용어를 공유하는 패턴 쌍을 이용한 의미 관계 추출

(Semantic Relation Extraction using Pattern Pairs Sharing a Term)

김 세 종 † 이 용 훈 ††
(Se-Jong Kim) (Yong-Hun Lee)

이 종 혁 †††
(Jong-Hyeok Lee)

요 약 대용량 코퍼스를 사용하여 온톨로지를 구축하는 것은 해당 코퍼스에서 등장하는 용어들과 이들 간의 의미 관계를 보다 자동화된 방법으로 추출하는 것으로부터 시작한다. 이때 주로 사용하는 방법이 용어들 사이에서 나타나는 문자열을 일종의 패턴으로 취급하여 특정 패턴과 함께 나타나는 용어들을 해당 패턴에 할당된 의미 관계로 설정하는 방법이다. 하지만 기존의 패턴 기반 의미 관계 추출 방법은 한 문장만을 대상으로 패턴을 추출 및 적용하기 때문에 서로 떨어진 용어에 대한 의미 관계를 추출할 수 없다는 단점을 가지고 있다. 본 논문은 이러한 한계점에 착안하여, 의미 관계를 대표하는 각각의 용어를 하나씩 포함하고 기타 용어를 공유하고 있는 서로 떨어진 패턴 쌍을 추출하여 확장된 패턴을 생성하고 이를 의미 관계 추출에 적용하였다. 본 방법론은 *is-a* 관계의 경우 기존 방법론 보다 7.5% 향상된 83.75%의 정확률을, *part-of* 관계의 경우에는 5% 향상된 동일한 83.75%의 정확률을 보였으며 상대적 재현율을 통해 실제 재현율의 향상 가능성도 함께 제시하였다.

키워드 : 의미관계추출, 패턴 쌍, 부트스트랩핑, 온톨로지

Abstract Constructing an ontology using a mass corpus begins with an automatic semantic relation extraction. A general method regards words appearing between terms as patterns which are used to extract semantic relations. However, previous approaches consider only one sentence to extract a pattern, so they cannot extract semantic relations for terms in different sentences. This paper proposes a semantic relation extraction method using pairs of patterns sharing a term, where each pattern is extracted using one of the seed term pair satisfying the target relation. In our experiments, we achieved the accuracy 83.75% improving previous methods by 7.5% in *is-a* relation and the accuracy 83.75% improved by 5% in *part-of* relation. We also present a possibility of improving the recall by the relative recall.

Key words : Semantic Relation Extraction, Pattern Pair, Bootstrapping, Ontology

1. 서 론

온톨로지란 실세계의 개념 및 사물들 간의 관계들을 체계적으로 정의하여 이러한 지식을 컴퓨터가 이해할 수 있는 형태로 표현한 것이다. 온톨로지를 구축하기 위한 방법론에 대한 연구는 1990년대 말부터 활발히 진행되고 있으며 이러한 방법론들은 크게 기존 온톨로지 및 시소러스를 통합하여 큰 규모의 통합 온톨로지를 구축하는 방법과, 대용량 코퍼스를 분석하여 자동/반자동으로 온톨로지를 구축하는 방법으로 나뉜다[1,2]. 첫 번째 방법은 온톨로지를 새롭게 구축하는 것이 아닌, 기존 자원을 활용하여 온톨로지를 확장하는데 초점을 맞추고 있는 반면, 두 번째 방법은 대용량 코퍼스를 통한 새로운 온톨로지를 구축하는데 목적을 두고 있기 때문에 새롭게 생성된 자원을 확보할 수 있다는 점에서 보다 유용한 방법으로 생각할 수 있을 것이다.

대용량 코퍼스를 사용하여 온톨로지를 구축하는 것은 해당 코퍼스에서 등장하는 용어들과 이들 간의 의미 관계를 보다 자동화된 방법으로 추출하는 것으로부터 시작한다. 이때 주로 사용하는 방법이 용어들 사이에서 나타나는 문자열을 일종의 패턴으로 취급하여 특정 패턴과 함께 나타나는 용어들을 해당 패턴에 할당된 의미 관계로 설정하는 방법이다[3]. 한 가지 예를 들어, 대상 문장에 "*Y consists of X*"라는 패턴을 만족하는 용어 *X*, *Y*가 존재할 경우, 본 용어들은 "*part-of*" 관계를 가진다. 하지만 기존의 패턴 기반 의미 관계 추출 방법은 한 문장만을 대상으로 패턴을 추출하고 적용하기 때문에 서로 떨어진 용어에 대한 의미 관계를 추출할 수 없다는 단점을 가지고 있다. 실제로 대용량 코퍼스에서 의미 관계를 대표할 수 있는 용어들의 쌍이 한 문장에 포

