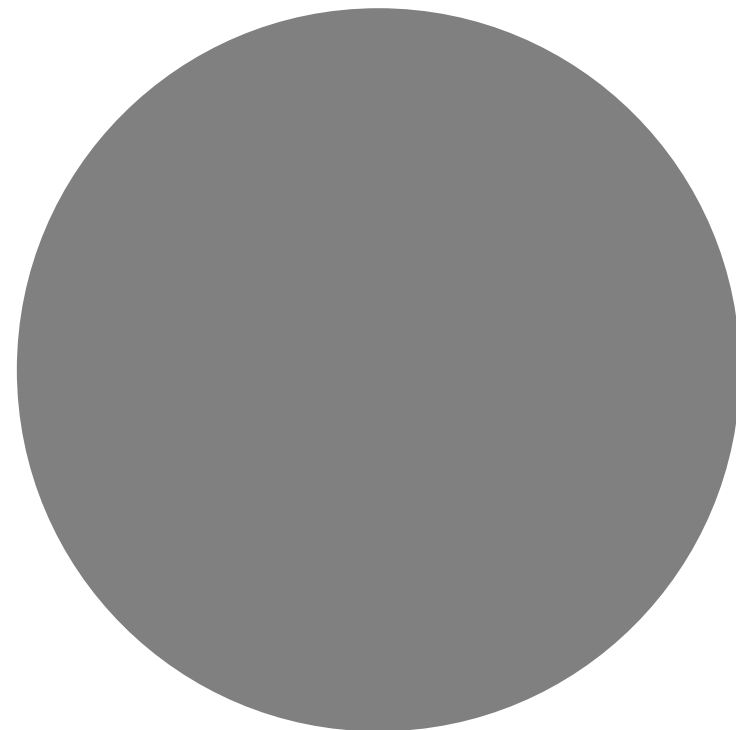


Open API 활용 세미나

인공지능 연구소 C'est La Vie 이준호



본 발표는 Etri의 공식 입장이 아닌
한 명의 API 사용자의 주관적인 생각이 담겨져 있음을 미리 밝힙니다.

이후에 나오는 소스 코드는 Python 3.0을 기준으로 작성되었습니다.

언어처리

- 언어 분석 기술
 - 형태소 분석 성능 분석
 - 개체명 분석
- 질의응답 기술
 - 위키백과 QA API
- ETRI API 활용 사례

Python

python 3.0을 기준으로 작성되었습니다.

HTTP 통신을 위해 urllib3 라이브러리를 사용하여 제공하고 있습니다. Python 3.0 이하의 버전에서 예제를 실행하기 위해서는 별도로 urllib3의 설치가 필요합니다. 설치에 대한 설명은 <https://pypi.python.org/pypi/urllib3> 를 참고하시기 바랍니다. urllib3 라이브러리에 대한 자세한 설명은 <https://urllib3.readthedocs.io/en/latest/> 에서 확인 하실 수 있습니다.

```
#-*- coding:utf-8 -*-
import urllib3
import json

openApiURL = "http://aiopen.etri.re.kr:8000/WiseNLU"
accessKey = "YOUR_ACCESS_KEY"
analysisCode = "ANALYSIS_CODE"
text = "YOUR_SENTENCE"

requestJson = {
    "access_key": accessKey,
    "argument": {
        "text": text,
        "analysis_code": analysisCode
    }
}

http = urllib3.PoolManager()
response = http.request(
    "POST",
    openApiURL,
    headers={"Content-Type": "application/json; charset=UTF-8"},
    body=json.dumps(requestJson)
)

print("[responseCode] " + str(response.status))
print("[responBody]")
print(str(response.data, "utf-8"))
```

