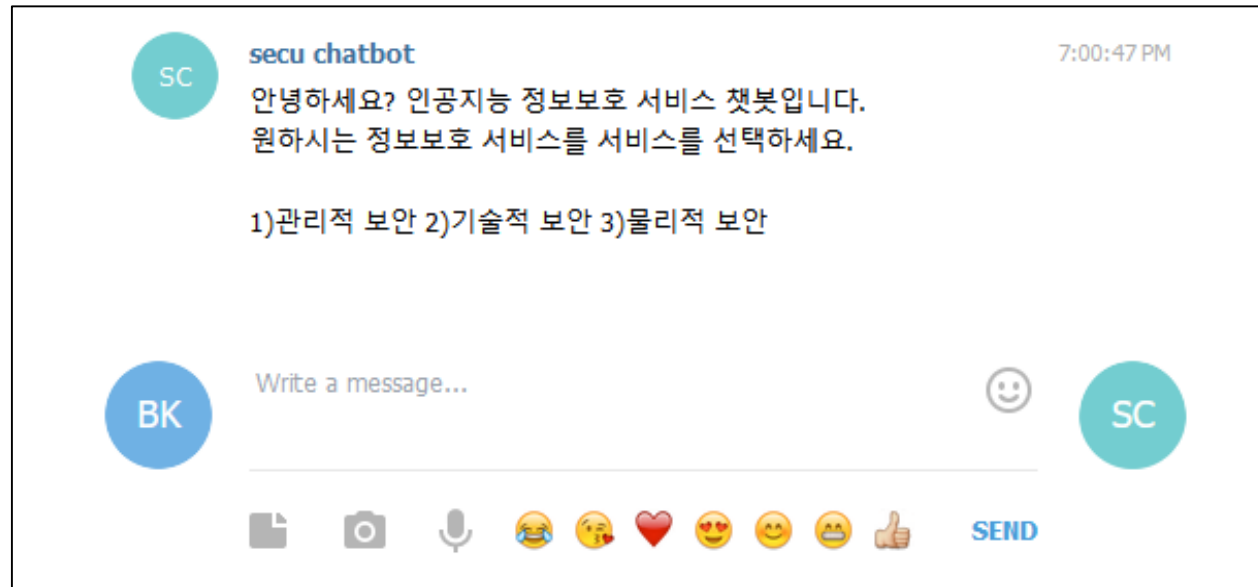


- 웰컴(Welcome) 메시지

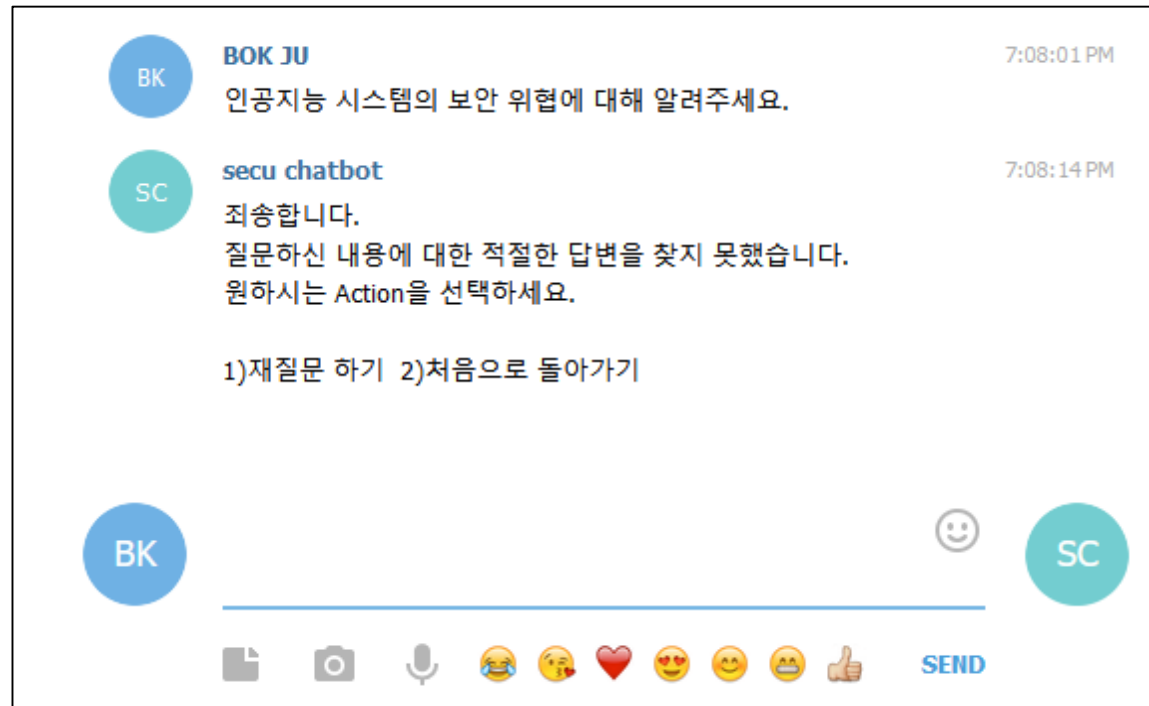
사용자가 챗봇 시스템에 처음 접속했을 때 챗봇이 사용자에게 보내는 환영 메시지

