

# 한국어 서술어와 지식베이스 프로퍼티의 연결에 관한 연구

원유성<sup>0</sup>, 우종성, 김지성, 함영균, 최기선  
한국과학기술원

{styner0305, woo88, jiseong, hahmyg, kschoi}@kaist.ac.kr

## Predicate Linking to Knowledge Base Property

Yousung Won, Jongseong Woo, YoungGyun Hahm, Key-sun Choi  
KAIST

### 요 약

본 논문은 문장 내에 있는 서술어(Predicate)와 지식베이스(Knowledge Base)의 프로퍼티(Property)를 연결하여 (Lexicalization), 최종적으로는 자연어 텍스트로부터 의미 있는 정보를 도출하기 위한 방향을 제시하고 있다.

### 1. 서 론

사람들 사이에 공유되는 생각이나 정보는 끊임없이 언어의 형태로 만들어지고 있다. 웹에는 이러한 언어 정보가 디지털 정보로써 끊임없이 표현되고 있고, 특히 텍스트의 형식을 빌려, 명실 상부 인간 역사상 가장 거대한 정보 구조물로서 계속해서 확장해 나가고 있다.

웹에 있는 방대한 텍스트를 기계가 읽고 이해하여, 최종적으로는 다시 사람에게 의미 있는 정보를 제공하기 위해서는, 먼저 인간이 만들어낸 자연어 텍스트(unstructured)를 지식베이스와 같은 구조화된(structured) 정보와 연결할 필요성이 있다. 이것은 텍스트가 가지고 있는 의미를 보다 명확하게 할 뿐만 아니라, 인간의 사고 과정을 모델링 할 수 있는 온톨로지(Ontology)로서 표현이 가능해 진다는 의미를 가지고 있다.

본 연구에서는 한국어 문장 내 두 개체(entity) 사이의 관계(relation), 즉 서술어에 집중하여 관련된 지식베이스의 프로퍼티를 찾아내는 것(Predicate Linking)을 목적으로 한다.

### 2. 관련 연구

디비피디아(DBpedia)와 같이 우리가 흔히 말하는 대부분의 지식베이스는, 위키피디아(Wikipedia)의 인포박스(InfoBox)와 같은 반구조적(semi-structured)인 데이터로부터 만들어진다. 하지만 웹상에서 오직, 약 15-20%의 정보만이 이러한 형식으로 표현되어 있고, 그외 대부분의 정보는 자연어 텍스트의 형태로 존재하고 있다.[1]. 자연어 텍스트의 개체(주로 체언)와 지식베이스의 리소스(Resource)와의 연결에 대한 연구와 관련하여, 한국어권과 영어권 연구 모두 진행중에 있다[2][3]. 반면, 자연어 텍스트 상에 나타난 서술어와 지식베이스 프로퍼티와의 연결에 대한 연구는 한국어권에서는 아직까지 미비한 상황이다. 영어권에서 대표적으로 Distant Supervision[4], 즉 두 개체를 포함하는 문장은 그 두 개체의 관계가 문장속에 어떻게든 나타난다는 아이디어를 바탕으로 Relation Extraction에 대한 연구가 진행되고 있고, 대표적으로는 BOA Framework[1]가 있다.

본 논문에서는 Distant Supervision의 가정을 한국어 텍스트에 적용하여, 한국어 문장과 이 문장의 내용을 잘 표현 할 만한 RDF를 활용하여 자연어 텍스트에 나타난 서술어 표현과, 이에 대응할 가능성이 높은 지식베이스의 프로퍼티를 찾아내고, 이의 가능성을 높일 수 있는 자질(feature)들을 파악한다.

### 3. 접근 방법

본 논문에서 서술어와 지식베이스의 프로퍼티에 대한 연결을 위한 입력 데이터로써 한국어 위키피디아 2014년 9월 11일자 덤프와 한국어 디비피디아 2014를 사용하였다. 즉, 디비피디아의 RDF(Resource Description Framework)와 이 RDF의 정보를 담고 있을 법한 문장들을 수집 하여, RDF의 P(프로퍼티)와 관련성이 높은 서술어를 문장속에서 찾아내는 접근 방법을 따른다.