<http://aiopen.etri.re.kr/guide wiseNLU.php#group06>

형태소 분석기 성능 비교

| Hannanum(ntags=22) | | Mecab-ko(ntags=43) | | Komoran(ntags=45) | | Eri API(ntags = 45) | | Kheili(ntags = 45) | | Kma(ntags=56) | |
|--------------------|-------------|--------------------|-------------|-------------------|-------------|---------------------|-------------|--------------------|-------------|---------------|---------------------|
| Tag | Description | Tag | Description | Tag | Description | Tag | Description | Tag | Description | Tag | Description |
| NC | 보통명사 | NNG | 일반 명사 | NNG | 일반 명사 | NNG | 일반명사 | NNG | 일반명사 | NNG | 보통명사 |
| NQ | 고유명사 | NNP | 고유 명사 | NNP | 고유 명사 | NNP | 고유명사 | NNP | 고유명사 | NNP | 고유명사 |
| NB | 의존명사 | NNB | 의존 명사 | NNB | 의존 명사 | NNB | 의존명사 | NNB | 의존명사 | NNB | 일반 의존 명사 |
| NN | 수사 | NR | 수사 | NR | 수사 | NR | 수사 | NR | 수사 | NNM | 단위 의존 명사 |
| NP | 대명사 | NP | 대명사 | NP | 대명사 | NP | 대명사 | NP | 대명사 | NP | 대명사 |
| PV | 동사 | VV | 동사 | VV | 동사 | VV | 동사 | VV | 동사 | VV | 동사 |
| PA | 형용사 | VA | 형용사 | VA | 형용사 | VA | 형용사 | VA | 형용사 | VA | 형용사 |
| PX | 보조 용언 | VX | 보조 용언 | VX | 보조 용언 | VX | 보조용언 | VX | 보조용언 | VXV | 보조 동사 |
| | | VCP | 궁정 지정사 | VCP | 궁정 지정사 | VCP | 궁정 지정사 | VCP | 궁정 지정사 | VXA | 보조 형용사 |
| | | VCN | 부정 지정사 | VCN | 부정 지정사 | VCN | 부정 지정사 | VCN | 부정 지정사 | VCP | 궁정 지정사, 서술격 조사 '이다' |
| | | | | | | | | | | VCN | 부정 지정사, 형용사 '아니다' |
| | | | | | | | | | | MDN | 수 관형사 |
| MM | 관형사 | MM | 관형사 | MM | 관형사 | MMN | 지시 관형사 | MM | 관형사 | MDT | 일반 관형사 |
| | | MAQ | 일반 부사 | MAQ | 일반 부사 | MMN | 주 관형사 | MM | 일반 부사 | MAQ | 일반 부사 |
| MA | 부사 | MAJ | 접속 부사 | MAJ | 접속 부사 | MAJ | 접속 부사 | MAJ | 접속 부사 | MAC | 접속 부사 |
| II | 감탄사 | IC | 감탄사 | IC | 감탄사 | IC | 감탄사 | IC | 감탄사 | IC | 감탄사 |
| | | JKS | 주격 조사 | JKS | 주격 조사 | JKS | 주격 조사 | JKS | 주격 조사 | JKS | 주격 조사 |
| | | JKC | 보격 조사 | JKC | 보격 조사 | JKC | 보격 조사 | JKC | 보격 조사 | JKC | 보격 조사 |
| | | JKG | 관형격 조사 | JKG | 관형격 조사 | JKG | 관형격 조사 | JKG | 관형격 조사 | JKG | 관형격 조사 |
| | | JKO | 목적격 조사 | JKO | 목적격 조사 | JKO | 목적격 조사 | JKO | 목적격 조사 | JKO | 목적격 조사 |
| | | JKB | 부사격 조사 | JKB | 부사격 조사 | JKB | 부사격 조사 | JKB | 부사격 조사 | JKB | 부사격 조사 |
| | | JKV | 호격 조사 | JKV | 호격 조사 | JKV | 호격 조사 | JKV | 호격 조사 | JKV | 호격 조사 |
| | | JKQ | 인용격 조사 | JKQ | 인용격 조사 | JKQ | 인용격조사 | JKQ | 인용격 조사 | JKQ | 인용격 조사 |
| JC | 격조사 | JC | 접속 조사 | JC | 접속 조사 | JC | 접속 조사 | JC | 접속 조사 | JC | 접속 조사 |
| JX | 보조사 | JX | 보조사 | JX | 보조사 | JX | 보조사 | JX | 보조사 | JX | 보조사 |
| JP | 서술격 조사 | | | | | | | | | | |
| EP | 선어말어미 | EP | 선어말어미 | EP | 선어말어미 | EP | 선어말 어미 | EP | 선어말 어미 | EPH | 존칭 선어말 어미 |
| | | | | | | | | | | EPT | 시제 선어말 어미 |
| | | | | | | | | | | EPF | 종류 선어말 어미 |
| | | | | | | | | | | EPN | 형식어 종결 어미 |
| | | | | | | | | | | EFO | 의문형 종결 어미 |
| | | | | | | | | | | EFQ | 명령형 종결 어미 |
| | | | | | | | | | | EFA | 칭용형 종결 어미 |
| | | | | | | | | | | EFB | 간접형 종결 어미 |
| EF | 종결 어미 | EF | 종결 어미 | EF | 종결 어미 | EF | 종결 어미 | EF | 종결 어미 | EFJ | 존칭형 종결 어미 |
| | | | | | | | | | | ECE | 대등 연결 어미 |
| | | | | | | | | | | ECS | 보조적 연결 어미 |
| EC | 연결 어미 | EC | 연결 어미 | EC | 연결 어미 | EC | 연결어미 | EC | 연결 어미 | ECD | 의존적 연결 어미 |
| | | ETN | 명사형 전성 어미 | ETN | 명사형 전성 어미 | ETM | 명사형 전성 어미 | ETM | 명사형 전성 어미 | ETN | 명사형 전성 어미 |
| ET | 전성 어미 | ETM | 관형형 전성 어미 | ETM | 관형형 전성 어미 | ETM | 관형형 전성 어미 | ETM | 관형형 전성 어미 | ETD | 관형형 전성 어미 |
| XP | 접두사 | XPN | 제언 접두사 | XPN | 제언 접두사 | XPN | 제언 접두사 | XPN | 제언 접두사 | XPN | 제언 접두사 |
| | | | | | | | | | | XPV | 용언 접두사 |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |
| | | | | | | | | | | | |

