

# Linked Data를 위한 한국어 자연언어처리 플랫폼

함영균<sup>†</sup>, 임경태<sup>†</sup>, Martin Rezk<sup>†</sup>, 박정열<sup>§</sup>, 윤용운<sup>†</sup>, 최기선<sup>†</sup>

<sup>†</sup> 한국과학기술원, Semantic Web Research Center § Les Editions an Amzer Vak, Lannion, France  
<sup>†</sup> {hahmyg, kyungtaelim, mrezk, yoon, kschoi}@kaist.ac.kr, § park@amzer-vak.fr

## Korean Natural Language Processing Platform for Linked Data

YoungGyun Hahm<sup>†</sup>, Kyungtae Lim<sup>†</sup>, Martin Rezk<sup>†</sup>, Jungyeul Park<sup>§</sup>, Yoon Yongun<sup>†</sup>, Key-Sun Choi<sup>†</sup>  
<sup>†</sup>한국과학기술원, Semantic Web Research Center. §Les Editions an Amzer Vak, Lannion, France

### 요약

본 논문에서는 한국어 자연언어처리를 위해 형태소분석기와 구구조 구문분석기와 의존구조 구문분석기를 통합한 하나의 플랫폼을 제공하고, 외국의 다양한 자연언어처리 도구들의 결과물과의 국제적 상호운용성 및 Linked Data를 위한 RDF 형태로의 변환 시스템을 제시한다.

**주제어:** 자연언어처리, 형태소분석, 구문분석, NLP2RDF, Linked Open Data

### 1. 서론

웹상에서의 LOD(Linked Open Data)<sup>1)</sup>에 관한 연구는 비교적 최근의 연구이지만 빠르게 진척되고 있다. 이는 다양한 형태로 존재하는 개방된 데이터들의 접근성을 높여 연결하고 그 과정에서 새로운 지식창출을 기대하는 정보관리 및 통합에 관한 이슈이다. 이러한 데이터 간의 연결을 위하여 각자 다른 형태로 존재하고 있는 데이터들에 대하여 상호운용성과 규격을 확보하고자 하는 노력이 진행되고 있다[1]. 이러한 노력의 일환으로 W3C<sup>2)</sup>에서 정의한 RDF는 메타데이터를 설명할 수 있는 상호운용적인 규격이며, 지능형 웹을 위한 언어이다.

특히, 자연어의 경우에는 다양한 국가의 다양한 언어가 비구조화된 텍스트로 존재하고 있으며, 이에 대한 자연언어처리의 결과 역시 서로 다른 포맷으로, 같은 언어에 대해서도 응용프로그램에 의존적인 포맷으로 제공되고 있다. 이는 다른 품사태그셋을 사용하거나 혹은 다른 자연언어처리 도구를 사용하면서 발생하는 포맷의 이질성의 문제이며, 특히나 언어의 다양성에 의해 자연언어처리 결과물에 대한 국제적 상호운용성과 접근성의 문제를 수반한다.

따라서 국제적으로 다양하게 사용되는 자연언어처리의 결과물들을 통합할 필요성이 있다. 이를 위하여서는 다양한 용어들로 출력되는 자연언어처리 결과물들을 설명할 수 있는 온톨로지와 용어사전이 필요하며, 이를 바탕으로 특정한 포맷으로 자연언어처리 결과물들을 바꾸어 줄 수 있는 도구 역시 필요하다.

또한, 한국어의 경우에는 영어나 다른 언어에 비교하여 자연언어처리 연구 성과가 완전히 공개되는 일이 미흡한 것이 사실이다. 본 논문에서는 한국어 자연언어처리를 위한 플랫폼 구축을 위한 과정이 담겨져 있다. 그리고 연구 성과를 공개한다. 이 과정에서 기존에 존재하

는 형태소분석기와 구문분석기와 같은 도구들을 사용하였고, 해당 도구들의 성능향상을 위한 연구 역시 병행되었다. 여기서 사용된 자연언어처리 도구들은 하나의 문장을 입력으로 받아, 품사태그 정보와 구구조형태, 의존형태의 결과물을 출력하는 플랫폼으로 통합되었다. 출력되는 결과물들은 세종코퍼스를 학습데이터셋으로 사용하였고, 품사태그 역시 세종품사태그셋을 사용하였다.

