

규칙 기반 한국어 시간 정보 추출

정영섭⁰¹, 도효진¹, 임준호², 최호진¹
한국과학기술원¹, 한국전자통신연구원²

pinode@kaist.ac.kr, dohyojin90@gmail.com, joonho.lim@etri.re.kr, hojinc@kaist.ac.kr

Rule-Based Temporal Information Extraction for Korean

Young-Seob Jeong⁰¹, Hyo-Jin Do¹, Joon-Ho Lim², and Ho-Jin Choi¹
KAIST¹, ETRI²

요 약

웹을 비롯한 다양한 곳에서 기하급수적으로 증가하고 있는 문서들로 인해, 자연어 텍스트로부터의 지식 추출의 중요성이 점차 커지고 있다. 이 연구에서는 한국어로 작성된 자연어 텍스트로부터의 시간 정보 추출을 위해 개발된 시스템을 소개하고, 직접 구축한 한국어 데이터셋에 대한 성능 분석을 제공한다. 이 시스템은 사람이 직접 작성한 규칙들에 기반하여 작동하지만, 질의응답시스템 등에 적용될 수 있는 수준의 성능으로 향상시키기 위해 기계학습 기반의 시스템으로 업그레이드하는 등의 작업을 계속할 것이다.

주제어: 시간 정보 추출, 규칙 기반 시스템, 한국어 자연어 텍스트

1. 서론

웹을 비롯한 다양한 곳에서 기하급수적으로 증가하고 있는 문서들로 인해, 비정형화된 자연어 텍스트로부터 지식을 자동으로 추출하는 기술 개발의 중요성이 점차 커지고 있다. 추출된 지식은 정보검색, 트렌드 분석, 질의응답시스템 등에 유용하게 사용될 수 있다. 이 연구에서는, 자연어 텍스트로부터의 시간 정보 추출에 초점을 맞추고 있으며, 시간 정보 추출을 위한 기술 개발 또한 그 필요성이 매우 크다. “현재 대한민국의 대통령은 누구인가?” 라는 질문을 예로 들면, 질문하는 시점과 현재의 시간에 대한 정보가 없이는 질의응답시스템이 올바른 답변을 도출해내기 어렵다. 이 연구의 목표는 한국어로 작성된 자연어 텍스트로부터의 자동화된 시간 정보 추출 모듈을 개발하는 것이며, 이를 통해 궁극적으로는 질의응답시스템에서 시간과 관련된 사건 간의 관계 추출 및 추론을 보다 지능적으로 수행할 수 있도록 돕는 것이다. 이를 위해 규칙 기반의 시간정보 추출 모듈을 개발하였으며, 직접 구축한 한국어 데이터셋에 대한 성능 분석을 제공한다.

본 논문의 구성은 다음과 같다. 제 2절에서는 시간정보 추출에 대한 배경 및 관련연구를 소개하고, 제 3절에서는 본 연구에서 개발된 시간정보 추출 시스템에 대하여 설명하며, 제 4절에서는 실험결과에 대한 분석, 그리고 제 5절에서는 결론 및 향후 연구계획을 요약하였다.

2. 배경

2.1 시간 정보

TimeML (Time Meta Language) [1]에서는 시간 표현, 사건 표현, 시간 관계라는 3가지 시간 정보에 대하여 정의하고 있다. 시간 표현은 시간과 관련된 내용을 텍스트

에서 직접 표현하며, 예로는 ‘9월 30일’, ‘어제’ 등이 있다. 사건은 시간과 밀접한 관련을 갖는 경우가 많으므로, 사건 표현은 텍스트에서 사건과 관련한 표현을 하는 경우이고, 그 예로는 ‘공부하다’, ‘태어나다’ 등이 있다. 시간 관계는 시간 표현과 사건 표현들 사이의 관계에 대한 것이다.