#### 3.1 Seed 프로퍼티와 S(주어), O(목적어) 쌍의 수집

한국어 디비피디아 2014는 총 약 1만여개의 한국어 프로퍼티(prop-ko)가 있고, 이 프로퍼티를 사용한 트리플이 약 215만개가 존재한다. 그 중에서 약 95퍼센트의 트리플(약 205만개)이 단지, 약 1500개의 프로퍼티로 이루어져 있다. 본 연구에서는, 영어나 숫자가 포함된 프로퍼티를 제외한 약 1300개의 Seed 프로퍼티를 대상으로 텍스트에 나타난 서술어와 연결하는 작업을 수행하였다.

먼저, Seed 프로퍼티를 사용한 디비피디아의 모든 트리플을 수집한다. 이때 S와 O의 Redirection 페이지도 고려하여, 임의의 Seed 프로퍼티에 해당되는 S, O 쌍을 많게는 수십만개 정도로 확보하였다.

연결 대상 프로퍼티 이름	트리플 개수	스티븐 스킬버그 오하이오 주
1 http://ko.dbpedia.org/property/이름	95,365	스티븐 스킬버그 오하이오 주
2 http://ko.dbpedia.org/property/계급	47,294	스티븐 스킬버그 오하이오주
3 http://ko.dbpedia.org/property/그림	37,882	스티븐 스킬버그 오하이오 주
4 http://ko.dbpedia.org/property/출생지	30,851	스티븐 스킬버그 오하이오 주
5 http://ko.dbpedia.org/property/장르	23,511	스티븐 스킬버그 오하이오주
6 http://ko.dbpedia.org/property/직업	23,084	스티븐 스킬버그 오하이오 주
7 http://ko.dbpedia.org/property/클럽연도	19,644	스티븐 스킬버그 오하이오주
8 http://ko.dbpedia.org/property/출장수(골)	18,868	스티븐 스킬버그 오하이오주
9 http://ko.dbpedia.org/property/클럽	18,391	스티븐 스킬버그 오하이오 주
10 http://ko.dbpedia.org/property/웹사이트	17,107	스티븐 스킬버그 오하이오주
...		

[그림 1] 사용 빈도가 높은 Seed 프로퍼티와 redirection을 고려한 SO쌍 예시(prop-ko:출생지)

#### 3.2 S, O 쌍을 포함하는 문장의 수집

한국어 위키피디아에는 총 약 270만개의 문장이 있다. 앞서 언급한 것처럼 하나의 P(프로퍼티)에 대해 다수의 S, O 쌍(Surface Form)을 확보하였고, 이렇게 확보한 S와 O를 모두 포함한 문장을 위키피디아 전체 문장속에서 찾는다. 이때, 수집할 문장의 형태소 분석 및 의존 구문 분석이 필요하고, 문장 수집의 질을 높이기 위해 다음과 같은 제약 사항을 둔다.

- 문장에 주어가 있는 것만 고려
- 문장은 트리플의 S와 O를 모두 포함
- S와 O의 바로 앞 글자는 공백
- S와 O의 바로 뒤 글자는 공백 또는 조사
- S와 O가 같은 단어인 경우 제외
- S와 O가 모두 한글이 아닌 경우 제외
- 문장 내 S의 위치 < O의 위치 < 서술어(P가 될 수 있는 후보)의 위치
- ...