이러한 자연언어처리의 결과물들의 국제적 상호운용성을 위하여 RDF(Resource Description Framework) 출력형태에 기반한 NIF(NLP Interchange Format)[2][3] 규격을 준수했다. 이를 위해 세종품사태그셋을 기반으로 한 한국어 언어태그 온톨로지를 구축하였고, OLIA(Ontologies of Linguistic Annotation)<sup>3)</sup>와의 매핑을 통하여 국제적 상호운용성을 확보하기 위해 노력하였다. 출력된 RDF Triple은 SPARQL 쿼리를 지원한다. 이를 통하여 자연언어처리 결과의 국제적 상호운용성을 확보하는 한편, 자연언어처리를 통한 Linked Data의 기초를 마련하고자 한다.

2장에서는 Linked Data를 위한 국제적 규격과 상호운용성을 위한 현재의 논의가 소개되고, 3장에서는 한국어 자연언어처리 플랫폼 구축을 위한 노력이, 4장에서는 자연언어처리 결과를 RDF형태로 변환하는 과정이 보고된다. 5장에서는 결론 및 향후 연구 방향에 대하여 논의한다.

### 2. 관련연구

현재 Linked Data를 위하여 RDF결과를 출력하는 도구는 Stanford Core-NLP<sup>4)</sup>가 대표적이다. Stanford Core-NLP의 경우 자연언어처리 도구들의 결합을 통해서 하나의 입력에 대하여 품사태그와 구구문분석, 의존구문분석 등의 다양한 자연언어처리 분석 결과물을 출력한

1) <http://lod2.eu>

2) <http://www.w3.org/>

3) <http://www.sfb632.uni-potsdam.de/~chiarcos/ontologies.xml>

4) <http://nlp.stanford.edu/software/corenlp.shtml>

다. 이 결과물은 NLP2RDF 프로젝트<sup>5)</sup>에서 개발된 Stanford Core-NLP Wrapper<sup>5)</sup>를 통해 NIF의 RDF결과로 변환되어 출력된다.

형태소 분석의 경우, 이미 충분한 연구가 진행되었고 한나눔<sup>6)</sup> [12]과 같은 몇몇 형태소분석 도구가 공개되어 있다. 구문분석은 한국어가 다른 언어에 비해 비교적 어순이 자유롭다는 특징이 있어서 의존구문분석에 관한 연구에 집중되어 있다[4][5]. 최근에는 세종 트리뱅크의 형태를 의존 구문 문법 형태로 변환하여 기계학습을 사용하기도 하였다[6]. 양질의 언어자원인 세종 트리뱅크를 사용하여 구문분석을 위한 연구가 진행되기는 하였으나[7], 현재까지는 그 연구의 공개가 활성화되어있지 않고, 또한 다양한 포맷으로 결과물을 출력하는 다양한 도구들이 존재하고 있어 상호운용성이 부족한 것이 사실이다. 따라서 본 논문에서는 하나의 플랫폼으로서 품사태그 정보와 구문분석정보, 의존구문분석정보를 모두 출력하는 시스템을 구축하고자 하고, 자연언어처리 결과물의 국제적 상호운용성을 위하여 NIF규격을 따르는 RDF를 통해 메타데이터를 기술하고자 한다.

### 3. 한국어 자연언어처리 플랫폼

한국어 자연언어처리를 위한 플랫폼을 구축하기 위하여 기존의 3가지 도구 및 방법들을 사용하였다. 이 과정에서 다음 3가지의 도구들의 결합 작업을 수행하였고, 성능 향상을 연구하였다. 1) 한나눔 2) 버클리 구문분석기<sup>7)</sup> 3) 의존형태 구문분석 결과 출력. 이 3가지의 도구들의 결합으로, 하나의 한국어 문장으로부터 품사태그 정보와 CFG(Context-Free Grammar), 그리고 DG(Dependency Grammar)형태의 결과물을 얻을 수 있다.