시간 표현은 TIMEX3 태그로 표시되며, 사건 표현은 EVENT 태그와 MAKEINSTANCE 태그로 표시되고, 시간 관계는 TLINK, ALINK, SLINK 로 표시된다. 예를 들어, “철수는 어제와 오늘 영어를 가르쳤다.” 는 문장에 시간정보 태깅을 수행할 경우, 시간 표현은 TIMEX3(어제) 와 TIMEX3(오늘)가 태깅되고 사건 표현은 EVENT(가르치다)가 태깅된다. EVENT 태그가 템플릿이라면 MAKEINSTANCE 태그는 템플릿에 의해 생성된 개체의 역할을 하며, 위 예시 문장에서는 한 개의 EVENT(가르치다)에 대하여 두 개의 MAKEINSTANCE(EVENT(가르치다))가 생성된다. 왜냐하면, EVENT(가르치다)는 실제로는 ‘어제’ 와 ‘오늘’ 을 합하여 두 번 발생한 사건이기 때문이다. 시간 관계는 TLINK(TIMEX3(어제) + MAKEINSTANCE(EVENT(가르치다)))와 TLINK(TIMEX3(오늘) + MAKEINSTANCE(EVENT(가르치다)))가 생성된다. 이 예시에서 TLINK 태그는 TIMEX3 태그와 EVENT 태그간의 관계가 아니라 TIMEX3 태그와 MAKEINSTANCE 태그간의 관계이다.

위 3가지 종류의 시간 정보를 추출하기 위한 연구는 TempEval [2-4] 에서 시간 표현 추출(Task A), 사건 표현 추출(Task B), 시간 관계 추출(Task C) 등으로써 여러 연구자들이 모여 주기적으로 경합하고 있으며 TimeML 에 기반한 TimeBank [5] 데이터셋도 제공하고 있다.

2.2 관련 연구

TimeML 의 시간 정보에 대한 정의를 기반으로 시간정보 추출에 대한 많은 연구가 수행되었다. 3가지 종류의 시간 정보를 한꺼번에 추출하는 것은 매우 어려우므로,

많은 연구들이 시간정보의 일부에 초점을 맞춰 수행되었다. 텍스트 상에서 시간 표현인 부분 찾기에 초점을 맞춘 [7]에서는 Begin-Inside-Outside (BIO) 방식을 사용하였고, 규칙에 기반하여 시간 표현을 추출하기도 하였다 [11]. 시간 표현과 사건 표현을 모두 추출하기도 했고 [10], 시간 표현과 시간 관계를 Markov Logic Networks (MLN)을 사용하여 추출하기도 했다 [13].

시간 관계 추출은 시간 표현과 사건 표현이 주어졌을 때 비로소 추출할 수 있다. 사건 표현 간의 시간 관계를 추출하기도 하였으며 [6], 시간 관계 간의 모순을 해결하기도 하였다 [8,9]. EM (Expectation Maximization) 알고리즘에 기반한 연구도 수행되었고 [15], EM 과 BCDC (Bootstrapped Cross-Document Classification)를 사용한 연구도 수행되었다 [16].

TIPSem (Temporal Information Processing based on Semantic) [14] 은 TempEval-2 의 Task 들을 모두 한번에 풀기 위해 소개되었으며 CRF (Conditional Random Fields) 학습에 SR (Semantic Role) 정보를 활용하였다. TimeTrails [12] 는 시간 정보를 UIMA (Unstructured Information Management Architecture)에 기반하여 분석하기 용이한 방식으로 표현하였다.

위 모든 연구들은 영어에 초점을 맞추어 수행되었지만, [17] 는 TempEval-2 의 모든 언어 (영어, 스페인어, 이탈리아어, 중국어, 프랑스어, 한국어)를 대상으로 시간 정보 추출을 수행하였으며 규칙 기반의 K-best parser를 사용하였다. 각 언어의 특성에 맞는 규칙을 설계하지 않는 한, 언어에 따라서 성능이 크게 달라지며, 한국어 데이터셋에 대하여 시간 표현의 타입 예측 성능은 약 82%, 시간 표현 값 예측 성능은 약 42% 였다.

이 연구는 한국어를 대상으로 3가지 종류의 시간정보를 모두 추출하는 규칙 기반 시스템을 소개하며, 직접 구축한 소규모 한국어 데이터셋에 대한 성능 분석을 제공한다.