· 본 논문은 지식경제부 및 정보통신연구진흥원의 정보통신선도기반기술개발사업과 2009년도 두뇌한국21사업의 지원을 받아 수행되었습니다.
· 이 논문은 제35회 추계학술대회에서 '용어를 공유하는 패턴 쌍을 이용한 의미 관계 추출'의 제목으로 발표된 논문을 확장한 것임

† 학생회원 : 포항공과대학교 컴퓨터공학과
sejong@postech.ac.kr
†† 학생회원 : 포항공과대학교 컴퓨터공학과
yhlee95@postech.ac.kr
††† 종신회원 : 포항공과대학교 컴퓨터공학과 교수
jhlee@postech.ac.kr

논문접수 : 2008년 12월 18일
심사완료 : 2009년 2월 6일

정보과학회논문지: 컴퓨팅의 실제 및 레터 제15권 제3호(2009.3)

합되어 있는 경우는 매우 적었다. 본 논문은 이러한 한계점에 착안하여, 의미 관계를 대표하는 각각의 용어를 하나씩 포함하고 기타 용어를 공유하고 있는 서로 떨어진 문장들에 대한 패턴을 추출하여 확장된 패턴을 생성하고 이를 의미 관계 추출에 적용하였다.

본 논문의 2장에서는 패턴 기반의 의미 관계 추출에 관한 기존 연구들을 살펴보고 3장에서는 제안하고자 하는 방법에 대해 자세히 설명한다. 4장에서는 위키피디아(Wikipedia)에서 추출한 문장들로 이루어진 코퍼스를 사용하여 실험결과를 제시하고 5장에서 결론을 맺는다.

2. 관련 연구

의미 관계 추출에 있어서 가장 일반적이고 많이 연구된 방법론은 "*is-a*" 및 "*part-of*" 관계를 패턴 기반으로 추출하는 방법이다. 이러한 *is-a* 및 *part-of* 관계는 온톨로지 뿐만 아니라 시소러스의 구축에도 중요한 역할을 한다. Hearst는 수작업으로 작성한 구문 패턴을 사용하여 상·하위어 관계, 즉 *is-a* 관계를 추출하는 방법을 제안하였다[4]. 그는 추출한 결과를 바탕으로 보다 많은 구문 패턴들을 학습하는 부트스트랩핑 알고리즘을 소개하였는데 본 알고리즘은 이후 대부분의 패턴 기반 의미 관계 추출 방법론에 적용되었다. Berland와 Charniak는 *part-of* 관계를 가진 용어들을 추출하는 시스템을 개발하였고, Girju는 기계학습 방법 및 워드넷(WordNet)을 활용하여 기존의 *part-of* 관계 추출 방법을 향상시켰다[5,6]. Ravichandran과 Hovy는 의미 관계를 대표할 수 있는 적은 양의 용어 쌍(seed)들을 사용하여 질의 응답 시스템에서 나타날 수 있는 다양한 의미 관계들을 추출하였는데 특정 의미 관계에 대해서만 높은 정확률(precision)을 보였다[7]. Kim은 패턴의 일반화를 통해 의미 관계 추출 결과의 재현율(recall)을 향상시켰으나 정확률에 대해서는 큰 영향을 미치지 못했다[8]. Pantel과 Pennacchiotti는 Espresso라는 부트스트랩핑 알고리즘을 개발하여 기존의 모든 연구들 중에서 가장 높은 정확률을 가져왔다[3]. 본 방법론은 Yang과 Su에 의해서 대용어를 찾아내기 위한 방법으로도 활용되었다[9].

Espresso 알고리즘은 먼저 각각의 의미 관계에 대한 seed를 수작업으로 작성하고 품사 태깅이 완료된 코퍼스로부터 이들이 나타난 문장들을 모두 추출한다. 그리고 해당 문장에서 seed 이외의 용어들을 모두 단일 레이블(TR)로 치환하는 일반화 작업을 거친 후에 각 seed에 포함된 용어들 사이에서 나타나는 문자열을 패턴으로서 추출한다. 추출된 모든 패턴은 식 (1)을 사용하여 신뢰도(reliability)를 측정하는데, p는 패턴을, $r_\pi(p)$는 해당 패턴의 신뢰도를, i는 용어 쌍을, I'는 입력 용어 쌍들의 집합을, $pmi(i, p)$는 해당 용어 쌍과 패턴 간의 연관성(pointwise mutual information)을, max_{pmi}는 최대 pmi를, $r_i(i)$는 해당 용어 쌍의 신뢰도를 나타낸다.