형태소 분석기 N-tags

Hannanum 9

Okt 19

Hannanum 22

Mecab-ko 43

Komoran 45

(세종 품사태그와 동일)

Etri_API 45

Kkma 56

인공지능 연구소 C'est La Vie 이준호

형태소 분석기 성능 비교

| Komoran(ntags=45) | | | Etri API(ntags = 45) | | |
|-------------------|-------------|--|----------------------|-------------|----|
| Tag | Description | | Tag | Description | |
| VCP | 긍정 지정사 | | VCP | 긍정 지정사 | VC |
| VCN | 부정 지정사 | | VCN | 부정 지정사 | VC |
| MM | 관형사 | | MMA | 성상 관형사 | MC |
| | | | MMD | 지시 관형사 | |
| | | | MMN | 수 관형사 | MC |
| MAG | 일반 부사 | | MAG | 부사 | MA |
| MAJ | 접속 부사 | | MAJ | 접속 부사 | MA |
| | | | | | |
| SN | 숫자 | | SN | 숫자 | ON |
| NF | 명사추정범주 | | NA | 분석 불능 범주 | UN |
| NV | 용언추정범주 | | | | |
| NA | 분석불능범주 | | | | |
| | | | | | |
| | | | | | |

Komoran 과 Etri API의 차이

하늘을 나는 자동차

| Hannanum | | Kkma | | Komoran | | Etri | | Mecab | | Okt | | Khaiii | |
|----------|---|------|-----|---------|-----|------|-----|-------|-----|-----|------|--------|-----|
| 하늘 | N | 하늘 | NNG | 하늘 | NNG | 하늘 | NNG | 하늘 | NNG | 하늘 | Noun | 하늘 | NNG |
| 을 | J | 을 | JKO | 을 | JKO | 을 | JKO | 을 | JKO | 을 | Josa | 을 | JKO |
| 나 | N | 날 | VV | 나 | NP | 나 | VV | 나 | NP | 나 | Noun | 나 | NP |
| 는 | J | 는 | ETD | 는 | JX | 는 | ETM | 는 | JX | 는 | Josa | 는 | JX |
| 자동차 | N | 자동차 | NNG | 자동차 | NNG | 자동 | NNG | 자동차 | NNG | 자동차 | Noun | 자동차 | NNG |
| | | | | | | 차 | NNG | | | | | | |
| | | | | | | | | | | | | | |

형태소 분석기 성능 비교(띄어쓰기 문제)

아버지가방에들어가신다

| Hannanum | | Kkma | | Komoran | | Etri | | Mecab | | Okt | | Khaiii | |
|-------------------|---|------|-----|---------|-----|------|-----|-------|-------|-----------|------|--------|-----|
| 아버지 가방에 들어가 | N | 아버지 | NNG | 아버지 | NNG | 아버지 | NNG | 아버지 | NNG | 아버지 | Noun | 아버지 | NNG |
| | | 가방 | NNG | 가방 | NNP | 가방 | NNG | 가 | JKS | 가방 | Noun | 가 | JKS |
| | | 에 | JKM | 에 | JKB | 에 | JKB | 방 | NNG | 에 | Josa | 방 | NNG |
| 이 | J | 들어가 | VV | 들어가 | VV | 들어가 | VV | 에 | JKB | 들어가 신다 | Verb | 에 | JKB |
| 시ㄴ다 | E | 시 | EPH | 시 | EP | 시 | EP | 들어가 | VV | | | 들어가 | VV |
| | | ㄴ다 | EFN | ㄴ다 | EC | ㄴ다 | EF | 신다 | EP+EC | | | 시 | EP |
| | | | | | | | | | | | | ㄴ다 | EC |