#### 3.1 형태소분석기

3.2장에서 사용된 버클리 구문분석기는 트리뱅크에서 품사 태거를 학습한다. 그러나 영어의 경우 품사태거 학습의 단위인 토큰이 띄어쓰기 단위와 같지만 한국어의 경우 한 개의 띄어쓰기 단위인 어절 내에 여러 토큰이 있으므로 이를 분리하는 형태소해석을 하여 토큰을 분해할 필요가 있다.

본 연구에서는 KAIST Semantic Web Research Center<sup>8)</sup>에서 자체개발한 한나눔 형태소분석기를 사용 하였다. 한나눔 형태소분석기는 1999년 C언어로 개발되었고 2010년 Java언어로 변환되었다. 독립적으로 사용되는 자연언어처리 도구로서 오픈소스로 공개되어 있으며 분석 수준에 따라서 크게 3가지로 나뉜다.

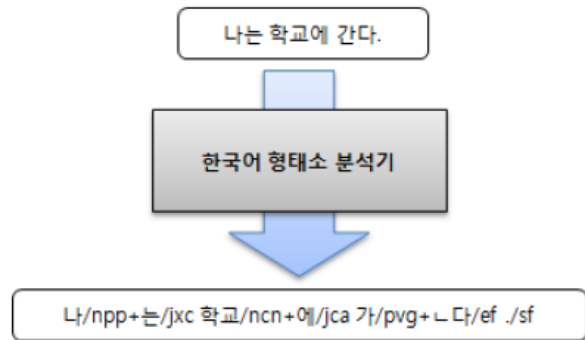


그림 1 형태소 분석기 입출력의 예

1. 전처리 단계: 문장 경계인식, 필터링, 자동띄어쓰기 등 형태소분석 이전에 필요한 전처리 작업을 수행한다.
2. 형태소분석 단계: 입력 문장에 대해서 어절 단위로 발생 가능한 모든 형태소분석 결과를 생성한다.
3. 품사태깅 단계: 가장 유망한 형태소분석 결과를 선택하여 입력 문장에 대한 최종 품사태깅 결과[8]를 반환한다.

한나눔이 본 연구를 위해 사용될 형태소 분석기로서 크게 두 가지 문제점이 있는 것으로 파악되었다. 첫 번째로 형태소 분석 단어 사전의 버전이 오래되어 미등록어 문제로 형태소분석 결과가 잘못 나와 구문분석기 오류를 증폭 시킬 수 있다는 점, 두 번째로 조사 어미 분석을 과도하게 세분화하여 구문분석에서 토큰 단위의 모호성을 만들 수 있다는 점이다. 두 문제점을 최소화하기 위해 사전 추가와 형태소분석 수준의 세밀함에 대한 조정이 필요할 것으로 보인다.

#### 3.2 구문분석기

한국어의 경우에는 주로 의존구조를 사용한 연구에 집중되어 있었기에 구구조 구문분석의 경우에는 영어에 비하여 연구가 거의 이루어져 있지 않다. 본 연구에서는 세종 트리뱅크<sup>9)</sup>라는 양질의 구구조 코퍼스를 사용하여 기존의 영어 구문분석기에 적용하여 실험하였고, 그 결과 다른 구구조 구문분석기 중에서 버클리 구문분석기의 성능이 가장 높게 나왔음을 확인할 수 있었다[7]. 세종코퍼스의 일부를 버클리 구문분석기의 학습 데이터셋으로 사용하기 위하여 간단한 전처리작업을 수행하였고[9], 이 과정에서 구문분석기의 성능 향상을 위하여 한국어의 특성에 맞추어 변환하는 연구가 병행되었다.