3. 한국어 시간정보 추출

이 연구는 기본적으로는 TimeML 의 시간정보 정의를 따르지만, 몇 가지 태그와 속성은 배제하였다. 표 1에 이 연구의 시간정보 범위를 표시하였다. 이 연구에서 배제된 태그는 SIGNAL 태그, SLINK 태그, ALINK 태그이며, 배제된 속성은 TLINK 태그의 signalID 속성, TIMEX3 태그의 beginPoint, endPoint, freq 속성, EVENT 태그와 MAKEINSTANCE 태그의 class, aspect, tense, polarity, modality, cardinality, signalID 속성이다. TLINK 태그는 한 문장 안에서의 TIMEX3 태그와 MAKEINSTANCE 태그 사이의 시간 관계로 국한하였다. 배제된 태그 및 속성들은 크게 두 가지 측면에서 문제가 되기 때문에 배제되었다. 첫째, 최종 서비스에서 크게 중요하지 않을 수 있는 너무 자세한 정보를 담는 태그나 속성들은 배제되었다. 예를 들어, signalID 속성은 최종 서비스 관점에서 보면 거의 정보가 없는 속성이나 마찬가지로이다. 둘째, 문장으로부터 시간 정보를 자동으로 추출하는 시스템 개발의 측면에서 보았을 때 비효율적인 속성이나 태그는

배제되었다. 예를 들어, modality, cardinality, signalID 속성 등을 자세하게 알아내는 것은 개발 측면에서 매우 어려운 부분이며, 이런 속성에서 얻는 정보들은 서비스 관점에서 봤을 때에도 무의미한 경우가 많다. 예외적으로, beginPoint, endPoint, freq 속성 등은 이번 연구에서는 일단 배제되었지만, 서비스 관점에서 유용한 정보를 담을 수 있으므로 향후 시스템에 다시 포함할 것이다.

시간정보 3가지를 모두 추출하는 것은 매우 어려운 작업이므로, 본 연구는 3가지 시간정보의 범위를 위와 같이 국한하여 시간정보 추출 시스템을 개발하였지만, 추후 시간정보 범위를 늘려가고 성능 개선을 지속할 것이다.

시간정보 추출 시스템의 입력은 ETRI 언어분석기 [18] 결과이며 그림 1 과 같이 JSON 포맷으로 제공된다. 항목 "text"는 원본 문장을 담고 있고, "morp"는 형태소 분석 결과를 담고 있으며 "word"는 띄어쓰기를 기준한 단어 구분 결과, "NE"는 개체명 인식 결과, "chunk"는 구(chunk) 구분 결과, "dependency"는 의존구문분석 결과를 담고 있다. 이 연구의 시간정보 추출 시스템은 ETRI 언어분석기에서 제공되는 이러한 모든 정보를 입력으로 받는다.

시간정보 추출 시스템의 출력은 그림 2 와 같이 JSON 포맷으로 출력되며, 추출된 시간정보는 "TIME" 항목에 들어있다. "TIME" 항목은 4가지의 하위 항목을 가지고 있는데, "timex3"는 TIMEX3 태그들을 담고 있으며, "event"는 EVENT 태그, "makeinstance"는 MAKEINSTANCE 태그, "tlink"는 TLINK 태그들을 담고 있다. 그림 2에서 "timex3" 에 있는 한 개의 TIMEX3 태그의 type 속성은 'Time'이며 value 속성은 '2014-05-24TAM' 이다. 여기서 value 속성은 주변 문맥 혹은 Document Creation Time (DCT) 와 같은 메타 정보를 활용하여 얻게 된다. 속성 tid 는 TIMEX3 태그의 고유 아이디이며 start 속성과 end 속성은 TIMEX3 태그가 텍스트 상에서 몇 번째 형태소부터 몇 번째 형태소에 걸쳐서 존재하는지를 나타낸다. 이 그림에서는 start 속성이 3 이고 end 속성이 6 이므로, TIMEX3 태그는 '24일오전' 에 걸쳐있으며 이는 value 속성의 값인 '2014-05-24TAM' 과 일치한다. 이 value 속성의 값은 표준 ISO-8601 에 기반하여 표시된다. EVENT 태그는 '발표하' 에 걸쳐있으며 이와 관련한 MAKEINSTANCE 태그가 있다. TLINK 태그는 위 TIMEX3 태그와 MAKEINSTANCE 태그 사이의 시간 관계를 나타내며, 이를 통해 '발표하다' 라는 사건이 2014년 5월 24일 오전에 발생하였음을 표현하게 된다.

