모의 통합 온톨로지를 구축하는 방법이고, 다른 하나는 대용량 코퍼스를 분석하여 자동/반자동으로 온톨로지를 구축하는 방법이다. 특히 대용량 코퍼스를 사용하여 온톨로지를 구축하는 방법은 기존의 온톨로지 자원들을 결합하는 것이 아닌 새로운 온톨로지 자원을 개발 및 확보할 수 있다는 장점을 가지고 있다[1,2].

온톨로지 구축에 있어서 대용량 코퍼스의 활용은 해당 코퍼스에서 등장하는 용어들과 이들 사이에서 나타나는 문자열을 일종의 패턴으로 취급하여 특정 패턴과 함께 나타나는 용어 쌍들을 해당 패턴이 대표하는 의미 관계로 설정하는 것으로부터 시작한다[3]. 예를 들어 설명하면, 대상 문장에 “*Y such as of X*” 라는 패턴

을 만족하는 용어 X, Y 가 존재할 경우, 본 용어들은 “*is-a*” 관계를 가진다. 그러나 기존의 패턴 기반 의미 관계 추출 방법에서 사용하는 패턴들은 주로 두 용어들 사이에서 나타나는 문자열만을 고려하여 패턴을 추출하기 때문에 해당 문장에 포함된 보다 다양한 문장 정보들을 활용할 수 없다. 즉 패턴에 포함되어 있지 않은 주변 동사와 용어들 간의 의존도, 용어 자체의 의미 제약 정보, “*X of Y*” 와 같은 다양한 의미 관계가 나타날 수 있는 패턴들을 재분류할 수 있는 주변 문자열 정보 등을 활용할 수 없는 것이다. 물론 일부 연구에서는 이러한 정보들을 부분적으로 고려하여 수작업으로 작성한 패턴들을 사용한 바 있으나 본 논문에서는 동사 정보를 활용한 자동화된 방법을 제안하고자 한다. 본 방법은 용어들 사이에서 나타나는 문자열과 주변 동사 정보를 함께 고려하여 패턴의 정교성을 향상시켰기 때문에 보다 향상된 정확률(precision)을 기대할 수 있다. 또한 패턴에 포함된 동사들의 동의어를 사용하여 패턴의 일반화를 수행하였고 이를 통해 다양한 용어들을 포괄할 수 있는 패턴들을 구축할 수 있도록 하였다.

본 논문의 2장에서는 패턴 기반 의미 관계 추출에 관한 기존 연구들을 살펴보고 3장에서는 제안하고자 하는

방법을 자세히 설명한다. 4장에서는 영문 위키피디아(Wikipedia)에서 추출한 문장들로 구성된 코퍼스를 사용하여 실험 결과를 제시하고 5장에서 결론을 맺는다.

2. 관련연구

Hearst는 수작업으로 작성한 구문 패턴을 사용하여 *is-a* 관계의 하나인 상·하위어 관계를 추출하는 방법을 제안하였다[4]. 그는 의미 관계를 추출하기 위한 패턴들을 학습하는 Bootstrapping 알고리즘을 함께 소개하였는데 본 알고리즘은 이후 대부분의 패턴 기반 의미 관계 추출 방법론에 적용되었다. 그림 1은 Hearst의 Bootstrapping 알고리즘을 간략히 소개한 것이다.

- 1 단계. 추출하고자 하는 의미 관계를 정의한다.
- 2 단계. 정의한 의미 관계를 나타낼 수 있는 용어 쌍들을 수집한다.
- 3 단계. 코퍼스에서 수집한 용어 쌍들이 함께 나타나는 표현들을 추출한다.
- 4 단계. 추출한 표현들의 공통점을 찾아 새로운 패턴을 만든다.
- 5 단계. 새 패턴을 사용하여 새로운 용어 쌍들을 추출하고 2 단계부터 반복한다.