본 연구에서는 한국어를 위해 변환된 버클리 구문분석기는 자연언어처리 플랫폼 구축을 위하여 한나눔 형태소분석기의 결과를 세종태그셋 형태로 변환하였고, 훈련모듈을 포함한 형태로 오픈소스로 공개되어 있다.<sup>10)</sup> 특히,

5) <http://code.google.com/p/nlp2rdf/source/browse/>

#hg%2Fimplementation%2Fstanford-core

6) <http://sourceforge.net/projects/hannanum/>

7) <http://code.google.com/p/berkeleyparser/>

8) <http://semanticweb.kaist.ac.kr/>

9) <http://www.sejong.or.kr>

10) <http://semanticweb.kaist.ac.kr/home/index.php/KoreanParser>

사용자 편의를 위하여 간단한 입출력을 이용할 수 있도록 구성되어 있다. 그림 2는 한국어파서 (Java 프로그램)11)과 입력문장 “나는 학교에 간다.”에 대한 구구조 출력을 보인다.

```
Java -jar Berkeleyparser_korV2.jar "나는 학교에 간다."
(ROOT
  (S (NP_SBJ (NP 나) (JX 는))
    (VP (NP_AJT (NNG 학교) (JKB 예))
      (VP (VV가) (EF ㄴ다) (SF .))))))
```

그림 2 구구조분석기 입출력의 예

“나는 학교에 간다.”

- |                       |   |        |      |
|-----------------------|---|--------|------|
| 1. 나/NP+ 는/JX         | 3 | NP_SBJ |      |
| 2. 학교/NNG + 예/JKB     | 3 | NP_AJT |      |
| 3. 가/VV + ㄴ다/EF + .SF | 0 | VP     | ROOT |

그림 3 의존구조 출력형태의 예

### 3.3 의존형태 구문분석

구구조형태 구문분석의 경우에는 세종 트리뱅크라는 사람의 손으로 만들어진 코퍼스가 존재하지만, 의존구조형태 구문분석의 경우에는 그러한 양질의 코퍼스가 존재하지 않는다. 이에 [10]에서 사용된 알고리즘만을 사용하여 자연언어처리도구를 사용하지 않고서도 구구조형태의 문법을 의존구조형태로 변환할 수 있는 도구를 추가하였다. 따라서 이 논문에서 제안하는 자연언어처리 플랫폼은 3.2에서 출력된 구구조형태의 자연언어처리 결과물을 입력으로 받아서 그림 3과 같은 의존구조형태의 결과물을 제공한다. 그림 3에서 “나/NP+는/JX” 3번 “가/VV+ ㄴ다/EF+.SF”에 NP\_SBJ (주격 명사구)의 형태로 걸림을 의미한다.

전체적으로 누적된 오류가 F1-score상으로 18% 선으로 영어 버클리파서의 성능보다 10%정도 미흡하지만, 품사태그의 조절과 원래 목적인 RDF변환과 LOD-cloud에 대한 링크발견에 대한 최적화로서 대상 개체의 획득 정확도 및 LOD property에 대한 매핑을 위한 평가척도가 마련되어야 한다.

## 4. RDF 변환

한국어 자연언어처리 결과의 형태는 다른 언어의 자연언어처리 결과와 구조와 어휘에 있어서 다르다. 이러한 상이한 포맷을 가지고 있는 메타데이터를 설명할 수 있는 방법으로서 RDF가 사용되며, 이를 통해서 서로 다른 자연언어처리도구의 결과물들을 상호운용할 수 있는 기

초를 마련할 수 있다. 또한 자연언어처리로 인해 RDF로 변환된 텍스트의 entity들은 LOD-cloud의 entity들과 연결됨으로서 Linked Data의 자원이 될 수 있다.

이 과정에서 한국어의 품사태그에 대한 온톨로지를 구축하고, RDF의 술부 어휘들에 대해서도 규격을 따랐다. 최종적으로는 3장에서 사용된 자연언어처리도구의 결과물들을 RDF triple로 추출할 수 있었다.

### 4.1 세종품사태그셋 온톨로지

자연언어처리의 상호운용성 확보를 위한 LOD2 서브프로젝트인 NLP2RDF12)에서는 나라마다 서로 상이하게 사용되는 품사태그를 설명할 수 있는 온톨로지로서 OLiA(Ontologies of Linguistic Annotation)를 사용한다. 영어의 경우 펜 트리뱅크의 품사들과 OLiA를 매핑시켜 줌으로서 이러한 상호운용성을 확보하고 있다. 따라서 한국어의 세종품사태그셋의 온톨로지를 구축하고, 그것을 OLiA와 매핑시켜 주었다. 이때, 세종품사태그셋 온톨로지는 텍스트에 주석으로 태깅된 정보를 의미하며, 그것의 언어적 개념을 설명하는 온톨로지로서 OLiA가 사용된다. 구축된 온톨로지와 OLiA와의 매핑은 아래와 같다.