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

식 (1) 패턴의 신뢰도 측정

$pmi(i, p)$는 식 (2)와 같이 정의되며 $|x, p, y|$는 코퍼스에서 용어 쌍 x, y와 패턴 p가 함께 나타난 빈도수를, $|x, *, y|$는 패턴 p와 상관없이 용어 쌍 x, y가 나타난 빈도수를, $|*, p, *|$는 용어 쌍 x, y와 상관없이 패턴 p가 나타난 빈도수를 말한다.

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

식 (2) 용어 쌍과 패턴 간의 연관성 측정

초기 $r_i(i)$는 모두 1이며 이후 $r_\pi(p)$에 의해 선택된 최적의 패턴을 사용하여 코퍼스로부터 용어 쌍들을 추출한 후 식 3을 통해 신뢰도가 높은 용어 쌍들을 선택한다. 선택된 용어 쌍들은 다시 새로운 패턴을 추출하기 위해 본 알고리즘의 처음 단계의 입력 용어 쌍으로서 사용한다. P'는 이러한 반복 수행 과정 중에서 누적된 최적의 패턴들의 집합을 나타낸다.

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

식 (3) 용어 쌍의 신뢰도 측정

3. 제안하는 방법

3.1 용어를 공유하는 패턴 쌍 활용

의미 관계의 종류는 해당 의미 관계를 사용하는 도메인에 따라 다양하게 정의될 수 있다. 물론 많은 사람들이 동의하는 기준에 따라 보다 정형화되고 일반화된 의미 관계를 정의하는 것이 바람직하겠지만 현실적으로 몇몇 의미 관계를 제외하고는 어려운 상황이다. 여기서 알 수 있는 사실은 기존에 정의된 의미 관계 이외에도 용어들 사이에는 다양한 의미 관계가 존재하며 이러한 의미 관계들을 활용하여 기존에 찾지 못했던 용어 쌍들을 발견할 수 있다면 보다 향상된 의미 관계 추출 알고리즘을 개발할 수 있다는 것이다.

그림 1은 한 문장 내에 *is-a* 나 *part-of* 관계가 존재하지 않는 문장들을 나열한 것이다. 하지만 예 1에는

예 1. a. A computer wastes electricity.
b. Machines use electricity.
예 2. a. Rain wet clothes.
b. Clothes cleaned by the water.

그림 1 *is-a* 관계가 잠재되어있는 문장 예시

is-a 관계를 가진 computer와 machine이라는 용어가 존재하고 예 2에도 역시 rain과 water라는 *is-a* 관계의 용어가 포함되어있다. 각각의 예를 자세히 살펴보면, 예 1은 a와 b문장에 electricity라는 용어를 공유하고 있는데 a문장은 computer가, b문장은 machine이 electricity와 "*use*" 관계를 가지고 있음을 알 수 있다. 즉 동일한 용어와 *use* 관계를 가진 용어들은 *is-a* 관계일 가능성이 있다는 것이다. 예 2는 a와 b문장에 clothes라는 용어를 공유하고 있고 a문장은 rain이 clothes와 "*wet*" 관계를, b문장은 water가 clothes와 "*clean*"관계를 가지고 있음을 알 수 있는데, 동일한 용어와 *wet* 또는 *clean* 관계를 가진 용어들도 *is-a* 관계일 가능성이 있다는 것을 보여주는 예라고 할 수 있겠다.