3.3 Distant Supervision

앞서, 하나의 프로퍼티 P에 대해 관련된 다수의 S, O 쌍을 확보했고, 각 쌍의 S, O를 모두 포함한 문장들을 수집하였다. Distant Supervision은 두 개체의 관계(relation)는 두 개체를 포함하는 문장속에 내재되어있음을 가정하는 Relation Extraction의 대표적인 접근 방법이다. 따라서 앞서 수집한 문장에는 프로퍼티 P에 대한 표현이 반드시 있을 것이라는 전제 하에 많은 횟수로 발견된 서술어를 프로퍼티 P의 Lexicalization으로 인정한다. 본 논문에서는 이러한 가정을 뒷받침 할 수 있는 지표, R을 도입한다.

$$R_i = \sum_k \frac{SF(S_{ijk}, O_{ijk})}{TF(T_i)}$$

- $P_i$  : 임의의 프로퍼티
- $T_i$  :  $P_i$ 를 사용하는 트리플
- $S_{ijk}$  :  $T_i$ 의 Subject의 Surface Form k
- $O_{ijk}$  :  $T_i$ 의 Object의 Surface Form k
- $TF(T_i)$  :  $P_i$ 를 사용하는 트리플의 총 개수
- $SF(S_{ijk}, O_{ijk})$  :  $S_{ijk}$ 와  $O_{ijk}$ 를 포함하는 문장의 개수

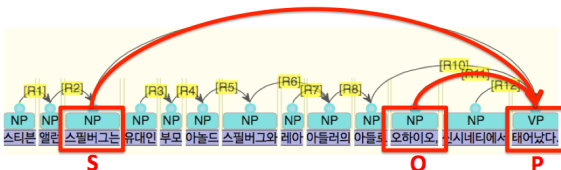
즉,  $R_i$ 는 임의의 프로퍼티  $P_i$ 로 이루어진 트리플과 대응하는 위키피디아 문장이 얼마나 존재하는가를 나타내고, 이 값이 작으면 프로퍼티  $P_i$ 는 문장으로 Distant Supervision의 가정으로 서술어를 획득하기에 적합하지 않다는 것을 의미한다. 예를 들어 “prop-ko:이름” 프로퍼티는 [그림 1]에서 알 수 있듯이, 한국어 디비피디아에서 가장 많이 사용된 프로퍼티로써, 쓰임새는 다음과 같다.

[스티브 잡스][prop-ko:이름]                      [스티브 잡스]  
 [스티브 잡스][prop-ko:이름]                      [Steve Jobs]

하지만, 위와 같은 트리플은 이 트리플의 정보와 대응 할 만한 위키피디아 문장(서술어로 표현되는)이 거의 없을 것이고, 따라서 R값이 작을 것이며, 이는 프로퍼티 “prop-ko:이름”은 서술어를 획득하기에 적절하지 않다는 것을 나타낸다. 위에서 이 값은 Weigted Scoring을 위한 하나의 자질로 사용된다.

3.4 프로퍼티 P에 대한 서술어 후보 수집

수집한 문장으로부터 두 개체의 관계에 해당하는 프로퍼티 P를 수집하기 위하여 본 논문에서는 문장에 대한 의존 구문 분석 나무(tree)를 확인한다, 나무에서 S와 O의 가장 가까운 공통 조상 노드(node)가 동사(VP) 또는 지정사(VNP)일 경우, 이를 서술어로 판단하고, 이에 대한 어근을 S, P, O의 프로퍼티 P에 대한 서술어 후보군으로 넣는다. 이때, 프로퍼티의 Domain, Range를 확인하여 서술어 후보군의 질을 높일 수 있다.



[그림 2] 의존 구문 분석 나무를 이용한 서술어의 선택

S의 Surface Form “스피버그”와 O의 Surface Form “오하이오”를 동시에 포함하고 있는 문장 “스티브 앨런 스피버그는 유대인 부모 아놀드 스피버그와 레아 아틀러의 아들로 오하이오, 신시네티에서 태어났다.”