이 연구의 시간정보 추출 시스템은 형태소 단위로 태깅을 수행한다. 예를 들어, TIMEX3 태깅은 3단계의 작업을 통해 이루어진다. 첫 번째, 각 형태소에 대하여 시간 표현인지 여부를 판단하여 TIMEX3 태그의 범위에 포함시킨다. 두 번째, TIMEX3 태그들의 type 속성을 결정한다. type 속성은 4가지(TIME, DATE, SET, DURATION) 중에서 하나가 된다. 세 번째, TIMEX3 태그들의 value 속성의 값을 ISO-8601 에 입력한다. EVENT 태그와 MAKEINSTANCE 태그의 태깅도 위와 비슷한 과정을 거쳐서 수행된다. TLINK 태깅은 2단계에 걸쳐 이루어진다. 첫 번째, 태깅

된 TIMEX3 태그들과 MAKEINSTANCE 태그들을 대상으로, 각 쌍에 대하여 TLINK 태그의 범위에 포함되는지 판단한다. 두 번째, TLINK 태그들의 relType 속성을 결정한다. 이 relType 속성은 8가지(BEFORE, AFTER, INCLUDES, DURING, SIMULTANEOUS, IDENTITY, BEGINS, ENDS) 중에서 하나가 된다.

표 1 : 시간정보 추출 시스템에서 추출되는 시간정보 태그 및 속성

태그	속 성	설 명
TIMEX3	tid	태그 아이디
	type	태그 타입
	value	시간 값 (ISO-8601 기반)
EVENT	eid	태그 아이디
MAKE- INSTANCE	eiid	태그 아이디
	eid	관련된 EVENT 태그 아이디
TLINK	lid	태그 아이디
	timeID	첫 번째 인자가 TIMEX3 태그일 때 태그 아이디
	eventInstanceID	첫 번째 인자가 MAKEINSTANCE 태그일 때 태그 아이디
	relatedToTime	두 번째 인자가 TIMEX3 태그일 때 태그 아이디
	relatedTo- EventInstance	두 번째 인자가 MAKEINSTANCE 태그일 때 태그 아이디
	relType	태그 타입 (관계 타입)

```
{
  "category": "",
  "category_weight": 0,
  "title": {"text": "", "NE": ""},
  "sentence": [{"id": 0, "reserve_str": ""},
    "text": "박근혜 대통령은 24일 오전 청와대 춘추관에서 대국민담화를 발표하였다.",
    "morph": [
      {"id": 0, "lemma": "박근혜", "type": "NNP", "position": 1},
      {"id": 1, "lemma": "대통령", "type": "NNP", "position": 11},
      {"id": 2, "lemma": "은", "type": "JK", "position": 20},
      {"id": 3, "lemma": "24", "type": "SN", "position": 24},
      {"id": 4, "lemma": "일", "type": "NNB", "position": 26},
      {"id": 5, "lemma": "오", "type": "NNG", "position": 30},
      {"id": 6, "lemma": "전", "type": "NNP", "position": 33},
      {"id": 7, "lemma": "청와대", "type": "NNP", "position": 37},
      {"id": 8, "lemma": "춘추관", "type": "NNP", "position": 47},
      {"id": 9, "lemma": "에서", "type": "JKB", "position": 56},
      {"id": 10, "lemma": "대국민", "type": "NNG", "position": 63},
      {"id": 11, "lemma": "담화", "type": "NNG", "position": 72},
      {"id": 12, "lemma": "를", "type": "JKO", "position": 78},
      {"id": 13, "lemma": "발표", "type": "NNG", "position": 82},
      {"id": 14, "lemma": "하", "type": "XSV", "position": 88},
      {"id": 15, "lemma": "였", "type": "EP", "position": 91},
      {"id": 16, "lemma": "다", "type": "EF", "position": 94},
      {"id": 17, "lemma": ".", "type": "SF", "position": 97} ],
    "word": [{"id": 0, "text": "박근혜", "type": "", "begin": 0, "end": 0}, ...],
    "NE": [{"id": 0, "text": "대통령", "type": "CV_POSITION",
      "begin": 1, "end": 1, "weight": 0.896012, "common_noun": 0}, ...],
    "chunk": [{"id": 0, "text": "박근혜", "type": "NP", "begin": 0, "end": 2}, ...],
    "dependency": [{"id": 0, "text": "박근혜", "head": 1, "label": "NP", "mod": {}}, ...]
}
```