형태소 분석기 성능 비교(신조어 분석)

트와이스 신곡 필 스페셜 집중조명

| Hannanum | | Kkma | | Komoran | | Etri | | Mecab | | Okt | | Khaiii | |
|----------|---|------|-----|----------|-----|----------|-----|-------|------------|----------|------|----------|-----|
| 트와 이스 | N | 트와 | UN | 트와 이스 | NNP | 트와 이스 | NNG | 트 | NNG | 트와 이스 | Noun | 트와 이스 | NNP |
| | | 이스 | UN | | | | | 와 | JC | | | | |
| 신곡 | N | 신곡 | NNG | 신곡 | NNP | 신곡 | NNG | 이스 | NNP | 신곡 | Noun | 신곡 | NNG |
| 필 | P | 필 | NNG | 필 | NNP | 필 | NNP | 신곡 | NNG | 필 | Noun | 필 | NNG |
| 르 | E | 스페셜 | NNG | 스페셜 | NNP | 스페셜 | NNG | 필 | VV+ ETM | 스페셜 | Noun | 스페셜 | NNG |
| 스페셜 | N | 집중 | NNG | 집중 | NNG | 집중 | NNG | 스페셜 | NNP | 집중 | Noun | 집중 | NNG |
| 집중 조명 | N | 조명 | NNG | 조명 | NNP | 조명 | NNG | 집중 | NNG | 조명 | Noun | 조명 | NNG |
| | | | | | | | | 조명 | NNG | | | | |

형태소 분석기 성능 비교(신조어 분석)

오늘 시험은 완전 멘붕이었는데, 완전 케바케더라

| Hannanum | | Kkma | | Komoran | | Etri | | Mecab | | Okt | | Khaiii | |
|------------|---|------|-----|---------|-----|-----------|-----|-------|------------|-----|------|----------|-----|
| 멘붕이 었는데 | N | 메 | VV | 멘붕 | NNP | 멘붕이 | NNG | 멘붕 | NNP | 멘붕 | Noun | 멘붕 | NNP |
| | | ㄴ | ETD | 게 | EC | 케바케 더라 | NNP | 케 | IC | 케바케 | Noun | 케바케 더 | NNG |
| 케바케 | N | 붕 | MAG | 바 | NNB | | | 바케 | NNP | 더 | Noun | 라 | EC |
| 이 | J | 게 | ECD | 하 | XSA | | | 더라 | VCP +EC | 라 | Josa | | |
| 더라 | E | 바 | NNG | 게 | EC | | | | | | | | |

[('오늘', 'N'),
('시험', 'N'),
('은', 'J'),
('완전', 'N'),
('멘붕이었는데', 'N'),
('완전', 'N'),
('케바케', 'N'),
('이', 'J'),
('더라', 'E')]

[('오늘', 'NNG'),
('시험', 'NNG'),
('은', 'JX'),
('완전', 'NNG'),
('메', 'VV'),
('ㄴ', 'ETD'),
('붕', 'MAG'),
('미', 'VCP'),
('였', 'EPT'),
('는데', 'ECD'),
(' ', 'SP'),
('완전', 'NNG'),
('하', 'XSA'),
('게', 'ECD'),
('바', 'NNG'),
('하', 'XSV'),
('게', 'ECD'),
('더', 'MAG'),
('라', 'NNG')]

[('오늘', 'NNG'),
('시험', 'NNG'),
('은', 'JX'),
('완전', 'NNG'),
('멘붕', 'NNP'),
('미', 'NNP'),
('미', 'VCP'),
('였', 'EP'),
('는데', 'EC'),
(' ', 'SP'),
('완전', 'NNG'),
('하', 'XSA'),
('게', 'EC'),
('바', 'NNB'),
('하', 'XSA'),
('게', 'EC'),
('덜', 'VV'),
('라', 'EC')]

[('오늘', 'NNG'),
('시험', 'NNG'),
('은', 'JX'),
('완전', 'NNG'),
('멘붕이', 'NNG'),
('미', 'VCP'),
('였', 'EP'),
('는데', 'EC'),
(' ', 'SP'),
('완전', 'NNG'),
('덜', 'VV'),
('케바케더라', 'NNP')]

형태소 분석기 성능 비교(실행시간)

결과물 : [(단어, 형태소 태깅)] (ex. [(token, tags)])