이렇게 임의의 프로퍼티  $P_i$ 에 대하여, 의존 구조 분석을 통해

수집한 서술어의 어근을 기준으로 가장 높은 빈도로 등장한 어근을 프로퍼티  $P_i$ 의 정답으로 인정 한다.

$$FS_{ij} = \frac{N_{ij}}{N_i}$$

- $N_{ij}$  : 프로퍼티  $P_i$ 를 통해 발견된 서술어 중에서, 어근  $j$ 를 가지는 서술어의 총 개수
- $N_i$  : 프로퍼티  $P_i$ 를 통해 발견된 서술어의 총 개수

다시 말하면, 임의의 프로퍼티  $P_i$ 를 통해 수집한 문장에서 나타난 서술어 목록 중에서  $FS_{ij}$  값이 큰 서술어가 연결될 가능성이 높다. 이후 과정에서는 [그림 3]를 통해 짐작 할 수 있듯이, FS값 순으로 나열된 후보들 중에서 가중치를 통해 순위를 끌어 올릴 수 있는 방법을 소개한다.

P(i) = prop-ko:출생지				
i	어근	빈도	FS(i,j)	서술어(어근 j)
1	태어나	23	0.081560284	태어나 태어났으며 태어난 태어났다
2	가	9	0.031914894	가 가는 간다 갔다 가기를 가서 가게
3	돌아오	8	0.028368794	돌아왔다 돌아와 돌아온다 돌아올
...	...	...	...	...
5	돌아가	5	0.017730496	돌아가고 돌아가 돌아가려 돌아가기 돌아가지
9	지내	4	0.014184397	지내 지냈다
10	활동	4	0.014184397	활동하던 활동했다 활동하고 활동하였다
...	...	...	...	...
15	귀국	3	0.010638298	귀국하여 귀국해 귀국했다
16	보내	3	0.010638298	보내 보내
17	이끌	3	0.010638298	이끌어 이끌었다 이끌었으며
...	...	...	...	...
22	출생	2	0.007092199	출생했다
23	가져	2	0.007092199	가져 가지고
24	전하	2	0.007092199	전하여 전한다
25	온신	2	0.007092199	온신했다 온신하였다
...	...	...	...	...

[그림 3] “prop-ko:출생지”에 대한 서술어 후보

3.5 의미적 유사도 기반한 가중치

[그림 3]의 예시에서 볼 수 있듯이 서술어 “태어나”를 프로퍼티 “prop-ko:출생지”에 연결 할 수 있는 근거를 확보하였다. 하지만 비단 “태어나” 뿐만이 아닌 “태어나”와 의미적으로 유사한 후보를 찾아 특정 자질을 부여하여 최종적으로 서술어와 프로퍼티 연결의 품질을 높일 수 있다. 이에 대한 방안으로 CoreNet<sup>1)</sup> 개념 번호에 기반한 어휘간 유사도를 Jaccard Similarity(JS)를 통하여 측정한다.

$$JS(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

- A : FS값이 가장 높은 서술어의 CoreNet 개념 번호 집합
- B : 후보 대상 서술어의 CoreNet 개념 집합

즉, 최상위 서술어 후보와 후보 대상 서술어 간에 많은 수의 CoreNet 개념 번호를 공유한다면 의미적으로 유사하다는 것을 가정한다. 이 값은 최종적으로 서술어 후보의 Weighted Scoring의 자질값으로 사용된다.

3.6 Weighted Scoring

지금까지 임의의 프로퍼티에 대해 다수의 서술어 후보를 수집하였다. 이때 적절한 서술어를 선택하기 위해서, 먼저 서술어를 수집할 만큼의 문장이 있는지(R), 그리고 서술어가 임의의 프로퍼티에 대해 다수의 문장속에서 높은 비율로 사용되는지(FS), 마지막으로, 최상위 서술어 후보와 의미적으로 유사한 또다른 후보 대상이 있는지(JS)를 고려하여, 이러한 자질별 가중치를 통해 최종 스코어(WS)를 산출한다.

$$WS_{ij} = (\omega_{fs} \times FS_{ij}) + (\omega_{js} \times JS_{kj}) + (\omega_r \times R_i)$$