Welcome Block, Welcome Intent 라고도 함



- 폴백(Fallback) 메시지

챗봇이 사용자의 의도를 파악하지 못하거나, 질문에 대한 적절한 답변을 생성하지 못한 경우 등 챗봇이 대화의 흐름을 자연스럽게 이어가기 어려운 상황을 대비한 메시지



- 인텐트(Intent)
자연어 대화를 전달하기 위한 목적과 의도, 화자의 자연어 발화의도
인텐트 파악의 정확도 답변의 품질과 챗봇의 성능에 결정적 영향을 미치는 요소

자연어	인텐트
안녕, 안녕하세요, 잘 있었어	인사하기
안전한 해쉬 알고리즘은? 어떤 해쉬 알고리즘을 사용해야 하지?	해쉬 알고리즘 질문
SHA-256 이상 사용하면 됩니다.SHA-256, SHA-384, SHA-512	해쉬 알고리즘 정보 제공
그만하자, 안녕히 계세요. 잘 있어	끝인사

- 인텐트(Intent)

정확한 인텐트 파악을 위해 의미, 도메인/클래스 및 컨텍스트 고려 필요

규칙 기반(인텐트 정의 및 검색), 학습 기반(인텐트 분류 학습), 하이브리드(규칙 + 학습) 등 존재

자연어	인텐트	비고
햄버거로 주문합니다	BURGER_ORDER	의미
햄버거가 먹고 싶어요		
해쉬 적용 방법이 궁금합니다	DB_JOIN_HASH	도메인/클래스
	SECURITY_PASSWORD	
	SECURITY_INTEGRITY	
버거 사세요	BURGER_SELL	컨텍스트
	BURGER_BUY	

▪ 엔티티(Entity)

도메인 대화모델에서 사용자와 시스템이 주고 받는 개념 및 속성(구성요소)

Entity 는 크게 Class 와 Slot 으로 구분
Class 로 특정 개념을 정의하고, Slot 으로 Class 의 속성 또는 구성요소 등을 정의
날씨 개념 Weather Class 에 City, Date, Info 슬롯(Slot) 들을 정의
복잡한 도메인에서는 여러 개의 Class 를 정의

엔티티(Entity)		설명
클래스(Class)	슬롯(Slot)	
날씨(Weather)	도시(City)	서울, 부산, 대구, 제주 등의 지역 날씨 정보
날씨(Weather)	날짜(Date)	오늘, 내일, 2020년 12월 25일 등의 날짜의 날씨 정보
날씨(Weather)	날씨 정보(Info)	맑음, 흐림, 비 등의 날씨 예보 정보

■ 슬롯 채우기(Slot filling)

정확한 답변 생성을 위해 모든 슬롯이 다 채워질 때 까지 질문을 반복하여 수행

Chat interface showing a conversation between BOK JU and secu chatbot.

BOK JU (8:32:16 PM): 햄버거를 주문하려고 합니다.

secu chatbot (8:37:51 PM): 주문하고자 하는 햄버거를 선택하세요.

1) 불고기 버거 2) 새우 버거 3) 치킨 버거

BOK JU (8:38:02 PM): 1

secu chatbot (8:38:10 PM): 불고기 버거의 수량을 입력하세요.

