

세종대 캡스톤 디자인 2조

언어 습관 추출

희소행렬 처리를 통한 연산속도 증가



01.

목차

1. 문제인식

- 개발 동기
- 개발 목적

3. 활용 및 시장성

- 시장 조사
- 활용성

2. 문제 해결

- 아이디어
- 기능 구성
- 시스템 구성도
- 개발 환경

4. 향후 일정

- 개발 일정
- 팀원 역할

문제 인식

개발 동기

유명인마다 언어습관,
자주 사용하는 단어들 존재

'great' - 'many' - 'good' - 'big' - ...

'kung flu' - 'nasty' - 'terrible' - 'unattractive' - 'loser' - 'fat'



문제 인식

문제 해결

활용 및 시장성

향후 일정

문제 인식

개발 목적

01.

기존 RNN,LSTM을
이용해 모델링시
연산과정이 오래 걸림



분석 모델의 연산량은 줄이고,
정확도는 높인다.

02.

습관적으로 사용하는 단어에서
평소 언어습관을 분석함



정치인, 셀럽들의 특정 게시글을
보고 파악하지 않음

01. 언어 습관 진단

언어 습관을 어떻게 판단할 수 있을까?

문장과 단어를 분석

트위터에서
인물이 작성한 내용을 수집

긍정적 단어, 부정적 단어, 기타로 분류



긍정비율과 부정비율 계산

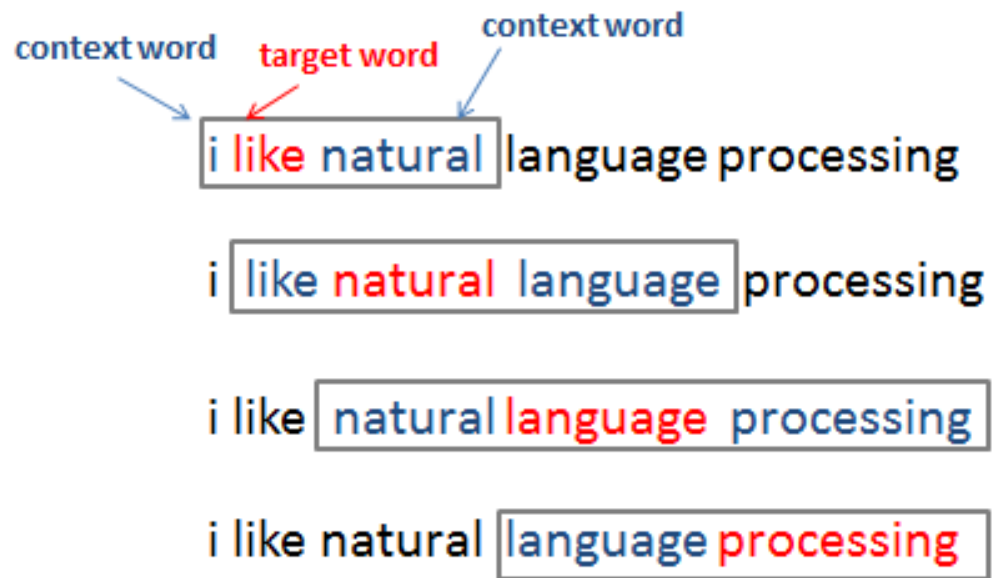
긍정적 단어 – 부정적 단어 비율 계산



5가지 언어 습관으로 분류

master of positive speech
sprout of positive speech
master of various speech
sprout of hate speech
master of hate speech