그림 1. Hearst의 Bootstrapping 알고리즘

Berland와 Charniak는 “*part-of*” 관계를 가진 용어들을 추출하는 시스템을 개발하였고[5], Girju는 워드넷(WordNet) 및 기계학습 방법을 활용하여 기존의 *part-of* 관계 추출 방법을 향상시켰다[6]. Ravichandran과 Hovy는 의미 관계를 대표할 수 있는 적은 양의 용어 쌍(seed)들을 사용하여 다양한 의미 관계들을 추출하였는데 *is-a* 및 *part-of* 관계를 제외한 특정 의미 관계에 대해서만 높은 정확률을 보였다[7]. Pennacchiotti와 Pantel은 Espresso라는 Bootstrapping 알고리즘을 개발하여 *is-a* 와 *part-of* 관계 및 기타 관계들에 대해서도 높은 정확률을 가져왔는데[3], 본 방법론은 Yang과 Su에 의해서 대응어를 찾아내기 위한 방법으로도 활용되었다[8].

그림 2는 Espresso 알고리즘을 단계별로 나타낸 것으로서, p 는 패턴을, $r_\pi(p)$ 는 해당 패턴의 신뢰도를, i 는 용어 쌍을, I '는 입력 용어 쌍들의 집합을, P '는 반복 수행 과정 중에서 누적된 최적의 패턴들의 집합을, $pmi(i, p)$ 는 해당 용어 쌍과 패턴 간의 연관성(pointwise mutual information)을, \max_{pmi} 는 최대 pmi 를, $r_i(i)$ 는 해당 용어 쌍의 신뢰도를 나타낸다.

pmi 는 식 3과 같이 정의되며 $|x, p, y|$ 는 코퍼스에서 용어 쌍 x, y 와 패턴 p 가 함께 나타난 빈도수를, $|x,$

$y|$ 는 패턴 p 와 상관없이 용어 쌍 x, y 가 나타난 빈도수를, $|*, p, *|$ 는 용어 쌍 x, y 와 상관없이 패턴 p 가 나타난 빈도수를 말한다.

- 1 단계. 추출하고자 하는 의미 관계에 대한 seed를 수작업으로 작성한다.
- 2 단계. 품사 태깅이 완료된 코퍼스로부터 해당 seed가 나타난 문장들을 모두 추출한다.
- 3 단계. 해당 문장에서 seed 이외의 용어들을 모두 단일 레이블(TR)로 치환하는 일반화 작업을 거친 후에 각 seed에 포함된 용어들 사이에서 나타나는 문자열을 패턴으로서 추출한다.
- 4 단계. 추출된 모든 패턴은 식 1을 사용하여 신뢰도(reliability)를 측정한다. (초기 $r_i(i)$ 는 모두 1이다.)

$$r_\pi(p) = \frac{\sum_{i \in I'} \left(\frac{pmi(i, p)}{\max_{pmi}} \times r_i(i) \right)}{|I'|}$$

식 1. 패턴의 신뢰도 측정

- 5 단계. $r_\pi(p)$ 에 의해 선택된 높은 신뢰도를 가진 최적의 패턴을 사용하여 코퍼스로부터 용어 쌍들을 추출한 후, 식 2를 통해 신뢰도가 높은 용어 쌍들을 선택한다.

$$r_i(i) = \frac{\sum_{p \in P'} \left(\frac{pmi(i, p)}{\max_{pmi}} \times r_\pi(p) \right)}{|P'|}$$

식 2. 용어 쌍의 신뢰도 측정

- 6 단계. 선택된 용어 쌍들은 다시 새로운 패턴을 추출하기 위해 본 알고리즘의 처음 단계의 입력 용어 쌍으로서 사용한다.

그림 2. Espresso 알고리즘

$$pmi(i, p) = \log \frac{|x, p, y|}{|x, *, y| \times |*, p, *|}$$

식 3. 용어 쌍과 패턴 간의 연관성 측정

$$df(i, p) = \frac{|x, p, y|}{|x, p, y| + 1} \times \frac{\min(|x, *, y|, |*, p, *|)}{\min(|x, *, y|, |*, p, *|) + 1}$$

식 4. 감소 인수

추가적으로, Espresso는 *pmi*의 저빈도 쓸림현상을 방지하기 위해 식 4와 같은 감소 인수(discounting factor)를 *pmi*에 곱한다[9].