- $\omega_x$  : x에 대한 가중치 (x = fs, js, r, ...)
- i : 프로퍼티
- j : 어근
- k : 프로퍼티 i에 관하여, FS값이 가장 높은 어근

1) 다국어 어휘 의미망

([http://semanticweb.kaist.ac.kr/org/bora/CoreNet\\_Project/index.html](http://semanticweb.kaist.ac.kr/org/bora/CoreNet_Project/index.html))

다시 말하면, 임의의 프로퍼티로 표현되는 다수의 트리플과, 이 트리플의 정보를 잘 담고 있다고 할 수 있는 다수의 문장으로부터 선택된 서술어는 높은 WS 값을 갖게 된다.

디비피디아 프로퍼티	FS(i,j)	JS(k,j)	R(i)	WS(i,j)	어근(i)	서술어(어근 j)
prop-ko:출생지	0.081560284	1	0.010210366	0.342822215	태어나	태어나 태어났으며 태어나 태어났다
prop-ko:출생지	0.007092199	1	0.010210366	0.305588172	출생	출생했다
prop-ko:출생지	0.031914894	0	0.010210366	0.01799952	가	가 가는 간다 갔다 가기를 가서 가게
prop-ko:출생지	0.028368794	0	0.010210366	0.01622647	좋아오	좋아왔다 좋아와 좋아온다 좋아올
prop-ko:출생지	0.024822695	0	0.010210366	0.014453421	떠나	떠날때 떠났다 떠나기로 떠나서 떠나 떠났으며

[그림 4] FS, JS, R의 Weighting Score

(가중치로는  $\omega_{fs} = 0.5$ ,  $\omega_{js} = 0.3$ ,  $\omega_r = 0.2$ 를 적용하였고 프로퍼티 "prop-ko:출생지"와 서술어 "태어나", "태어났으며" 등을 연결하였다.)

#### 4. 실험 결과 및 성능 평가

본 연구에서는 임의의 프로퍼티와 이에 대한 연결 가능성이 있는 서술어 후보를 두고, 세 가지의 자질 값 R, FS, JS의 Weighted Scoring<sup>2)</sup>을 통하여 가능성이 높은 서술어를 선택하였다. 프로퍼티-서술어 간 정답셋이 존재하지 않기 때문에 디비피디아 2014에서 가장 빈번하게 사용된 상위 200개의 프로퍼티에 대해서, 이 실험에 의해 연결된 서술어의 적합성을 수작업으로 평가하였다. 평가는 세 사람의 평가자가 프로퍼티의 쓰임새(관련 트리플)를 보고, 자동적으로 연결된 서술어와 의미적 타당성 여부를 검토한 뒤, 다수의 판단 결과를 정답으로 인정하였다.

정밀도는 이 실험의 접근 방법에 의해 수집된 서술어(어근) 후보군 중에서 평가자가 적절하다고 판단된 서술어(어근)의 비율을 측정하였고, 재현율은 실험 특성상 정답셋이 명확하게 존재하지 않으므로, 앞선 결과를 토대로 평가 대상 프로퍼티 중에서 적절한 서술어(어근)을 연결할 수 있는 프로퍼티의 비율로 정의하여 측정하였다.

$$\text{정밀도(Precision)} = \frac{\text{타당하다고 판단된 프로퍼티-서술어(어근) 연결 갯수}}{\text{각 평가 대상 프로퍼티에 연결된 후보 서술어(어근)의 총 갯수}}$$

$$\text{재현율(Recall)} = \frac{\text{타당하다고 판단된 서술어가 연결된 프로퍼티의 갯수}}{\text{평가 대상 프로퍼티의 총 갯수}}$$