BOK JU (8:38:14 PM): 2

secu chatbot (8:38:20 PM): 사이드 메뉴를 선택하세요.

1) 음료 2) 파스타 3) 포테이토

BK (Write a message...)

SC

SEND

Chat interface showing a conversation between BOK JU and secu chatbot.

secu chatbot (8:38:29 PM): 음료의 종류를 선택하세요.
1) 콜라 2) 사이다 3) 커피

BOK JU (8:38:36 PM): 1

secu chatbot (8:38:42 PM): 배달지 주소를 입력하세요.

BOK JU (8:38:52 PM): 서울시 마포구 상암동 1585

secu chatbot (8:38:57 PM): 버거종류 : 불고기 버거 2
사이드 메뉴 : 콜라 2
배달지 주소 : 서울시 마포구 상암동 1585

주문하시겠습니까?

1) 주문 및 결제 2) 주문 취소

BK (Write a message...)

SC

SEND

- 개체명 인식(NER, Named Entity Recognition)

입력된 코퍼스로 부터 개체명을 추출하여 인명(person), 지명(location), 기관명(organization), 시간(time) 등으로 분류하여 인식

형태소 분석기를 이용한 방법, OpenAPI를 이용한 방법, 딥러닝을 학습을 이용한 방법 등이 존재

- 형태소 분석기를 이용한 NER

Mecab 등과 같은 형태소 분석기와 딕셔너리를 이용하여 개체명 인식

간단하게 적용할 수 있으나, 딕셔너리에 존재하지 않는 개체명에 대한 추가 및 지속적 업데이트 필요

- OpenAPI를 이용한 NER

공공 인공지능 오픈 API·DATA 서비스 포털(<http://aiopen.etri.re.kr>)에서 제공하는 Open API를 이용하는 방법

5,000건/1일 이내, 1회 사용시 입력은 1만글자 이하로 제한, 오프라인에서는 사용 불가

- 딥러닝 학습을 이용한 NER

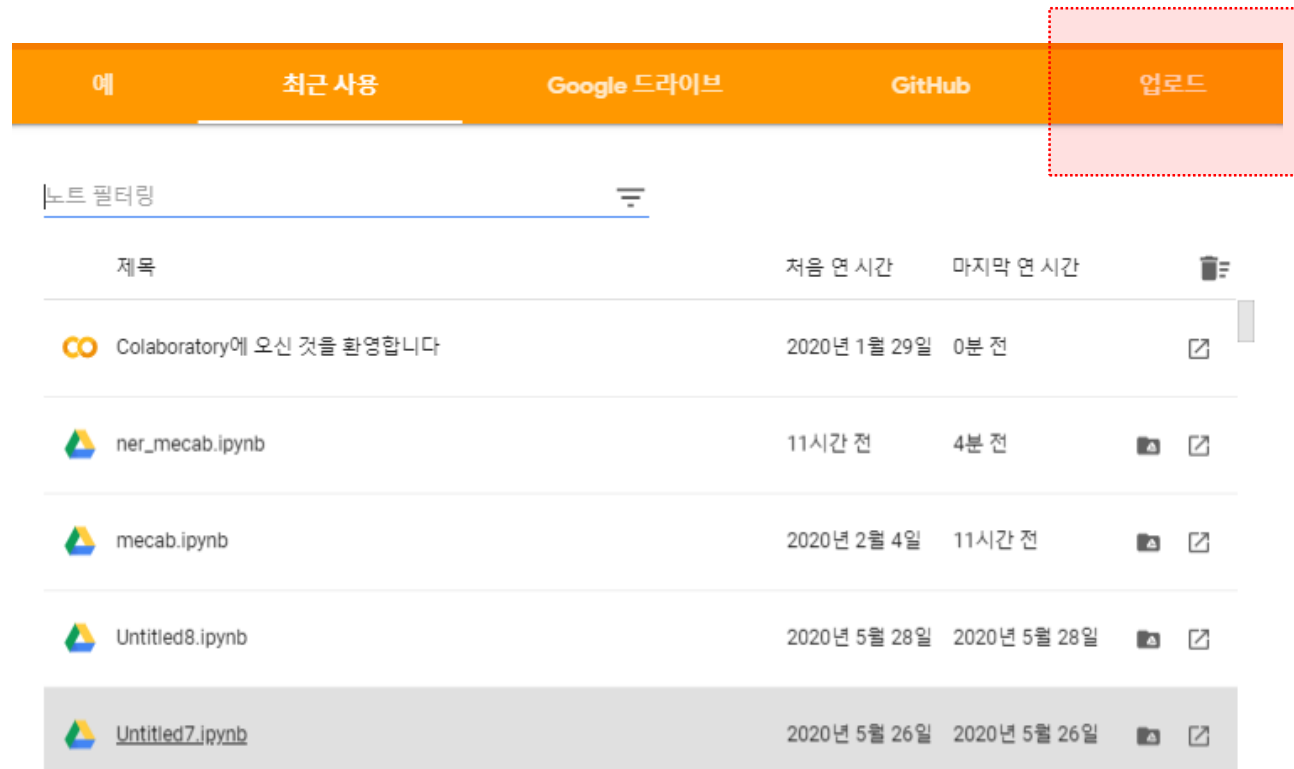
딥러닝 학습을 통한 개체명 분류, Bi-directional LSTM과 CRF(Conditional Random Field) 모델을 이용한 개체명 인식