3. 제안하는 방법

3.1 동사를 포함한 패턴 구축

예 1. He <i>makes X using Y</i> . ⇒ <i>X using Y</i> : use 관계 ⇒ <i>makes X using Y</i> : made-of 관계
예 2. That is <i>Y which X uses</i> . ⇒ <i>Y which X</i> : is-a 관계 ⇒ <i>Y which X uses</i> : use 관계
예 3. I <i>eat X of Y cooked</i> by him. ⇒ <i>X of Y</i> : 소유 관계 ⇒ <i>eat X of Y cooked</i> : part-of 관계

그림 3. 패턴에 따른 의미 관계 예시

그림 3은 동일한 문장에 대해서 패턴으로 인식할 문자열의 범위에 따라 서로 다른 의미 관계를 추출할 수 있음을 나타낸 것이다. 예 1은 용어 쌍 *X, Y*에 나타나는 문자열만을 패턴으로서 고려할 경우 해당 용어 쌍을 “use” 관계로 판단하는 반면, *X* 이전에 나타난 동사 *make*를 함께 고려할 경우 “made-of”라는 올바른 관계를 추출할 수 있음을 보여준다. 예 2는 *X* 이후에 나타나는 동사를 고려함으로써 *use* 관계를 추출하고, 예 3은 *X* 이전과 *Y* 이후에 등장하는 각각의 동사를 함께 고려함으로써 *part-of* 관계를 올바르게 추출하고 있다.

일반적으로 한 문장에는 하나 이상의 동사가 포함되며 각 동사는 해당 문장의 주 의미를 결정함과 동시에 주변 명사구, 즉 용어의 의미를 제약하거나 용어 간의 관계를 결정하는 중요한 단서가 된다. 본 논문에서는 동사의 이러한 역할에 초점을 맞춰 기존 패턴의 범위를 용어 쌍의 이전.이후에 나타나는 동사도 함께 고려할 수 있도록 확장하였다. 추출할 패턴들은 용어 쌍 사이, 용어 쌍 이전과 이후 중 최소한 한 곳에 동사를 포함하고 있어야 하며 해당 동사의 위치 정보를 유지해야 한다.

표 1. 동사를 포함하는 패턴

의미관계	패턴
<i>is-a</i>	(SE) X Y (look/vbp)
<i>part-of</i>	(make/vbd) X of/in his/prp\$ Y (EE)
<i>made-of</i>	(take/vbd) Y X (give/vbd)
<i>use</i>	(SE) Y take/vbd X (EE)

표 1은 제안한 방법을 적용하여 추출한 확장된 패턴들을 일부 기록한 것이다. ‘(SE)’와 ‘(EE)’는

각각 용어 쌍 이전에 동사가 존재하지 않을 경우, 용어 쌍 이후에 동사가 존재하지 않을 경우를 뜻한다.

용어 쌍 이전.이후에 나타나는 동사 이외의 문자열은 정교한 패턴을 구축하는데 유용하게 사용될 수 있으나 이를 함께 적용하면 패턴의 다양성이 극도로 심화되어 다양한 용어들을 포괄할 수 있는 패턴들을 추출할 수 없으므로 본 방법에서는 고려하지 않는다.

3.2 동사의 동의어를 활용한 패턴 일반화

앞에서도 언급했듯이 패턴을 구성하는 정보가 복잡해지면 찾고자하는 의미 관계를 만족하는 용어 쌍들을 보다 정확하게 추출할 수 있으나 그 패턴의 빈도수가 감소하여 다양한 용어들을 추출할 수 없게 된다. 이러한 한계점을 개선하기 위해 본 논문에서는 용어 쌍 이전.이후에 나타나는 동사만을 고려하는 것 이외에도, 각 패턴에 포함된 동사들의 동의어를 워드넷으로부터 추출하여 유사 패턴들을 하나로 묶는 작업을 수행하였다. 즉 동의어로 판단된 동사에 단일 아이디(Synset ID)를 부여하여 더욱 일반화된 패턴을 구축하였다[10]. 표 2는 동사 동의어를 고려하여 일반화한 패턴들을 나타낸다.