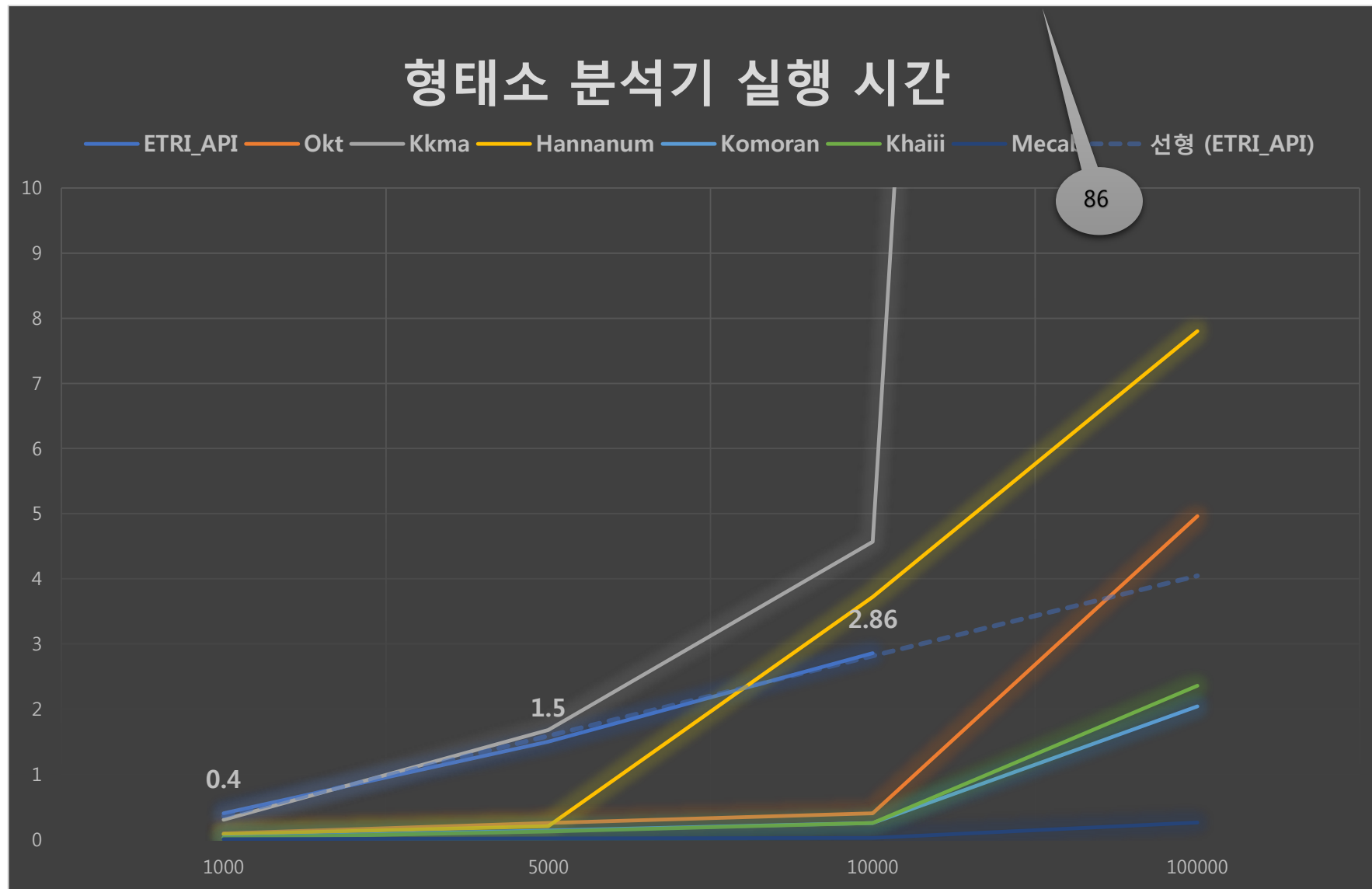
```
In [124]: etri_api(text)
Out[124]: [('출범', 'NNG'),
            ('과', 'JC'),
            ('성장', 'NNG'),
            ('위키백', 'NNP'),
            ('과', 'JKB'),
            ('는', 'JX'),
            ('2001', 'SN'),
            ('년', 'NNG'),
            ('1', 'SN'),
            ('월', 'NNG'),
            ('15', 'SN'),
            ('일', 'NNG'),
            ('서비스', 'NNG'),
            ('를', 'JKO'),
            ('개시', 'NNG'),
            ('하', 'XSV'),
            ('었', 'EP'),
            ('다', 'EF'),
            ('.', 'SF'),
            ('도메인', 'NNG').]
```