01. 단어 분석



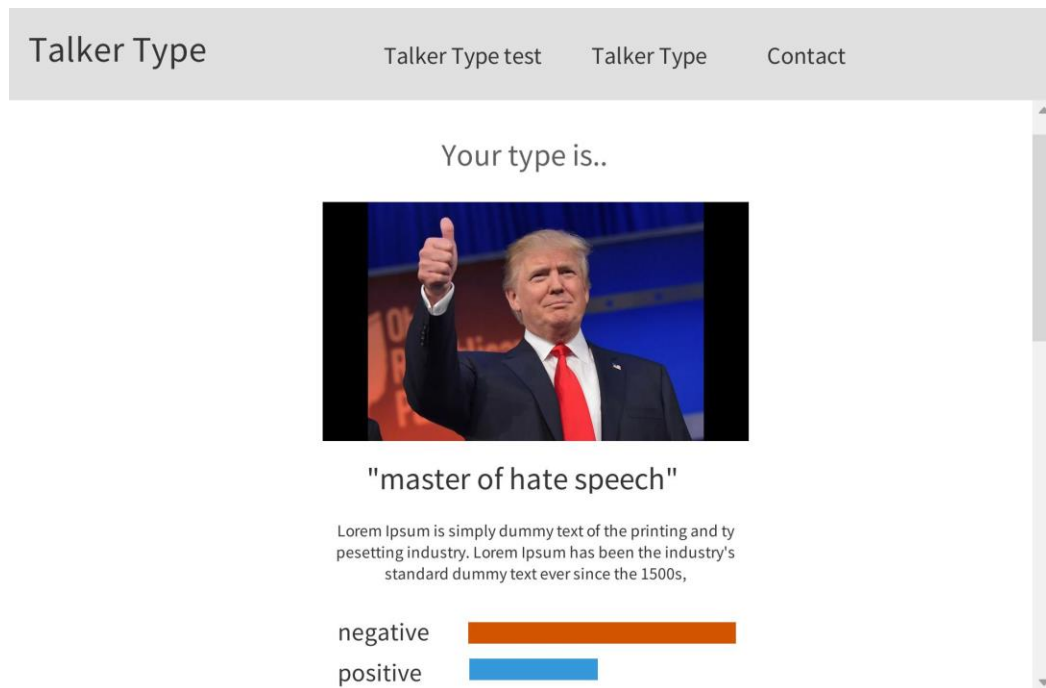
두 가지 모델 사용

- 기존 Word2vec모델만으로 분석
- Word2vec모델 연산 과정에 nzb 방식을 적용하여 분석

두 모델 성능 비교

- 기존 모델보다 메모리 접근 횟수가 줄어듦
- 인덱스 사이즈 감소

02. 웹 분석 결과