표 2. 동사 동의어를 활용한 일반화된 패턴

의미관계	패턴
<i>is-a</i>	(SE) X Y (SID-02130524-V/vbp)
<i>part-of</i>	(SID-02560585-V/vbd) Y ./, in/in the/dt X (EE)
<i>made-of</i>	(SID-02599636-V/vbd) Y X (SID-02316868-V/vbd)
<i>use</i>	(SID-00149583-V/vb) X of/in the/dt Y (EE)

3.3 Espresso의 *pmi* 개선

본 논문은 패턴의 형태를 제외하고 Espresso의 알고리즘을 따른다. Espresso에서 제안한 식들 중 용어 쌍과 패턴 간의 연관성을 나타내는 *pmi*는 mutual information의 정의에 의해 음수가 될 수 없다.

하지만 기존의 식은 *log* 안의 분모가 분자보다 항상 크거나 같기 때문에 위 정의를 만족할 수 없으므로 본 논문에서는 식 5와 같이 보완한 *pmi*를 제안한다. 보완된 *pmi*는 이후 비교 실험의 각각의 모델에 동일하게 적용한다.

$$\begin{aligned}
 pmi(i, p) &= \log \frac{P(x, p, y)}{P(x, y) \times P(p)} \\
 &\approx \log \frac{\frac{|x, p, y|}{|*, *, *|}}{\frac{|x, *, y|}{|*, *, *|} \times \frac{|*, p, *|}{|*, *, *|}} = \log \frac{|x, p, y| \times |*, *, *|}{|x, *, y| \times |*, p, *|}
 \end{aligned}$$

식 5. *pmi* 보완

4. 실험결과 및 분석

본 논문은 영문 위키피디아에서 추출한 총 947,625개의 문장들로 이루어진 코퍼스를 사용하며[11], 각 문장에는 스탠포드 태거를 사용한 품사 정보가 부착되어있다[12]. 용어로 인식할 명사들은 워드넷 3.0에 포함된 모든 명사를 대상으로 적용하였고 최장 명사구를 우선 선택하도록 하였다. 원활한 실험을 위해 모든 용어 쌍에 대한 패턴들을 코퍼스로부터 미리 추출하여 빈도수가 10이 넘는 패턴에 대해서만 실험을 수행하였다. 실험에서 평가한 의미관계는 *is-a*, *part-of*, *made-of*, *use* 관계로서, 각 관계에 대하여 Espresso 알고리즘을 적용한 방법과 본 논문에서 제시한 동사를 포함한 패턴을 활용한 방법(모델 A), 그리고 동사 동의어를 통해 일반화한 패턴을 추가적으로 적용한 방법(모델 B)을 비교 실험하였다.

표 3. 의미 관계별 seed 모음

의미관계	seed
<i>is-a</i>	(wheat, crop), (Miami, city), (shark, fish), (apple, fruit)
<i>part-of</i>	(cpu, computer), (drawer, desk), (roof, house), (hydrogen, water)
<i>made-of</i>	(bread, wheat), (radio, transistor), (bone, protein), (wine, grape)
<i>use</i>	(human, tool), (user, software), (consumer, product), (designer, brush)

표 3은 각 관계를 추출하기 위해 사용한 일부 seed들을 기록한 것으로서, 각 관계별로 10 개의 seed들을 수집하여 이를 통해 최적의 패턴을 2개 선택하고 본 패턴을 사용하여 신뢰도가 높은 10개의 용어 쌍을 추출한다. 이러한 과정을 총 10회 수행하면서 각 과정에서 추출한 용어 쌍들을 모아 그 정확률을 측정한다. 보다 정확한 성능 평가를 위해서는 신뢰도의 경계값을 설정하여 이를 만족하는 동안 각 과정을 반복 수행해야하나 해당 경계값을 설정하기 위한 평가 자원이 미구축되어 있고 반복 수행 자체에 소모되는 시간 및 자원이 비교적 크므로 본 횟수를 10회로 한정하기로 한다. 정확률은 수작업으로 판단하며 *is-a* 관계는 “*x is a y*”, *part-of* 관계는 “*x is a part of y*”, *made-of* 관계는 “*x is made of y*”, *use* 관계는 “*x uses y*”란 문장 구조에 용어 쌍을 대입하여 자연스러움의 유무에 따라 판단한다. 또한 각 과정에서 추출한 용어 쌍들은 해당 과정에서 새롭게 선택된 패턴을 통해서 획득한 것이므로 용어 쌍의 중복은 고려하지 않는다. 표 4와 표 5는 Espresso 알고리즘과 제안한 방법을 비교 실험한 결과이다.