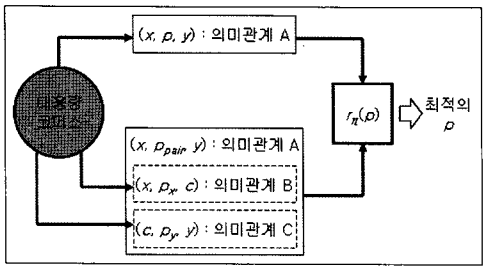

그림 2 용어(c)를 공유하는 패턴 쌍 활용

본 논문은 이렇게 두 개의 의미 관계를 결합하여 하나의 의미 관계를 추출할 수 있다고 가정한다. 그리고 이러한 가정을 적용하기 위한 한 가지 방법으로서 용어를 공유하는 패턴 쌍을 코퍼스로부터 추출하여 한 문장에서 추출한 패턴과 동일한 방법으로 그 신뢰도를 측정한다. 그림 2는 본 방법론을 시각적으로 표현한 것이다.

p_{pair}는 동일한 용어(c)를 공유하고 있는 패턴 p_x와 패턴 p_y, 그리고 각 패턴의 공유 용어에 대한 위치 정보를 포함한 확장된 패턴을 말한다. p_{pair}의 c는 공유하고 있는 용어의 위치 정보만을 나타낼 뿐, 공유 용어의 내용은 고려하지 않는다. 즉 p_{pair}의 종류는 패턴 p_x와 패턴 p_y의 내용 이외에도 c의 위치에 따라 4종류로 나누어질 수 있으며 본 논문은 이러한 경우를 모두 포함시켜 실험을 진행하였다.

앞에서도 언급했듯이 기존의 패턴 기반 의미 관계 추출 알고리즘은 한 문장 내에서 나타나는 패턴만을 사용

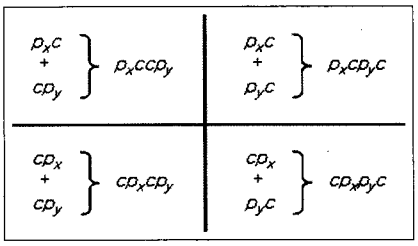

그림 3 c의 위치에 따른 p_{pair}의 종류

하여 용어 쌍들을 추출하였기 때문에 다양한 용어들을 추출할 수 없었던 반면, 본 방법론은 서로 떨어진 문장들 속에서 각각 나타난 용어들에 관해서도 이를 추출할 수 있는 패턴을 생성하여 적용하였기 때문에 반복 수행을 통한 정확률의 향상을 꾀할 수 있을 뿐만 아니라 특히 재현율 면에서 유용하게 사용할 수 있는 방법이라고 하겠다.

3.2 Espresso 알고리즘 적용

본 방법론을 Espresso 알고리즘에 적용하기 위해서는 제안한 패턴들을 처리할 수 있는 *pmi*를 설계해야한다. 식 (4)는 이러한 사항을 반영하여 기존의 *pmi*를 변경한 것이다.

$$pmi(i, p_{pair}) = \log \frac{|x, p_{pair}, y|}{|x, *, y| \times |*, p_{pair}, *|} = \log \frac{\sum_{c \in T} |x, p_x, c, p_y, y|}{\sum_{c' \in T} |x, *, c', *, y| \times \sum_{c'' \in T} |*, p_x, c'', p_y, *|}$$

식 (4) 용어 쌍과 제안한 패턴 간의 연관성 측정

T는 용어 쌍 x, y를 제외한 모든 용어들의 집합을 나타내며 c, c', c''는 각각의 빈도수를 측정할 때 사용하는 공유 용어들을 말한다. 패턴은 그 종류에 따라서 기존의 *pmi*와 변경한 *pmi*를 각각 다르게 적용하고 그 이외의 식들에 대해서는 동일하게 적용한다. 2장에서는 언급하지 않았지만, Espresso는 *pmi*의 저빈도 쏠림현상을 방지하기 위해 Pantel과 Ravichandran의 감소 인수(discounting factor)를 *pmi*에 곱한다[10]. 식 (5)는 제안한 패턴을 적용할 수 있도록 감소 인수를 변경한 것이다. 또한 기존의 *pmi*는 로그(log) 안의 값이 항상 1보다 작거나 같아서 상호 연관성을 나타내는데 한계가 있으므로 실제 실험에서는 이를 고려한 정형화를 수행한다. $r_i(i)$를 구할 때 사용하는 P'는 기존의 패턴과 제안한 패턴을 함께 적용한 최적의 패턴들의 집합을 나타내도록 한다.