이 과정에서 명사나 부사 같은 영어에도 존재하는 품사태그의 경우에는 최대한 영어의 매핑 형태를 따르고자 하였으나, 한국어 형용사의 경우에는 동사의 성격도 있어서 동사와도 매핑되었으며, 조사의 경우에는 OLiA의 MorphosyntacticCategory의 하위 클래스가 아닌, MorphologicalCategory의 하위클래스로 속하게 매핑되었다.

Tag	Sejong	OLiA
superclass	LinguisticAnnotation/Tag/	LinguisticConcept/MorphosyntacticCategory/
MA	MA	Adverb
	MAJ	Adverb/ConjunctiveAdverb
	MAG	Adverb/GeneralAdverb
SN, XN	CardinalNumber	Quantifier/Numeral
MM	Determiner	PronounOrDeterminer/Determiner
SH, SL	ForeignWord	Residual/Foreign
IC	Interjection	Interjection
XR	Noun/BaseMorpheme	Noun/CommonNoun
NN	NN	Noun
	NNB, NNG	Noun/CommonNoun
	NNP	Noun/ProperNoun
	NA, NF	LikelyNoun
NP	Pronoun	PronounOrDeterminer/Pronoun
SE, SF, SO, SP, SS	Symbol	Punctuation
V	NV, V	Verb
	VA	Verb/Adjective
	VX	Verb/AuxiliaryPredicate
	VC, VCN, VCP	Verb/Copula
	VV	Verb/VerbalPredicate
		Verb
E, JK, XP, XS	E, JK, XP, XS	Particle
	JC, JK	Particle/AuxiliaryPostposition
	JKB, JKC, JKG, JKO, JQJ, JKS, JKV	Particle/CaseMarker
	XPN	Particle/Prefix
	XSA, XSN, XSV	Particle/Suffix
	EC, EF, EP, ETM, ETN	Particle/VerbalEnding
		Particle/VerbalEnding

그림 4 세종품사태그셋과 OLiA와의 매핑

### 4.2 RDF 온톨로지

NIF에서는 자연언어처리의 결과들을 RDF로 기술하는 데 있어서 사용되는 술부 어휘들에 대한 규격을 만들고 있다. 가장 최근까지 진행된 NIF1.0의 규격을 따른 과정

11) <http://semanticweb.kaist.ac.kr/home/index.php/KoreanParser>

12) <http://nlp2rdf.org/>

은 아래와 같다.

1) 문장 내 구성요소 문자열 (스트링) URI

String URI는 입력문장에 대하여 그 정해진 문헌 내에서 의미 있는 단위인 형태소, 어절, 구, 문장마다의 유일한 URI를 지정하고자 하는 것이다. 즉 Linked Data에 있는 어느 개체가 실제 문장의 어느 부분에서 연결되는지를 표시하기 위하여, 입력문장의 해당하는 부분도 유일하게 표시하는 것이 String URI의 역할이다. 이 때, 해당 문장의 단위의 위치 지정은 문헌URI의 처음부터 몇 번째 글자인지를 나타내는 offset방법과 context-Hash 방법이 있다. NIF에서는 offset과 context-Hash URI 스키마를 제공한다. 본 논문연구에서는 context-Hash URI를 사용하였다. 이 URI는 5가지 구성요소를 갖는다.

- (1) “hash” 단어,
- (2) 문자열의 왼쪽과 오른쪽에 위치한 글자 수의 길이,
- (3) 문자열의 전체 길이,
- (4) 문자열을 md5 hash로 변환,
- (5) 사람이 읽을 수 있도록 원래의 문자열을 포함한다.

“나는 학교에 간다.”