실험 결과를 살펴보면, 모델 A의 경우 *is-a* 관계는 기존 방법보다 3% 향상된 64%의 정확률을, *part-of* 관

계는 9% 향상된 83%의 정확률을, *made-of* 관계는 35% 향상된 70%의 정확률을, *use* 관계는 7% 향상된 72%의 정확률을 보였다. 특히 *made-of* 관계에 대해서 큰 성능 향상을 보였는데 이는 기존 방법을 통해 추출한 패턴이 해당 관계를 표현하는데 충분한 정보를 포함하고 있지 못해서 지속적인 정확률 하락을 가져왔기 때문이라고 할 수 있겠다.

표 4. Espresso와 모델 A 비교 실험 결과(단위:%)

반복수행횟수	Espresso	모델 A	성능 향상	
<i>is-a</i> 관계	1	60.00	70.00	10.00
	2	60.00	75.00	15.00
	3	60.00	66.67	6.67
	4	60.00	67.50	7.50
	5	64.00	70.00	6.00
	6	65.00	68.33	3.33
	7	64.29	64.29	0.00
	8	60.00	62.50	2.50
	9	61.11	63.33	2.22
	10	61.00	64.00	3.00
<i>part-of</i> 관계	1	80.00	80.00	0.00
	2	80.00	80.00	0.00
	3	80.00	83.33	3.33
	4	75.00	80.00	5.00
	5	76.00	82.00	6.00
	6	75.00	83.33	8.33
	7	75.71	82.85	7.14
	8	75.00	82.50	7.50
	9	74.44	83.33	8.89
	10	74.00	83.00	9.00
<i>made-of</i> 관계	1	60.00	60.00	0.00
	2	60.00	70.00	10.00
	3	53.33	70.00	16.67
	4	45.00	65.00	20.00
	5	46.00	68.00	22.00
	6	46.67	68.33	21.66
	7	44.29	71.43	27.14
	8	42.50	72.50	30.00
	9	38.89	71.11	32.22
	10	35.00	70.00	35.00
<i>use</i> 관계	1	80.00	80.00	0.00
	2	80.00	80.00	0.00
	3	76.67	80.00	3.33
	4	75.00	80.00	5.00
	5	74.00	78.00	4.00
	6	71.67	73.33	1.66
	7	68.57	68.57	0.00
	8	66.25	70.00	3.75
	9	65.56	71.11	5.55
	10	65.00	72.00	7.00

모델 B는 *made-of* 관계와 *use* 관계에 대해서만 각각 38%와 5% 향상된 73%의 정확률과 70%의 정확률을 보였다. 동사의 동의어를 고려하여 패턴을 일반화할 경우 한정된 동사와 동사 이외의 정보에 비교적 의존적인

is-a 관계 및 *part-of* 관계는 기존 방법보다 성능이 향상되지 않거나 오히려 감소하는 결과를 가져왔다. 이러한 사실은 패턴을 정교하게 하거나 일반화하는 작업을 의미 관계의 종류에 따라 다르게 적용할 필요가 있음을 알려준다.

표 5. Espresso와 모델 B 비교 실험 결과(단위:%)

반복수행횟수	Espresso	모델 B	성능 향상	
<i>is-a</i> 관계	1	60.00	70.00	10.00
	2	60.00	70.00	10.00
	3	60.00	63.33	3.33
	4	60.00	67.50	7.50
	5	64.00	64.00	0.00
	6	65.00	63.33	-1.67
	7	64.29	60.00	-4.29
	8	60.00	58.75	-1.25
	9	61.11	58.89	-2.22
	10	61.00	61.00	0.00
<i>part-of</i> 관계	1	80.00	80.00	0.00
	2	80.00	80.00	0.00
	3	80.00	63.33	-16.67
	4	75.00	60.00	-15.00
	5	76.00	60.00	-16.00
	6	75.00	55.00	-20.00
	7	75.71	50.00	-25.71
	8	75.00	51.25	-23.75
	9	74.44	52.22	-22.22
	10	74.00	53.00	-21.00
<i>made-of</i> 관계	1	60.00	80.00	20.00
	2	60.00	80.00	20.00
	3	53.33	76.66	23.33
	4	45.00	70.00	25.00
	5	46.00	72.00	26.00
	6	46.67	73.33	26.66
	7	44.29	72.86	28.57
	8	42.50	72.50	30.00
	9	38.89	72.22	33.33
	10	35.00	73.00	38.00
<i>use</i> 관계	1	80.00	80.00	0.00
	2	80.00	75.00	-5.00
	3	76.67	76.67	0.00
	4	75.00	77.50	2.50
	5	74.00	76.00	2.00
	6	71.67	71.67	0.00
	7	68.57	67.14	-1.43
	8	66.25	68.75	2.50
	9	65.56	70.00	4.44
	10	65.00	70.00	5.00