그림 1 : 시간정보 추출 시스템의 입력 예시. (ETRI 언어분석기 결과)

```
{ "TIME": {
  "timex3": [{"tid": "D2_$S_t1", "type": "TIME",
    "value": "2014-05-24TAM", "start": 3, "end": 6},
  "event": [{"eid": "D2_$S_e1", "start": 13, "end": 14},
  "makeinstance": [{"eiid": "D2_$S_ei1", "eventID": "D2_$S_e1"},
  "tlink": [{"lid": "D2_t0", "eventInstanceID": "D2_$S_ei1",
    "relatedToTime": "D2_$S_t1", "relType": "DURING"}]
}}
```

그림 2 : 시간정보 추출 시스템의 출력 예시

이 연구의 시간정보 추출 시스템은 사람이 직접 작성한 규칙들에 기반하여 작동되며, 이 규칙들의 일부가 그림 3에 표시되어 있다. 그림에서 extent 가 의미하는 것은 태그가 텍스트 상에서 가지는 범위를 의미한다. 예를 들어, EVENT extent 는 EVENT 태그가 텍스트에서 몇 번째 형태소부터 몇 번째 형태소까지 걸쳐있는지를 찾는 작업이고, 비슷하게 TIME extent 는 TIMEX3 태그가 텍스트 상에서 가지는 범위를 의미한다. MORP_x 는 위치 x 에 있는 형태소를 의미한다. 예를 들어, MORP₋₁ 은 현재 태깅을 수행중인 형태소보다 1칸 이전에 위치한 형태소를 의미한다. 비슷하게, POS_x 는 위치 x 에 있는 형태소의 POS (Part-Of-Speech) 태그를 의미하며, NE_x 는 위치 x 에 있는 형태소의 개체명 인식 결과를 의미한다. POS 태그의 기호(PV, NC, XSV 등)와 개체명 인식 결과 기호(DT, TI, DURATION 등)들이 모두 태깅을 위해 자질로써 사용되며, 이는 ETRI 언어분석기의 결과로써 제공된다 [18, 19].

```
* EVENT extent
If POS0==PV
{ If POS0==NC and
  ( MORP0=[하|위하|대하|관하|통하|멈추|] or ( MORP-1==에 and MORP0==따르 ))
  {MORP0==EVENT expression}
  else{ MORP0==EVENT expression}
}
.....
* TIME extent
If NE0=[DT|TI] or MORP0=[이번|현행|과거|이날|당일|한때]
{ MORP0==TIME expression}
.....
* TIME type
TE=Time extent
MORP_TE=a list of morphemes of TE
M:=an arbitrary morpheme of the MORP_TE
If ∃M such that NEM==DURING {Type=Duration}
else if ∃M such that NEM==DT {Type=Date}
else if ∃M such that NEM==TI {Type=Time}
else {Type=Set}
.....
```

그림 3 : 시간정보 추출 시스템의 규칙 일부

4. 실험

실험에 사용할 한국어 데이터셋을 구축하기 위해, TempEval-2에서 제공되는 한국어 데이터셋을 일부 수정하여 소규모로 구축하였다. 구축된 데이터셋은 40개의 문장, 52개의 시간 표현, 191개의 사건 표현, 152개의 시간 관계를 가지고 있다. 이 데이터셋을 대상으로 측정된 시간정보 추출 시스템의 성능이 표 2에 요약되어 있