- google Colab 접속

google Cloud python 개발환경인 Colab 접속

<https://colab.research.google.com/>

업로드를 클릭한 후
ner_mecab.ipynb 파일
업로드



새 노트

취소

- google Colab에 konlpy 및 mecab 설치

Colab에서 konlpy, mecab, mecab-ko-dic, mecab-python 설치하기 위한 스크립트 실행

```
[26] !set -x #  
    && pip install konlpy #  
    && curl -s https://raw.githubusercontent.com/konlpy/konlpy/master/scripts/mecab.sh | bash -x
```

출처 : https://zetawiki.com/wiki/Colab%EC%97%90_konlpy_%2B_mecab_%EC%84%A4%EC%B9%98

Colab은 리눅스 환경이므로 Mecab 사용 가능

- Konlpy 및 Mecab import

```
from konlpy.tag import Mecab  
from MeCab import Tagger
```

- Mecab 객체 생성

```
mecab = Mecab()
```

- 분석할 text에서 명사 추출

```
text = '스티브잡스가 미국 캘리포니아에 있는 스탠포드대학에서 졸업 연설을 했다'  
nouns = mecab.nouns(text)  
print(nouns)
```

➡ ['스티브잡스', '미국', '캘리포니아', '스탠포드', '대학', '졸업', '연설']

- 개체명 인식을 위한 딕셔너리 위치 지정 및 Tagger 객체 생성

```
tagger = Tagger('-d /usr/local/lib/mecab/dic/mecab-ko-dic')
```

- 리스트를 text 형태로 변환

```
nouns = ' '.join(nouns)
```

```
print(nouns)
```

- 분석할 text에서 명사 추출

```
text = '스티브잡스가 미국 캘리포니아에 있는 스탠포드대학에서 졸업 연설을 했다'
```

```
nouns = mecab.nouns(text)
```

```
print(nouns)
```

📄 스티브잡스 미국 캘리포니아 스탠포드 대학 졸업 연설

- Mecab의 Tagger로 파싱 수행

```
parsed_data = tagger.parse(nouns)
```

```
print(parsed_data)
```

```
text = '스티브잡스가 미국 캘리포니아에 있는 스탠포드대학에서 졸업 연설을 했다'
```

```
스티브잡스 NNP, 인명, F, 스티브잡스, Preanalysis, NNP, NNP, 스티브/NNP/인명+잡스/NNP/인명
미국 NNP, 지명, T, 미국, *, *, *, *
캘리포니아 NNP, 지명, F, 캘리포니아, *, *, *, *
스탠포드 NNP, 인명, F, 스탠포드, *, *, *, *
대학 NNG, *, T, 대학, *, *, *, *
졸업 NNG, 행위, T, 졸업, *, *, *, *
연설 NNG, 행위, T, 연설, *, *, *, *
EOS
```

```
[36] from konlpy.tag import Mecab
      from MeCab import Tagger
```

```
[37] mecab = Mecab()
```

```
[38] text = '스티브잡스가 미국 캘리포니아에 있는 스탠포드대학에서 졸업 연설을 했다'
      # text = '방탄소년단은 한국 가수 최초로 영국 웨블리 스타디움에서 공연을 했다'
```

```
[47] nouns = mecab.nouns(text)
      print(nouns)
```