각각의 의미 관계는 반복 수행을 거치면서 전반적으로 그 성능이 감소하게 되는데, 이러한 결과는 이전 수행을 통해 추출한 용어 쌍의 일부가 해당 관계를 만족하지 못할 경우 이후 반복 과정에서 유용하지 못한 패턴을 추출할 가능성이 높아지기 때문에 발생한 것으로 판단된다.

5. 결론

본 논문은 용어 쌍 사이에서 나타나는 문자열과 주변 동사 정보를 함께 고려하여 패턴의 정교성을 향상시키는 방법을 제안하였고, 각각의 의미 관계에 대해서 안정적인 성능 향상을 가져왔다. 또한 패턴에 포함된 동사들의 동의어 정보를 사용하여 다양한 용어들을 포괄할 수 있는 일반화된 패턴을 구축하고자 하였다.

하지만 일반화를 통해 발생한 정확률의 하락은 패턴의 정교함과 보편성에 대한 보다 깊은 연구가 요구됨을 시사해 주었다. 이후 연구에서는 의미 관계에 따라 패턴 구축의 기준을 차별화하고 이러한 패턴을 만족하는 용어 쌍들을 가능한 한 많이 추출하여 재현율(recall)에 대한 기여도를 함께 평가해보아야 할 것이다.

감사의 글

본 논문은 지식경제부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업과 2008년도 두뇌한국21사업의 지원을 받아 수행되었습니다.

참고 문헌

- [1] A. Maedche, S. Staab, "Measuring similarity between ontologies," EKAW, LNAI 2473, pp. 251-263, 2002.
- [2] M. Kavalec, V. Svatek, "A Study on Automated Relation Labelling in Ontology Learning," Ontology Learning and Population from Text: Methods, Evaluation and Applications, IOS Press, pp. 44-58, 2005.
- [3] P. Pantel, M. Pennacchiotti, "Espresso: Leveraging generic patterns for automatically harvesting semantic relations," Proceedings of ACL, pp. 113-120, 2006.
- [4] M.A. Hearst, "Automatic acquisition of hyponyms from large text corpora," Proceedings of the 14th conference on Computational linguistics, vol. 2, pp. 539-545, 1992.
- [5] M. Berland, E. Charniak, "Finding parts in very large corpora," Proceedings of ACL, pp. 57-64, 1999.
- [6] R. Girju, et al., "Learning semantic constraints for the automatic discovery of part-whole relations," Proceedings of HLT/NAACL, pp. 1-8, 2003.
- [7] D. Ravichandran, E. Hovy, "Learning surface text patterns for a question answering system," Proceedings of ACL, pp. 41-47, 2002.
- [8] X. Yang, J. Su, "Coreference resolution using semantic relatedness information from automatically discovered patterns," Proceedings of ACL, pp. 528-535, 2007.

- [9] P. Pantel, D. Ravichandran, "Automatically labeling semantic classes," Proceedings of HLT/NAACL, pp. 321-328, 2004.
- [10] WordNet 3.0, "<http://wordnet.princeton.edu>"
- [11] Wikipedia, "<http://www.wikipedia.org>"
- [12] Stanford tagger, "<http://nlp.stanford.edu/software/tagger.shtml>"

부록

부록 1. 표 2의 Synset ID에 속한 동사

Synset ID	동사
SID-00149583-V	become, go, get
SID-02130524-V	look
SID-02316868-V	give
SID-02560585-V	make, do
SID-02599636-V	take