다. 성능 측정은 F1 measure를 사용하였고, P 는 precision, R 은 recall 값이다. TIMEX3 태그의 type 속성과 value 속성에 대한 추출은 TIMEX3 태그가 텍스트 상에서 위치하는 범위를 맞추지 못하면 제대로 수행될 수 없으므로, type 속성과 value 속성에 대한 성능은 TIMEX3 태그의 텍스트 범위에 대한 성능에 의존적이다. TLINK 태그는 TIMEX3 태그들과 MAKEINSTANCE 태그들 사이의 시간 관계에 대한 것이므로, TIMEX3 태그 추출 성능과 MAKEINSTANCE 태그 추출 성능이 TLINK 태그 추출 성능(아래 표에서 TLINK 태그의 searching links 성능)에 큰 영향을 미친다. 마찬가지로, TLINK 태그의 relType 속성은 TLINK 태그가 올바르게 추출되었을 때 비로소 제대로 속성 값을 예측할 수 있으므로, TLINK 태그의 searching links 성능은 relType 속성에 대한 성능에 영향을 미친다.

시간정보 추출 시스템의 시간 표현 추출 성능은 F1 measure 0.69, 사건 표현 추출 성능은 F1 measure 0.85 이며, 시간 관계 추출 성능은 searching links 와 type 에 대하여 각각 F1 measure 0.61 과 0.34 이다. 현재의 전반적인 성능은 질의응답시스템에 적용되기에는 부족하지만, 추후 기계학습 기반으로 시스템을 업그레이드하며 성능개선을 지속할 것이다. 특히, 성능개선이 이루어져야 할 부분은 precision 보다는 recall 이며 TLINK 태그의 searching link부터 recall 성능이 급감되고 있다. 낮은 성능의 주 원인은 데이터의 부족이다. 데이터가 부족하기 때문에 불충분한 정보를 토대로 규칙이 생성되었기 때문이다. 따라서, 더 많은 데이터를 확보하고, 보다 다양한 상황에서 TIMEX3 태그 및 MAKEINSTANCE 태그들 간의 TLINK 태그 생성에 대한 자질을 발굴하는 작업에 집중할 것이다.

표 2 : 시간정보 추출 시스템 성능

태그	추출 종류	성능
TIMEX3	extent	F1 0.97 (P 0.94, R 1.0)
	type	F1 0.69 (P 0.67, R 0.71)
	value	F1 0.69 (P 0.67, R 0.71)
EVENT	extent	F1 0.85 (P 0.89, R 0.81)
TLINK	searching links	F1 0.61 (P 0.82, R 0.48)
	type	F1 0.34 (P 0.79, R 0.22)

5. 결론

이 논문에서는 한국어 텍스트로부터 규칙 기반으로 시간정보를 추출하는 시스템을 소개하였다. 이 시간정보 추출 시스템은 기본적으로 TimeML 에 기반하여 시간 표현, 사건 표현, 그리고 시간 관계를 모두 추출하였으며, 구축한 한국어 데이터셋에 대한 실험을 수행하였다. 궁극적으로 한국어 기반의 질의응답 시스템에 적용될 수 있는 수준의 시간정보 추출 시스템을 개발하기 위해 연구를 지속할 것이다.

6. 감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 산업융합원천기술개발사업(정보통신)의 일환으로 수행하였음. [10044577, 휴먼 지식증강 서비스를 위한 지능진화형 WiseQA 플랫폼 기술 개발]