[48] ner_mecab.ipynb 실습

```
[49] nouns = ' '.join(nouns)
      print(nouns)
```



스티브잡스 미국 캘리포니아 스탠포드 대학 졸업 연설

```
[ ] parsed_data = tagger.parse(nouns)
      print(parsed_data)
```

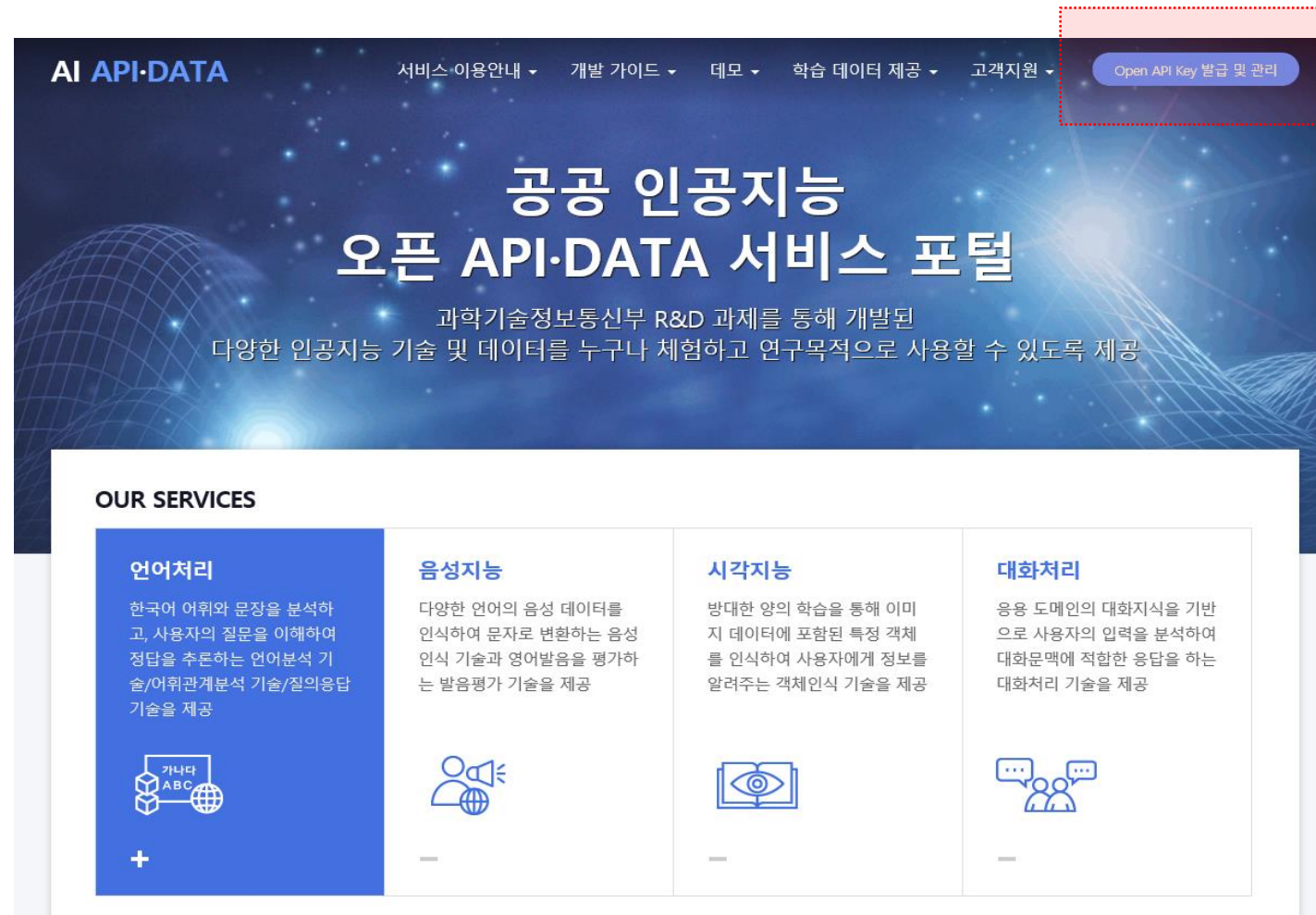


```
스티브잡스      NNP,인명,F,스티브잡스,Preanalysis,NNP,NNP,스티브/NNP/인명+잡스/NNP/인명
미국            NNP,지명,T,미국,*,*,*,*
캘리포니아      NNP,지명,F,캘리포니아,*,*,*,*
스탠포드        NNP,인명,F,스탠포드,*,*,*,*
대학           NNG,*,T,대학,*,*,*,*
졸업           NNG,행위,T,졸업,*,*,*,*
연설           NNG,행위,T,연설,*,*,*,*
EOS
```

