

## Fasttext : 페이스북에서 개발한 자연어 처리 라이브러리

- Word2Vec의 확장으로 Word2Vec는 단어를 쪼갤 수 없지만 Fasttext는 하나의 단어 안에  
도 여러 단어가 존재하는 것으로 간주하여 학습함.
- Word2Vec는 등장 빈도 수가 적은 단어는 임베딩 정확도가 높지 않고 Fasttext는 빈도 수  
가 적어도 정확도가 높음
- 한국어에 적용하고자 하는 시도들 존재 : 초성, 중성, 종성 분리하여 처리

## 가짜뉴스 판별

1. 제목과 본문이 부적합한 경우
2. 본문 중 맥락에 관계없는 내용 존재 경우

ref.딥러닝 기법을 이용한 가짜뉴스 탐지

## 가짜뉴스 판별 결과 라벨링하는 기술들

### 1. Politifact

- Poynter(언론인 배출을 위한 미디어 교육기관)에서 운영하는 팩트 체크 사이트
- 진실 정도를 6단계로 나눠 라벨링
  - 1) True : 주장이 정확하고, 설명이 충분함
  - 2) Mostly True : 주장은 정확하지만 추가 설명이 필요함
  - 3) Half True : 주장이 부분적으로 정확하고, 추가 설명이 없거나, 문맥이 약간 이상함
  - 4) Mostly False : 진실을 포함하고 있지만 정작 중요한 내용은 무시함
  - 5) False : 주장이 정확하지 않음
  - 6) Pants On Fire(새빨간 거짓말) : 주장과 근거가 정확하지 않고 터무니 없음
- 라벨링 기준
  - ㉠ 기사의 내용이 사실인가?
  - ㉡ 다른곳에서 기사를 접할 방법이 있는가? 해당 기사의 해석이 개방적인가?
  - ㉢ 기사내용에 대한 증거가 있는가? 그 증거가 사실인가?
  - ㉣ 과거에 비슷한 기사내용을 어떻게 다루었는가? 해당 기사의 내용에 대해 판단 할 수 있는  
규정이 있는가?

### 2. Snopes

- 진실 정도를 12단계로 나눠 결과 도출  
: True, Mixture, False, Mostly True, Outdated, Misattributed, Mostly False,  
Miscaptioned, Scam, unproven, Legend, Correct Attribution

### 3. Fiskkit

- 기사 링크를 입력하면 기사 안 문장에 대한 태그가 달림
- 태그도 기준에 따라 다른 형태로 나타남  
: True/False(사실), Fallacy(오류), Descriptive(편파적인 표현), Complimentary

## BRENDA : 브라우저에 설치하여 팩트체크를 수행할 수 있는 프로그램

### 가짜뉴스 판별

1. 팩트체크할 내용을 온라인에서 탐색하여 진실성 확인
  2. DNN 모델 사용
  3. 팩트체크 후 증거 자료 제공
- Snopes, PolitiFact와 같은 대부분의 팩트체크 사이트는 수동 팩트체크를 수행함.
  - CredEye는 일부 자동화된 팩트체크 시스템을 사용함
  - CredEye는 단어만 탐색하여 팩트체크(저작권 등)
  - **BRENDA**는 단어와 문장 모두에서 증거를 제공할 수 있음
  - 기사의 주제, 작성자, 내용 등 각 측면을 바라보고 팩트체크
  - FactMata라는 자동화 팩트체크 시스템이 있지만 검출 알고리즘에 대한 설명이 없으며, 팩트체크 후 증거를 제공하지 않음
  - Times, FakerFact 등도 구글 크롬 스토어에 등록되어 있지만 기술에 대한 주장을 찾을 수 없음

### BRENDA 설계

- 프론트엔드는 Javascript, 백엔드는 파이썬으로 구현된다.

#### 1) 프론트엔드

- 브라우저를 통해 사용 가능하며 Javascript로 구현됐다.

#### 2) 백엔드

- RESTful API를 통해 사용자가 선택한 URL 또는 텍스트를 서버로 전송함
- 서버는 구글 등 검색엔진을 통해 관련 기사를 검색하여 텍스트를 분석함
- 머신러닝 모델을 적용해 분석하며, 정확성을 예측함
- 발견된 증거에 근거한 정확성 점수가 브라우저로 전송됨

#### 1] Querying the Web

- ① 구글 API를 사용하여 관련된 상위 10개의 웹페이지를 검색
- ② Newspaper3k9 파이썬 라이브러리를 이용하여 가짜뉴스 판단을 할 텍스트의 날짜, 작성자 등을 추출
- ③ 유사성을 검사하여 유사점수 0.75를 넘는 관련 기사들을 추출

#### 2] SADHAN Mode

- ① deep neural network를 형성한 SADHAN 모델을 사용
- ② 계층적 신경망을 통해 텍스트와 그 외 기사들의 단어, 문장 수준을 학습
- ③ 판단할 텍스트와 증거 기사를 입력으로 포함
- ④ 결과 추출

ref.BRENDA: Browser Extension for Fake News Detection