



Step-1 - 데이터 수집

<https://kin.naver.com/people/expert/index.nhn?type=DOCTOR&orgId=2> (네이버 하이닥)에서 트리 첫단 10개 관련 게시물들 크롤링 필요. 크롤링된 데이터의 정리 형태는 다음과 같아야 함.

	A	B	C	D	E	F
1	질병명(query)	병원분류	질문제목	질문내용	답변제목	답변내용
2	외이도염곰팡이	이비인후과	외이도염 곰팡이들은 귀안에서 뭘 먹고 크는 거예요	None	최혁기 의사 님 답변	(생략)

 외이도염 곰팡이들은 귀안에서 뭘 먹고 크는 거예요?


 이비인후과

 비공개 · 16시간 전 · 조회수 35

A 1개

채택순 ▾ 전체보기 ▾ **최적** 원문

 **최혁기 의사 님 답변**
지안이비인후과의원 · 전문의

Step-1 - Embedding & Classifier

1. 띄어쓰기 보정용 의학 기사들 1만건 정도가 크롤링되면 좋음
(뉴스 크롤링 코드는 내가 만들어둠)
2. SentencePiece로 띄어쓰기/오타 robust한 POS 생성
3. 앞 단계에서 수집된 지식인 질문 데이터랑 위 의학 기사 1만건이랑 섞어서 embedding
4. 트리 첫 단으로 넣어줄 Classifier 제작

Step-2 - 트리 구조 챗봇 구현

“tree_example.ipynb”참고해서, str(input()) 함수 사용해서 실시간으로 트리 구조 활용할 수 있게 구현. CSV파일에서 tree 읽어오는건 내가 깃헙에 굴러다니는 코드 대충 수정해서 갖다둬. (tree_maker.py의 csv_to_jsontree함수)

```
Font Size Edit Code.ipynb x tree_example.ipynb x
+ ✂ 📄 📌 ▶ ■ ↺ Code v

[2]: import numpy as np, pandas as pd
     from tree_maker import csv_to_jsontree

[3]: csv_to_jsontree('disease_example.csv')

[3]: [{ 'name': '열이 나나요?',
      'children': [{ 'name': '예',
                    'children': [{ 'name': '<end>',
                                  'children': [{ 'name': '</end>',
                                                  'children': [{ 'name': '', 'children': [{ 'name': '' } ] } ] } ] } ] } ] } ] },
    { 'name': '아니오',
      'children': [{ 'name': '<end>',
                    'children': [{ 'name': '</end>',
                                  'children': [{ 'name': '', 'children': [{ 'name': '' } ] } ] } ] } ] } ] },
    { 'name': '언제부터아팠나요?',
      'children': [{ 'name': '1일',
                    'children': [{ 'name': '<end>',
                                  'children': [{ 'name': '</end>',
                                                  'children': [{ 'name': '', 'children': [{ 'name': '' } ] } ] } ] } ] } ] },
    { 'name': '1-7일',
```