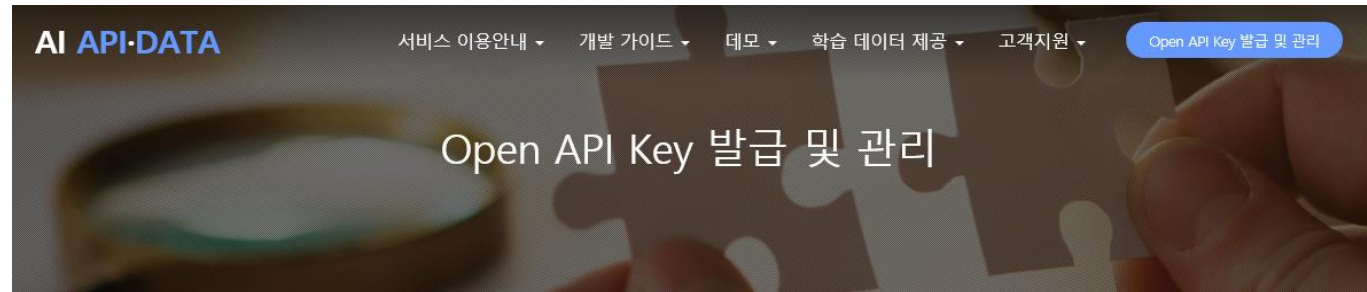
- OpenAPI를 이용하여 NER을 수행 위해 OpenAPI 키 발급

<http://aiopen.etri.re.kr> 접속

클릭한 후
OpenAPI 발급 화면 이동

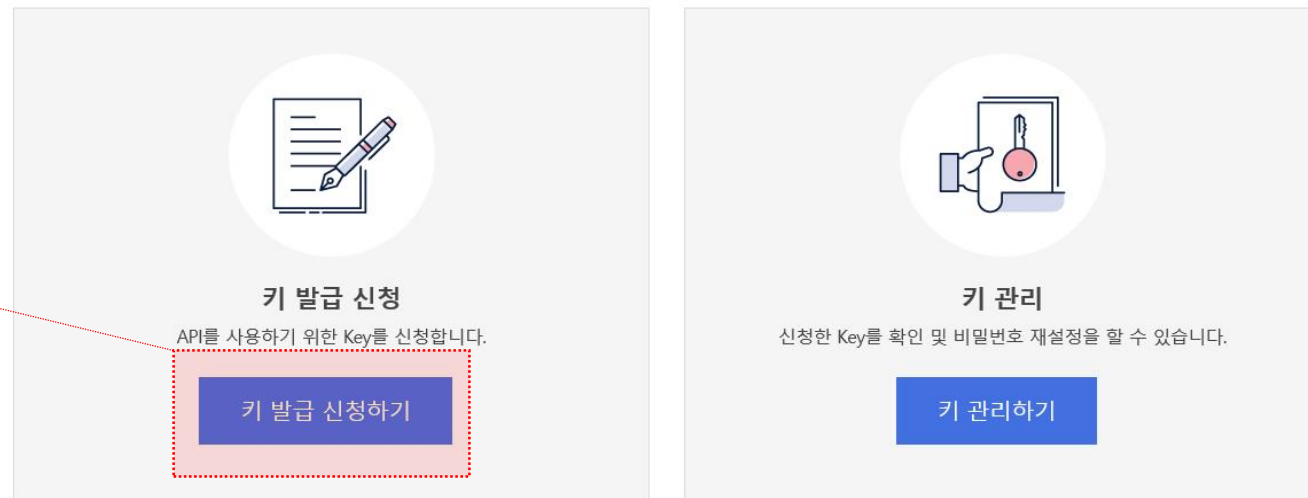


- OpenAPI를 이용한 NER을 위해 OpenAPI 키 발급



Open API Key 발급 및 관리

📍 Open API Key 발급 및 관리



클릭한 후
키 발급 신청화면
이동

▪ OpenAPI를 이용한 NER을 위해 OpenAPI 키 발급

AI API·DATA

서비스 이용안내 ▾ 개발 가이드 ▾ 데모 ▾ 학습 데이터 제공 ▾ 고객지원 ▾ Open API Key 발급 및 관리

STEP1.
동의 및 인증

STEP2.
정보 입력

STEP3.
신청 완료

STEP1. 동의 및 인증

개인정보취급방침

1. 개인정보의 처리 목적

'ETRI 인공지능 Open API 서비스 홈페이지'(aiopen.etri.re.kr)는(은) 개인정보를 다음의 목적을 위해 처리합니다. 처리한 개인정보는 다음 목적 이외의 용도로는 사용되지 않으며 이용 목적이 변경될 시에는 사전 동의를 구할 예정입니다.

가. 홈페이지 회원가입 및 관리

회원 가입의사 확인, Open API 서비스 제공에 따른 본인 식별·인증, 회원자격 유지·관리, 제한적 본인 확인제 시행에 따른 본인확인, 서비스 부정이용 방지, 각종 고지·통지, 고충처리, 분쟁 조정을 위한 기록 보존 등을 목적으로 개인정보를 처리합니다.

나. 재화 또는 서비스 제공

API 서비스 제공, 콘텐츠 제공 등을 목적으로 개인정보를 처리합니다.

☒ 상기 내용을 읽었으며 개인정보취급방침에 동의합니다. (필수)

이메일 인증

* 는 필수 항목입니다.

이메일 입력 *

인증메일 받기

인증키 입력 *

인증키 확인

이메일 인증

* 는 필수 항목입니다.

이메일 입력 *

인증메일 받기

인증키 입력 *

인증키 확인


이메일 입력 후
인증메일 받기 클릭

이메일 인증키 수신 후
인증키 입력


- OpenAPI를 이용한 NER을 위해 OpenAPI 키 발급