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

$$= \frac{\sum_{c \in T} |x, p_x, c, p_y, y|}{\sum_{c \in T} |x, p_x, c, p_y, y| + 1} \times \frac{\min\left(\sum_{c \in T} |x, *, c', *, y|, \sum_{c' \in T} |*, p_x, c'', p_y, *|\right)}{\min\left(\sum_{c \in T} |x, *, c', *, y|, \sum_{c'' \in T} |*, p_x, c'', p_y, *|\right) + 1}$$

식 (5) 제안한 패턴을 적용한 감소 인수

4. 실험결과 및 분석

본 논문은 영문 위키피디아에서 추출한 문장들로 이루어진 코퍼스를 사용한다[11]. 이 코퍼스는 총 947,625개의 문장으로 구성되어있고 각 문장에는 스탠포드 태거를 사용한 품사 정보가 부착되어있다[12]. 용어로 인식할 명사들은 워드넷 3.0에 포함된 모든 명사를 대상으로 적용하였고 존재할 수 있는 모든 용어 쌍에 대한 패턴들을 코퍼스로부터 미리 추출하여 빈도수가 10이 넘는 패턴에 대해서만 실험을 수행하였다[13]. 최종적으로 정제된 용어의 개수는 32,553개이고 용어 쌍의 개수는 1,243,068개, 패턴의 종류는 25,289개이다. 실험 대상으로 삼은 의미 관계는 *is-a* 와 *part-of* 관계로서, 표 1은 각 관계를 추출하기 위해 사용한 seed들을 기록한 것이다.

실험은 Espresso 알고리즘을 사용한 방법과 본 논문에서 제안한 방법, 즉 용어를 공유하는 패턴 쌍을 함께 활용한 방법을 각각의 의미 관계에 대해 적용하였다. 10개의 seed로부터 최적의 패턴을 2개 선택하고 본 패턴을 사용하여 신뢰도가 높은 10개의 용어 쌍을 추출한다. 여기서 Espresso 알고리즘의 경우 $r_\pi(p)$값에 의해 상위 2개의 패턴을 선택하고, 제안한 방법의 경우 $r_\pi(p)$값을 그대로 사용하되 p와 p_{pair}형태의 패턴을 하나씩 선택하도록 한다. 만일 p_{pair}형태의 패턴이 해당 용어 쌍들을 처리하는 과정에서 발견되지 않으면 p형태의 패턴을 대신 선택한다. 이러한 과정을 총 8회 수행하면서 각 과정에서 추출한 용어 쌍들을 모아 그 정확률을 측정한다. 정확률은 수작업으로 판단하며 *is-a* 관계는 "*x is a y*" 란 문장 구조에 각 용어를 대입하고, *part-of* 관계는 "*x is a part of y*" 란 문장 구조에 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 2는 Espresso 알고리즘과 제안한 방법을 수행한 결과이다.

표 1 의미 관계별 seed 모음

의미관계	*is-a*	*part-of*
seed	wheat, crop	cpu, computer
	Miami, city	drawer, desk
	shark, fish	roof, house
	apple, fruit	hydrogen, water
	man, human	head, body
	milk, beverage	branch, tree
	flower, plant	wing, airplane
	computer, machine	sea, earth
	desk, table	player, team
	noise, sound	wheel, car

표 2 기존 방법과 제안 방법의 정확률 비교(단위:%)

반복수행횟수		Espresso	제안 방법	성능 향상
is-a 관계	1	80.00	100.00	20.00
	2	80.00	90.00	10.00
	3	73.33	86.67	13.34
	4	70.00	82.50	12.50
	5	72.00	82.00	10.00
	6	75.00	83.33	8.33
	7	74.29	85.71	11.42
	8	**76.25**	**83.75**	**7.50**
part-of 관계	1	80.00	70.00	-10.00
	2	85.00	80.00	-5.00
	3	80.00	83.33	3.33
	4	80.00	82.50	2.50
	5	78.00	84.00	6.00
	6	76.67	81.67	5.00
	7	78.57	82.86	4.29
	8	**78.75**	**83.75**	**5.00**

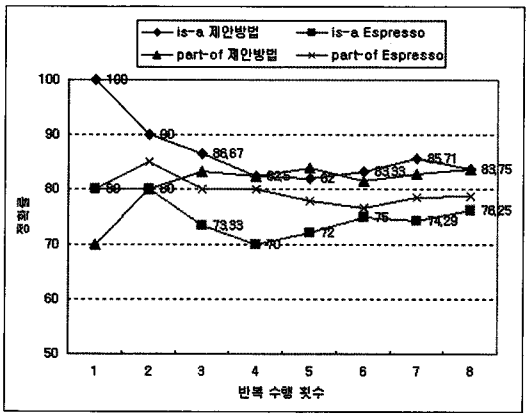

그림 4 정확률 비교 그래프(단위:%)