전체 실험 결과의 일부를 수록한 [그림 5]에서, WS값이 특정 임계값 이상인, 프로퍼티-서술어 연결 결과 및 타당성 여부를 찾아 볼 수 있다. 구체적인 평가 방식으로는, 한국어 디비피디아 2014에서 가장 빈번하게 사용된 상위 200개의 프로퍼티 중에서, 문장 수집이 가능했던 137개의 프로퍼티와 연결된 서술어(어근) 235개<sup>3)</sup>에 대해서 세 사람이 직접 그 타당성을 검증하였다. 이때 평가자 중 다수가 적절하다고 판단한 서술어를 프로퍼티와 실제로 연결할 필요성이 있는 정답으로 간주하였다. 검증 결과, 본 논문의 방법으로 찾아낸 총 235개의 서술어(어근) 중에서 99개가 적절한 연결 결과로 판단 되었고, 이는 약 42.12%의 정밀도(Precision)을 보여준다. 또한 137개 프로퍼티중에서 75개의 프로퍼티에 적절한 서술어가 연결되어, 약 55.56%의 재현율(Recall)을 보였다(F1-score:47.91).

본 실험 결과는 가정한 가중치( $\omega_{fs} = 0.5$ ,  $\omega_{js} = 0.3$ ,  $\omega_r = 0.2$ )와 최종 서술어 선택을 위한 임계값(Threshold > 0.1)에 따라 정밀도 및 재현율이 영향을 많이 받기 때문에 이에 대한 추가 실험이 필요하고 가중치 및 임계값에 대한 최적화가 향후 연구 과제이다. 서술어-프로퍼티 연결 결과는 데모 웹사이트<sup>4)</sup>에서 결과를 확인할 수 있고, 그 활용 방안도 살펴볼 수 있다.

2) 가중치로는  $\omega_{fs} = 0.5$ ,  $\omega_{js} = 0.3$ ,  $\omega_r = 0.2$  를 적용  
 3) 프로퍼티 137개에 대해 임계값이 0.1이상인 어근 235개가 연결됨 (프로퍼티 1개당, 약 1.7개의 어근 연결)  
 4) http://143.248.135.187:11111

#### 5. 결론 및 향후 계획

본 연구는 자연어 텍스트에서 나타난 서술어와 이미 존재하고 있는 지식베이스의 프로퍼티와의 연결을 시도하였다. 이것은 문장의 단순화(Sentence Simplification)를 통해 발견할 수 있는 체언과 용언(Entity Linking + Predicate Linking)의 관계를 통하여 자연어 텍스트로부터 지식베이스 확장의 가능성을 열었다는데 그 의의가 있다. 또한, 서술어로 표현 될 수 있는 프로퍼티와 그렇지 않은 프로퍼티를 통해 디비피디아와 같은 지식베이스의 한계를 가능해 볼 수 있는 근거를 제시했다고 볼 수 있다. 이것은 현존하는 지식베이스가 어느정도 수준의 정보를 대변할 수 있는지에 대한 의문과, 프레임넷(FrameNet)과 같은 의미 체계 또는 진화된 온톨로지 스키마(Schema)의 필요성을 보여주는 접근이라고 볼 수 있다[5].

향후 계획으로는 문장 속에서 서술어 후보를 수집하는 데, 의존 구문 분석 결과 뿐 만이 아닌, 서술어의 용언 격률 정보, 문장 구구조 패턴의 다각화를 통해 문장 단순화 작업의 질을 높여, 더 가능성이 높은 서술어 후보를 수집하는 것을 목적으로 할 것이다. 또한 한국어 프로퍼티를 온톨로지 프로퍼티와 맵핑(Mapping)하여 가져올 수 있는 온톨로지 스키마를 통해 문장 단순화와 프로퍼티-서술어 연결의 질을 높이는 선순환 구조를 마련할 계획이다.