키 발급 신청


📍 Open API Key 발급 및 관리 > 키 발급 신청



STEP1.
동의 및 인증



STEP2.
정보 입력



STEP3.
신청 완료

STEP2. 정보 입력

사용자 기본 정보

* 는 필수 항목입니다.

패스워드 *	<input type="password"/> <small>① 패스워드는 4~12자리 숫자와 영문자 특수문자 사용이 가능합니다.</small>
패스워드 확인 *	<input type="password"/>
이름 *	<input type="text"/>
전화번호 *	<input type="text"/> - <input type="text"/> - <input type="text"/>
소속기관 분류 *	<small>① 통계정보 수집을 위해 키발급 신청자를 소속별로 구분하여 관리하고 있기 때문에 개인 목적으로 사용하는 경우에도 소속기관 정보를 여쭙보고 있으니 확인을 부탁드립니다.</small> <input checked="" type="radio"/> 대기업 <input type="radio"/> 중소기업 <input type="radio"/> 대학교 <input type="radio"/> 연구기관(민간) <input type="radio"/> 정부/공공기관 <input type="radio"/> 개인사용자 <input type="radio"/> 기타

정보입력 후
API 키 발급 신청 완료

- OpenAPI를 이용한 NER을 위해 OpenAPI 키 발급

API 키 수신 이메일 수신

OpenAPI에서 사용할
API Key

안녕하세요. _____ 님,

공공 인공지능 오픈 API-DATA 서비스에 관심을 가져주셔서 감사드립니다.

회원님께서 신청하신 API Key 정보는 아래와 같습니다.

신청자 이메일: | _____@gmail.com

API Key: 4409fdb6-0408-48b6-9119-1cf245| _____

발급일: 2020-06-15 13:48:10

이제부터 발급받으신 API Key를 사용하여 인공지능 Open API 서비스를 이용하실 수 있습니다.

발급된 API Key는 제3자가 이용할 수 없습니다.