ETRI_API

```
In [18]: twitter.pos(text)
Out[18]: [('출범', 'Noun'),
            ('과', 'Josa'),
            ('성장', 'Noun'),
            ('위키', 'Foreign'),
            ('위키', 'Noun'),
            ('백', 'Suffix'),
            ('과는', 'Josa'),
            ('2001년', 'Number'),
            ('1월', 'Number'),
            ('15일', 'Number'),
            ('서비스', 'Noun'),
            ('를', 'Josa'),
            ('개시', 'Noun'),
            ('하였다', 'Verb'),
            ('.', 'Punctuation'),
            ('도메인', 'Noun'),
            ('은', 'Josa'),
            ('www.wikipedia.com', 'URL'),
            ('을', 'Josa'),
            ('사용', 'Noun').]
```

Okt

형태소 분석기 성능 비교



■ 개체명 인식(1/3)

개체명 인식 Input → Output

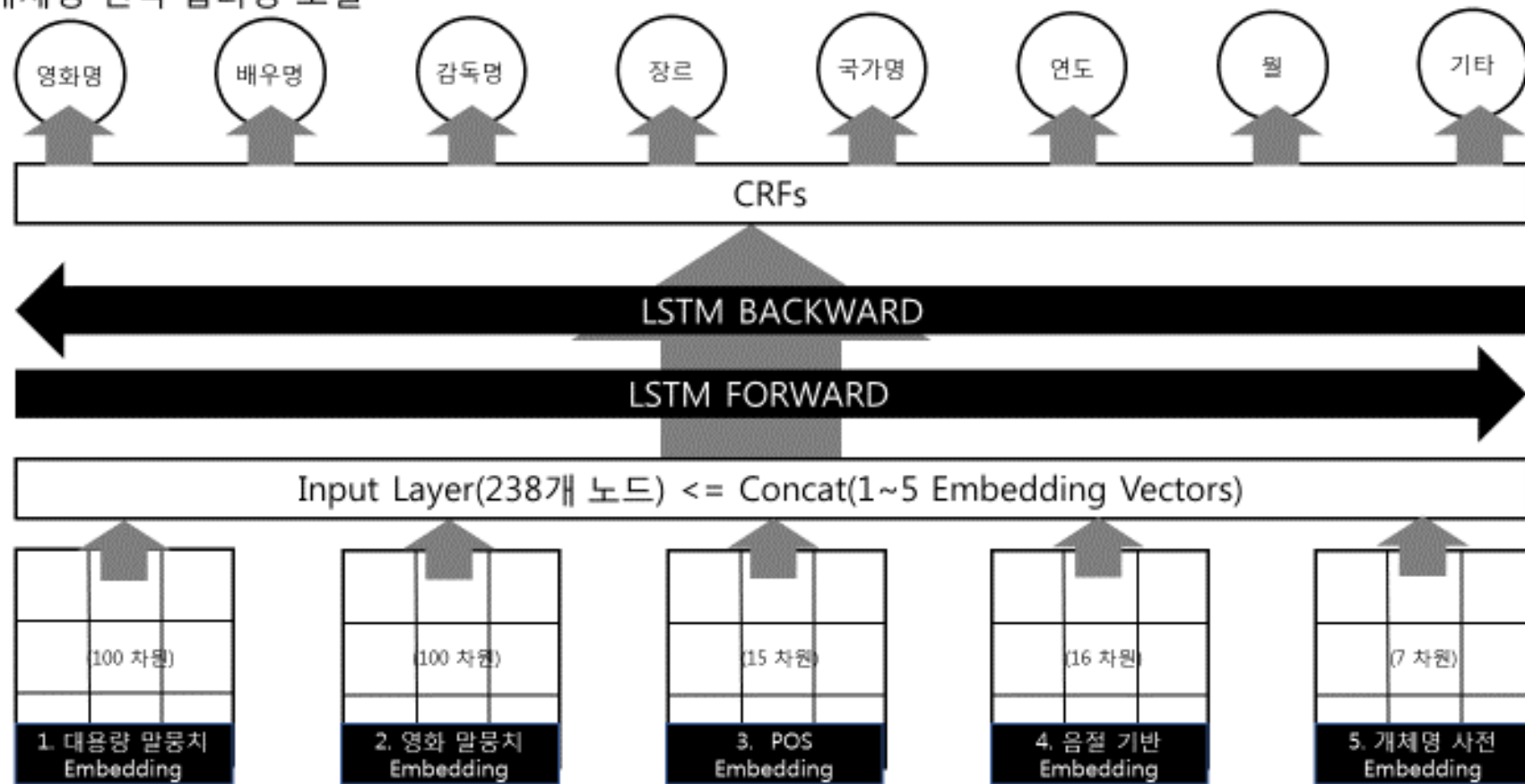
'2017년에 개봉한 미국 스릴러 영화 알려줘'