참고문헌

- [1] Pustejovsky, J., Castano, J., Ingria, R., Sauri, R., Gaizauskas, R., Setzer, A., Katz, G., "TimeML: Robust Specication of Event and Temporal Expressions in Text", in Proc. of the 5th Int'l Workshop on Computational Semantics, Tilburg, Netherlands, pp.1-11, 2003.
- [2] Verhagen, M., Gaizauskas, R.J., Schilder, F., Hepple, M., Moszkowicz, J., Pustejovsky, J., "The TempEval Challenge: Identifying Temporal Relations in Text", Language Resources and Evaluation, 43(2), pp.161-179, 2009.
- [3] Verhagen, M., Sauri, R., Caselli, T., Pustejovsky J., "SemEval-2010 task 13: TempEval-2", in Proc. of the 5th Int'l Workshop on Semantic Evaluation, Uppsala, Sweden, pp.57-62, 2010.
- [4] UzZaman, N, Llorens, H, Derczynski, L., Verhagen, M., Allen, J., Pustejovsky, J., "SemEval-2013 Task 1: TEMPEVAL-3: Evaluating Time Expressions, Events, and Temporal Relations", in Proc. of the 7th Int'l Workshop on Semantic Evaluation, Atlanta, Georgia, 2013.
- [5] Pustejovsky, J., Hanks, P., Sauri, R., See, A., Day, D., Ferro, L., Gaizauskas, R., Lazo, M., Setzer, A., Sundheim, B., "The Timebank Corpus", Corpus Linguistics, pp.647-656, 2003.
- [6] Chambers, N., Wang, S., Jurafsky, D., "Classifying Temporal Relations Between Events", in Proc. of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, Prague, Czech Republic, pp.173-176, 2007.
- [7] Poveda, J., Surdeanu, M., Turmo, J., "A Comparison of Statistical and Rule-Induction Learners for Automatic Tagging of Time Expressions in English", in Proc. of the 14th Int'l Symposium on Temporal Representation and Reasoning, Alicante, Spain, pp.141-149, 2007.
- [8] Chambers, N., Jurafsky, D., "Jointly Combining Implicit Constraints Improves Temporal Ordering", in Proc. of the Conference on Empirical Methods in Natural Language Processing, Hawaii, USA, pp.698-706, 2008.
- [9] Yoshikawa, K., Riedel, S., Asahara, M., Matsumoto, Y., "Jointly Identifying Temporal

- Relations with Markov Logic", in Proc. of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th Int'l Joint Conference on Natural Language Processing of the AFNLP, Suntec, Singapore, pp.405-413, 2009.
- [10] UzZaman, M., Allen, J., "Event and Temporal Expression Extraction from Raw Text: First Step towards a Temporally Aware System", Int'l Journal of Semantic Computing, 4(4), pp.487-508, 2010.
- [11] Strotgen, J., Gertz, M., "HeidelTime: High Quality Rule-based Extraction and Normalization of Temporal Expressions", in Proc. of the 5th Int'l Workshop on Semantic Evaluation (SemEval), Uppsala, Sweden, pp.321-324, 2010.
- [12] Strotgen, J., Gertz, M., "TimeTrails: A System for Exploring Spatio-Temporal Information in Documents", in Proc. of the VLDB Endowment, Singapore, pp.1569-1572, 2010.
- [13] Ling, X., Weld, D.S., "Temporal Information Extraction", in Proc. of the 24th AAAI Conference Artificial Intelligence, Atlanta, USA, 2010.
- [14] Llorens, H., Saquete, E., Navarro, B., "TIPSem (English and Spanish): Evaluating CRFs and semantic roles in TempEval-2", in Proc. of the 5th Int'l Workshop on Semantic Evaluation, Uppsala, Sweden, pp.284-291, 2010.
- [15] Mirroshandel, S.A., Ghassem-Sani, G., "Temporal Relation Extraction Using Expectation Maximization", in Proc. of Recent Advances in Natural Language Processing, Hissar, Bulgaria, pp.218-225, 2011.
- [16] Mirroshandel, S.A., Ghassem-Sani, G., "Towards Unsupervised Learning of Temporal Relations between Events", Journal of Artificial Intelligence Research, 45(1), pp.125-163, 2012.
- [17] Angeli, G., Uszkoreit, J., "Language-Independent Discriminative Parsing of Temporal Expressions", in Proc. of the 51th Annual Meeting of the Association for Computational Linguistics, Soa, Bulgaria, pp.83-92, 2013.
- [18] 임준호, 윤여찬, 배용진, 김현기, 이규철, "지배소 후위 제약을 적용한 트랜지션 시스템 기반 한국어 의존 파싱 모델", 한글 및 한국어 정보처리 학술대회, 2014.
- [19] Korean Language Institute, "Sejong Treebank", <http://www.sejong.or.kr>