실험결과를 통해 알 수 있듯이, *is-a* 관계는 기존 방법보다 7.5% 향상된 83.75%의 정확률을 보였으며, *part-of* 관계는 5% 향상된 83.75%라는 동일한 정확률을 보였다. 이러한 성능 향상은 기존에 추출할 수 없었던 용어 쌍들을 제안한 패턴을 사용하여 추출하고, 추출된 용어 쌍들이 보다 유용한, 또는 기존 방법으로 찾을 수 없는 높은 신뢰도를 가진 패턴을 찾아냄으로써 올바른 용어 쌍들을 추가적으로 추출할 수 있었기 때문이라고 할 수 있겠다. 그리고 기존의 방법은 초기 수행 과정

에서의 정확률이 상승과 하락을 반복하는 안정적이지 못한 결과를 보이지만, 제안한 방법은 비교적 짧은 반복 수행 과정을 거치면서 안정적인 정확률에 다가가고 있음을 알 수 있다. 이는 제안한 패턴이 "기존 패턴이 고려하는 용어 쌍들 이외의 용어 쌍들"을 신뢰도를 측정하는데 함께 고려함으로써 상대적으로 안정된 정확률의 변화를 가져온 것으로 판단된다.

본 실험결과를 통해 각 방법론의 재현율을 올바르게 측정할 수는 없지만 상대적인 재현율은 이미 측정된 정확률을 활용하여 구할 수 있다. 기존 방법론의 재현율이 "1"이라고 할 때, 제안한 방법론의 상대적 재현율은 "제안한 방법론의 정확률"을 "기존 방법론의 정확률"로 나눈 값이다. 다시 말해서, 제안한 방법론은 기존 방법론의 재현율보다 *is-a* 관계에 대해서는 1.1배, *part-of* 관계에 대해서는 1.06배의 재현율을 보일 가능성이 있다는 것이다.

5. 결 론

온톨로지 구축에 활용할 수 있는 용어들 간의 의미 관계 추출 방법은 찾고자하는 의미 관계를 대표할 수 있는 용어 쌍들을 수작업으로 구축하여 해당 용어들 사이에서 나타나는 문자열을 패턴으로서 추출하고 이를 다른 용어 쌍들을 추출하는데 활용하는 방법이 일반적이다.

본 논문은 이러한 패턴 기반 의미 관계 추출 방법이 한 문장에 포함된 패턴만을 사용함으로써 서로 떨어진 용어 쌍들에 대한 의미 관계를 추출할 수 없다는 한계점을 인식하고, 이를 보완하기 위해 용어를 공유하는 패턴 쌍을 이용한 의미 관계 추출 방법을 제안하였다. 본 방법론은 *is-a* 및 *part-of* 관계에 대해서 기존 방법론보다 각각 7.5%와 5%의 성능 향상을 보였으며 재현율의 향상에 대해서도 잠재적 가능성을 제시하였다. 그리고 자동화된 패턴 추출로 인해 코퍼스의 성격, 즉 도메인에 의존적이지 않은 의미 관계 추출 방법으로서 활용될 수 있다.

하지만 본 방법론은 패턴 쌍을 추출하거나 이와 관련된 다양한 통계치를 측정하는데 있어서 높은 복잡도가 요구되며 실제 실험에서도 많은 시간과 자원이 소모됨으로써 반복 수행에 대한 어려움이 발생하였다. 앞으로의 연구에서는 이러한 문제점을 감안하여 보다 최적화되고 개선된 알고리즘을 개발하고 유용한 통계 정보가 포함된 기반 자원들을 확보함과 동시에 *is-a*나 *part-of* 관계 이외의 의미 관계에 대해서도 본 방법론을 적용해 보아야 할 것이다.

참 고 문 헌

[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] 김혜민, 최익규, 김민구, "일반화된 패턴을 이용한 관계 추출 시스템", 한국컴퓨터종합학술대회논문집, Vol.32, No.1(B), pp. 658-660, 2005.

[9] X. Yang, J. Su, "Coreference resolution using semantic relatedness information from automatically discovered patterns," Proceedings of ACL, pp. 528-535, 2007.

[10] P. Pantel, D. Ravichandran, "Automatically labeling semantic classes," Proceedings of HLT/NAACL, pp. 321-328, 2004.

[11] Wikipedia, "http://www.wikipedia.org"

[12] Stanford tagger, "http://nlp.stanford.edu/software/tagger.shtml"

[13] WordNet 3.0, "http://wordnet.princeton.edu"