개체명 인식 Model

{ 'actor': [], 'country': ['미국'], 'director': [], 'genre': ['스릴러'],
'month': [], 'movie': [], 'year': ['2017년'] }

■ 개체명 인식(2/3)

개체명 인식 딥러닝 모델



- 참고 논문: 송치윤 외, 「개선된 워드 임베딩 모델과 사전을 이용한 Bidirectional LSTM CRF 기반의 한국어 개체명」, 한국소프트웨어종합학술대회 논문집, 2017.



대분류 : 15개

하위 분류 : 총 146개

Q : 광도원이 등장한 액션 영화가 뭐야?

| No | Word | 형태소 | | | 동음이의어 | | | 개체명 | | |
|----|---------|------------------|-----------------------|---|------------------|-----------------------|----------------------|-------|---------|-----------------------|
| | | 단어 | 태그 | 설명 | 단어 | 형태소 태그 | 의미번호 | 단어 | 태그 | 설명 |
| 0 | 광도원이 | 광도원 이 | NNG JKS | 일반명사 주격조사 | 광도원 이 | NNG JKS | 00 00 | 광도원 | PS_NAME | 사람 이름 |
| 1 | 등장 한 | 등장 하 ㄴ | NNG XSV ETM | 일반명사 동사파생 접미사 관형형전 성어미 | 등장하 ㄴ | VV ETM | 01 00 | | | |
| 2 | 액션 | 액션 | NNG | 일반명사 | 액션 | NNG | 00 | 액션 영화 | FD_ART | 예술관련 학문 분야 및 학파 |
| 3 | 영화 가 | 영화 가 | NNG JKS | 일반명사 주격조사 | 영화 가 | NNG JKS | 01 00 | | | |
| 4 | 뭐야? | 뭐 이 야 ? | NP VCP EF SF | 대명사 긍정지정 사 종결어미 마침표, 물음표, 느낌표 | 뭐 이 야 ? | NP VCP EF SF | 00 01 00 00 | | | |

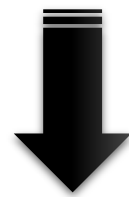
ETRI_API_CALL

Q : 태양계에서 가장 큰 행성 이름이 뭐야?

태양계

태양계(太陽系, 독일어: Sonnensystem 조넨쥬스뎀, 이탈리아어: Sistema solare 시스테마 솔라레, 영어: Solar System 솔라 시스템)는 대략 46억 년 전 분자운의 중력붕괴로 형성된 태양과, 항성인 태양의 중력에 이끌려 있는 주변 천체가 이루는 체계를 말한다. 태양을 중심으로 공전하는 행성은 소행성대를 기준으로 안쪽에 있는 네 개의 고체 행성인 수성, 금성, 지구, 화성, 즉 지구형 행성과, 바깥쪽에 있는 네 개의 유체 행성인 목성, 토성, 천왕성, 해왕성, 즉 목성형 행성으로 알려져 있다. 행성 외에도...(생략)

| Rank | Answer | confidence |
|------|--------|------------|
| 1 | 목성 | 0.763373 |
| 2 | 달 | 0.374684 |
| 3 | 명왕성 | 0.2953 |



“목성”

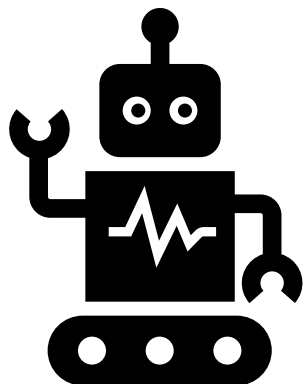
위키 백과 QA(문제)

가스파초는 어느나라
음식이야?

에스파냐

가스파초는 어느나라
음식이야?

??