- OpenAPI를 호출을 위한 모듈 import

```
import urllib3  
import json
```

- OpenAPI URL, Open API Key, NER API

```
openApiURL = "http://aiopen.etri.re.kr:8000/WiseNLU"  
accessKey = "4409fdbe-0408-48b6-9119-1cf245"  
analysisCode = "ner"
```

- 분석할 text

```
text = '스티브잡스가 미국 캘리포니아에 있는 스탠포드대학에서 졸업 연설을 했다'
```

- OpenAPI를 호출을 request, response

```
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)  
)
```

■ 개체명 인식 API

공공 인공지능 오픈 API·DATA 서비스 포털 개체명 인식 API의 개체명 태그셋은 15개 대분류 및 146개 세분류로 구성된 TTA 표준 개체명 태그셋 (TTAK.KO-10.0852)을 사용

세분류	정의
PS_NAME	사람 이름
LC_OTHERS	LC계열의 세부 유형이 아닌 기타 장소
LCP_COUNTRY	국가명
LCP_PROVINCE	도, 주 지역명
LCP_COUNTY	군, 면, 읍, 리, 동 등과 같은 세부 행정구역명
LCP_CITY	도시명
LCP_CAPITALCITY	수도명
LCG_RIVER	강, 호수, 연못
LCG_OCEAN	해양, 바다
LCG_BAY	반도, 만
LCG_MOUNTAIN	산, 산맥, 능선, 고개/재, 봉
LCG_ISLAND	섬, 제도
LCG_CONTINENT	대륙
LC_TOUR	관광명소
LC_SPACE	천체 명칭
OG_OTHERS	기타 기관/단체
OGG_ECONOMY	경제 관련 기관/단체, 기업
OGG_EDUCATION	교육 기관/단체, 교육관련 기관

- TTA 표준 개체명 태그셋 (TTAK.KO-10.0852)을 pandas로 저장한 Pickle 파일 로드

```
import pandas as pd
entity = pd.read_pickle('entity.pkl')
entity.head(5)
```

	entity_key	entity_name
0	AFW_ART_CRAFT	[미술 작품명(그림/미술품)]
1	AFW_DOCUMENT	[도서/서적 작품명, 지리서, 지도, 의학 관련 서적, 종교 서적, 각 종교의 경전...]
2	AFW_MUSIC	[음악 작품명]
3	AFW_PERFORMANCE	[춤/무용 작품명 및 춤 종류, 연극/공연명/가극]
4	AFW_VIDEO	[영화 작품명, TV 프로그램 이름]

- TTA 표준 개체명 태그셋 (TTAK.KO-10.0852) 개체명 리턴 함수

```
def getEntity(key):
    return entity[entity.entity_key==key].tail(1).values[0]
```

- OpenAPI의 response를 json 파싱을 통해 출력

```
rep_data = json.loads(str(response.data,"utf-8"))
for ne in rep_data['return_object']['sentence'][0]['NE']:
    print(ne['text'], ne['type'], getEntity(ne['type']))
```

스티브잡스 PS_NAME ['PS_NAME' list(['사람 이름'])]

미국 LCP_COUNTRY ['LCP_COUNTRY' list(['국가명'])]

캘리포니아 LCP_PROVINCE ['LCP_PROVINCE' list(['도, 주 지역명'])]

스탠포드대학 OGG_EDUCATION ['OGG_EDUCATION' list(['교육 기관/단체, 교육관련 기관'])]

```
import pandas as pd
entity = pd.read_pickle('entity.pkl')
entity.head(5)
```

	entity_key	entity_name
0	AFW_ART_CRAFT	[미술 작품명(그림/미술품)]
1	AFW_DOCUMENT	[도서/서적 작품명, 지리서, 지도, 의학 관련 서적, 종교 서적, 각 종교의 경전...
2		
3	AF	
4		

ner_openapi.ipynb 실습

```
def getEntity(key):
    return entity[entity.entity_key==key].tail(1).values[0]
```

```
rep_data = json.loads(str(response.data,"utf-8"))
for ne in rep_data['return_object']['sentence'][0]['NE']:
    print(ne['text'], ne['type'], getEntity(ne['type']))
```

```
스티브잡스 PS_NAME ['PS_NAME' list(['사람 이름'])]
미국 LCP_COUNTRY ['LCP_COUNTRY' list(['국가명'])]
캘리포니아 LCP_PROVINCE ['LCP_PROVINCE' list(['도, 주 지역명'])]
스탠포드대학 OGG_EDUCATION ['OGG_EDUCATION' list(['교육 기관/단체, 교육관련 기관'])]
```