‘학교’  
hash\_15\_2\_c3508b1509ed7789297de77cfe9fb14f\_학교

그림 5 예문 “나는 학교에 간다.”에서 ‘학교’ 문자열에 대한 context-Hash URI의 예

2) RDF 술부 어휘 온톨로지

NIF에서는 자연언어처리 결과물을 RDF로 출력할 때 사용할 수 있는 Object property와 Data property에 대한 술부 어휘에 대한 규격이 지정되어 있다. 본 연구에서는 이러한 규격을 따랐고, 사용된 술부 어휘는 표1과 표2와 같다.

표 1 Object property

nextSentenceTrans	nextSentence	
nextWordTrans	nextWord	
oliaLink		
previousSentenceTrans	previousSentence	
previousWordTans	previousWord	
sourceUrl		
subStringTrans	subString	
superStringTrans	superString	parent sentence
'temporarily added this declartion		

표 2 Data property

anchorOf	sourceString
beginIndex	
endIndex	
leftContext	
lemma	
posTag	
rightContext	
stem	

4.3 RDF 변환

상기의 온톨로지와 어휘를 바탕으로 자연언어처리 결과물을 입력으로 받고 RDF triple을 결과로 보이는 시스템을 구현하였다[13]. 그림 6은 이러한 시스템을 도식화한 것이다.

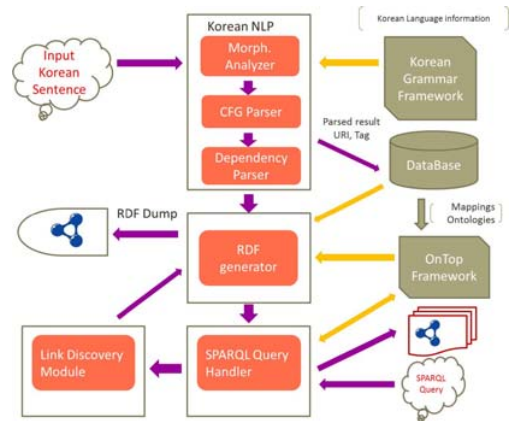


그림 6 전체 시스템 프로토타입

3장에서 구축된 자연언어처리 플랫폼에서 출력되는 자연언어처리 결과물들은 DB에 저장된다. DB는 실제 상황에서 NIF형태가 다음 프로세스로 넘어가므로 필요하지 않아 보이지만, 현재는 하나의 한국어 문장으로부터 RDF triple을 추출하는 시스템을 구현하는 것을 목표로 하고 있기 때문에 우선적으로 DB를 사용하는 시스템을 구축하였다.

DB에 저장된 자연언어처리 결과물은 OnTop Framework에 기반 한 Ontology Based Data Access 모델 [11]을 사용한다. OnTop의 기능은 본 연구에서 크게 두 가지로 사용되는데 첫 번째로 DB의 데이터를 SPARQL 쿼리를 통해 접근가능 하도록 만들어 주는 것. 두 번째로 DB의 데이터를 세종태그셋 온톨로지와 OLiA 온톨로지를 바탕으로 RDF형태로 변환 하는 역할을 한다. 본 시스템으로 출력되는 RDF 결과물들은 본 연구소 홈페이지의 데모사이트[13]에서 확인할 수 있다. 본 데모사이트에서는 위에 설명한 SPARQL 쿼리를 통해서 자연언어처리

결과를 검색하는 시스템도 제공하고 있다.

## 5. 결론 및 향후연구

웹의 발달로 인해 데이터관리에 대한 이슈가 새롭게 강조되고 있다. 현실적으로 대부분의 데이터들이 응용프로그램에 의존하여, 특정 응용프로그램에만 국한되어 사용가능하도록 제공되는 경우가 많다. 자연언어처리의 경우에도 다양한 언어에 대한 다양한 자연언어처리 도구들이 존재하여, 상이한 포맷으로 결과물이 도출되고 있다.