음식점 챗봇

예약하기

다음주 월요일 강남역지점 7시 20분에 30명 이준호로 예약해줘



WiseNLU (언어분석기술 API)

개체명
인식

DATE (DT)

LOCATION (LC)

TIME (TI)

QUANTITY (QT)

PERSON (PS)

DT : 다음주 월요일

LC : 강남역

TI : 7시 20분

QT : 30명

PS : 이준호

ETRI API 활용 사례

```
def etri_ner_api(text):
    requestJson = {
        "access_key" : ACCESSKEY,
        'argument' : {
            "text" : text,
            "analysis_code" : 'ner'
        }
    }
    http = urllib3.PoolManager()
    response = http.request("POST", URL, headers={
        "Content-Type": "application/json; charset=UTF-8"
    }, body = json.dumps(requestJson))
    data = json.loads(str(response.data, "utf-8"))['return_object']['sentence'][0]['NE']

    entity = [(data[i]['text'], data[i]['type'][:2]) for i in range(len(data))]

    print(entity)
```

```
etri_ner_api("다음주 월요일 강남역 7시 20분에 30명 이준호로 예약해줘")
```

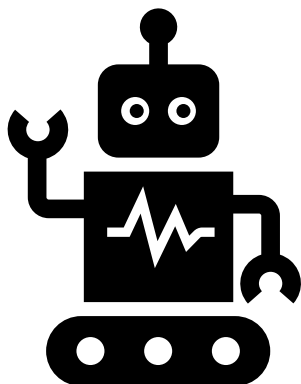
```
[('다음주 월요일', 'DT'), ('강남역', 'LC'), ('7시 20분', 'TI'), ('30명', 'QT'), ('이준호', 'PS')]
```

예약하기

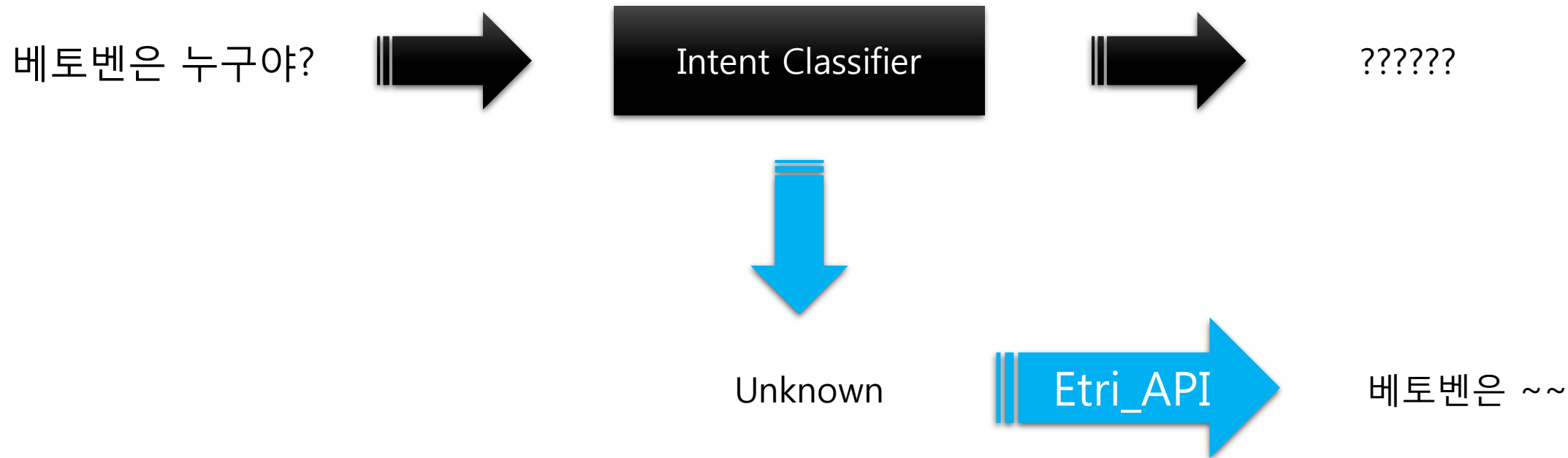
강남역점 7시 20분에 예약해줘

WiseNLU (언어분석기술 API)

DT : ???
LC : 강남역
TI : 7시 20분
QT : ???
PS : ???



예약하시는 날짜 와
방문하시는 인원수 와
예약하시는 분의 성함
을 알려주세요



ETRI

박은정, 조성준, “[KoNLPy: 쉽고 간결한 한국어 정보처리 파이썬 패키지](#)”, 제 26회 한글 및 한국어 정보처리 학술대회 논문집, 2014.
개체명 인식 API [[TTA 표준 개체명 태그셋 \(TTAK.KO-10.0852\)](#)]

Khایی

[[황용주 2016](#)] 21세기 세종 말뭉치 제대로 살펴보기
[[한경은 2017](#)] 공개와 협업을 통한 세종 형태 분석 말뭉치 오류 개선 방법

Q/A