Rank	프로퍼티(P)	W <sub>i</sub>	어근(i)	적합성	어근(j)로 이루어진 서술어(k)
2	http://ko.dbpedia.org/property/제목	0.365661	방송	X	방송하였으나 방송하였는데 방송중으로 방송되 방송되었고 방송하여 방송하여 방송하여 방송되
3	http://ko.dbpedia.org/property/그림	0.188262	결정	X	결정되어 결정 결정되 결정되 결정되 결정되
		0.188262	말어나	X	말어나기도 말어나지를 말어나는 말어나
		0.152968	개롱	X	개롱하여 개롱하여 개롱하여 개롱하여 개롱하여
		0.152968	크	X	크루게이다 크루 크루게이다 크루 크루
		0.117673	폐쇄	X	폐쇄하였으나 폐쇄당하여 폐쇄되었으며 폐
4	http://ko.dbpedia.org/property/출생지	0.342822	태어나	O	태어나고 태어나지 태어나라고 태어나오
		0.305588	출생	O	출생하였지만 출생이지만 출생지이다 출생했
5	http://ko.dbpedia.org/property/장르	0.327941	방탄	X	방탄할 방탄되어 방탄 있다 방탄되고 방탄하다
		0.107592	말리	X	말리 말리여 말리여 말리여 말리여 말리여
6	http://ko.dbpedia.org/property/직업	0.354732	활동	O	활동중이던 활동중이라는 활동중인 활동중
		0.112156	활약	O	활약하고 활약하다 활약하여 활약하다 활
7	http://ko.dbpedia.org/property/클럽연도	0.360316	이적	O	이적한후 이적한지 이적해 이적했다 이적해
8	http://ko.dbpedia.org/property/출장수(골)	0.352613	기록	O	기록되어있을 기록하기 기록하다 기록하
9	http://ko.dbpedia.org/property/클럽	0.547814	이적	O	이적한후 이적한지 이적해 이적했다 이적해
11	https://ko.dbpedia.org/property/노선	0.417945	운행	O	운행되고 운행되었지만 운행하지는 운행하
12	http://ko.dbpedia.org/property/소재지	0.344744	위치	O	위치시켜 위치시켰다 위치한상태기록할이
13	http://ko.dbpedia.org/property/그림설명	0.194295	우승	X	우승이며 우승이라는 우승된 우승하였다
14	http://ko.dbpedia.org/property/전임자	0.327376	해임	O	해임하는 해임당하여 해임당하여 해임당
		0.120942	내내	O	내내하고 내내하였지만 내내하는 내내
		0.117511	내놓	O	내놓기 내놓아내놓고 내놓았고 내놓아내
15	http://ko.dbpedia.org/property/후임자	0.336804	즉위	O	즉위시키고 즉위하여 즉위함으로써 즉위
		0.314205	확정	O	확정되었고 확정되었을 확정하였던 확정
		0.313489	등기	O	등기하면서 등기하는 등기했다 등기할

[그림 5] 프로퍼티-서술어 연결 결과

#### 사사

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임. (No. R0101-15-0054, WiseKB: 빅데이터 이해 기반 자가학습형 지식베이스 및 추론 기술 개발)

#### 6. 참고 문헌

[1] Daniel Gerber and Axel-Cyrille Ngonga Ngomo, Bootstrapping the Linked Data Web, 2011  
 [2] Youngsik Kim, Younggyun Hahm, Jisung Kim, Dosam Hwang, Key-Sun Choi, 한국어 텍스트의 개체 URI 탐지: 품사 태깅 독립적 개체명 인식과 중의성 해소, 제 26회 한글 및 한국어 정보처리 학술대회, 2014  
 [3] Mendes P.N., Jakob M., Garcia-Silva A., Bizer C., DBpedia Spotlight: Shedding Light on the Web of Document, I-Semantics, 2011  
 [4] Mike Mintz, Steven Bills, Rion Snow, Dan Jurafsky, Distant supervision for relation extraction without labeled data, ACL, 2009  
 [5] Younggyun Hahm, Jiwoo Seo, Dosam Hwang, Key-Sun Choi, 프레임넷을 통한 디비피디아 온톨로지 인스턴스 생성의 커버리지 개선, 제 26회 한글 및 한국어 정보처리 학술대회, 2014