이러한 문제의식을 바탕으로, 하나의 문장으로부터 품사태그와 구구조형태, 의존구조형태의 자연언어처리 결과물을 모두 출력하는 하나의 플랫폼을 구축할 필요성이 있었다. 또한 이렇게 출력된 자연언어처리 결과물들의 국제적 상호운용성을 위하여 RDF로 변환된 출력물을 제공할 필요성도 있었다. 이에 자연언어처리 플랫폼을 구축하고 오픈소스로 공개하였으며, 또한 자연언어처리 결과물을 RDF로의 변환 제공하는 기초 시스템을 구축하였고 데모사이트로 공개하였다. 이 모든 과정은 단순 텍스트 구문분석을 떠나 텍스트 정보에 대한 LOD의 링크 디스커버리를 밝히고 LOD-cloud에 연결가능 하도록 만들었다는데 큰 의미가 있다.

그러나 현재의 시스템은 중간처리단계에서 요구하는 다양한 입출력으로 대규모 텍스트에 대한 자연언어처리 RDF출력에 있어서 빠른 속도를 제공하지 못하는 것이 사실이다. 향후에는 동시에 대규모의 한국어 텍스트에 대한 자연언어처리 결과물을 효율적으로 RDF로 출력하는 도구를 제공하고자 한다. 또한 출력된 RDF triple을 LOD-cloud, 특히 DBpedia와 연결하는 시스템 구현을 연구 중에 있다.

## 사사

본 논문은 지식경제부 산업원천기술개발사업(10035348, 모바일 플랫폼 기반 계획 및 학습 인지 모델 프레임워크 기술 개발)의 지원으로 수행되었음

## 참고문헌

[1] Christian Bizer, Tom Heath, Tim Berners-Lee, Linked Data - The Story So Far, *Journal on Semantic Web and Information Systems (IJSWIS): Special Issue on Linked Data 5(3)* 1-22, 2009

[2] S. Hellmann, J. Lehmann, and S. Auer. Linked-data aware uri schemes for referencing text fragments. In *EKAW 2012, LNAI*. Springer, 2012.

[3] G. Rizzo, R. Troncy, S. Hellmann, and M. Bruemmer. NERD meets NIF: Lifting NLP extraction results to the linked data cloud. In *LDOW*, 2012.

[4] H. Chung, Statistical Korean Dependency Parsing Model based on the Surface Contextual Information, *Ph.D. thesis*, 고려대학교, 2004

[5] 오진영, 차정원, 다단계 구단위화를 이용한 고속 한국어 의존구조 분석, *한국시물레이션학회 논문지*, 2010.03.

[6] J. D. Choi and M. Palmer, Statistical dependency parsing in Korean: From corpus generation to automatic parsing, *In the Second Workshop on SPMRL*, pp. 1-11, 2011

[7] DongHyun Choi, Jungyeul Park, Key-Sun Choi, *Korean Treebank Transformation for Parsr Training*, ACL - SPMRL 2012

[8] 신중호, 한영석, 박영찬, 최기선, 어절구조를 반영한 은닉 마르코프 모델을 이용한 한국어 품사태그, *한글 및 한국어 정보처리 학술대회*, pp. 389-394, 1994.

[9] D. Smith and N. Smith, "Bilingual parsing with factored estimation: Using English to parse Korean", *In Proceedings of EMNLP 2004*, pp. 49-56, 2004

[10] Jinho D. Choi and Martha Palmer. 2010. Robust Constituent-to-Dependency Conversion for English. *In Proceedings of 9th Treebanks and Linguistic Theories Workshop (TLT)*, pages 55 - 66.

[11] Mariano Rodriguez-Muro and Diego Calvanese. Quest, an owl 2 ql reasoner for ontologybased data access. *In Proc. of the 9th Int. Workshop on OWL: Experiences and Directions (OWLED 2012)*, volume 849 of CEUR Electronic Workshop Proceedings, <http://ceur-ws.org/>, 2012.

[12] 박상원, 최동현, 김은경, 최기선, 플러그인 컴포넌트 기반의 한국어 형태소 분석기, *한글 및 한국어 정보처리 학술대회 (HCLT) Poster*, pp. 197-201, 2010.

[13] M. Rezk, J. Park, Y. Youngun, K. Lim, J. Larsen, Y. Hahm, K-S. Choi, Korean Linked Data on the Web: From Text to RDF. *JIST Conference*. 2012 (submitted for publication)

[14] <http://semanticweb.kaist.ac.kr/nlp2rdf>