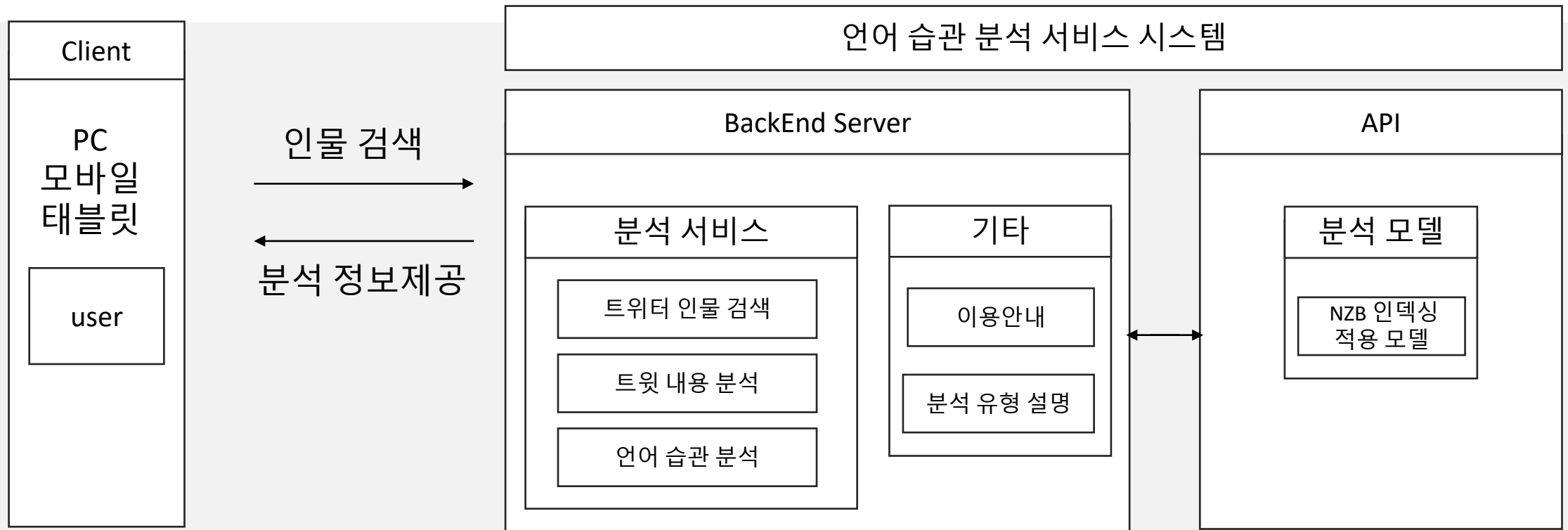


분석 결과

- 언어 습관 유형
- 유형 설명
- 긍정, 부정 단어 비율
- 자주 사용하는 단어 및 언어 습관
- 분석 데이터 출처 (twitter 계정)
- 공유하기

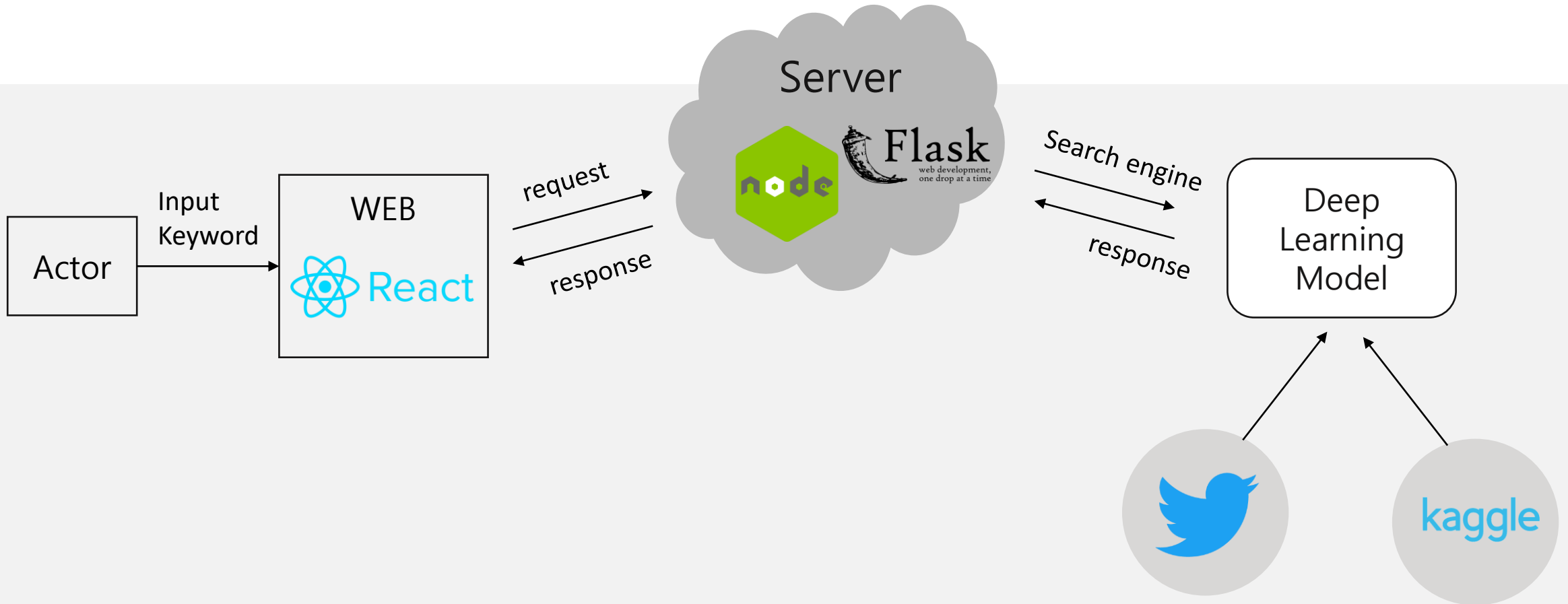
문제 해결

시스템 구성도



문제 해결

개발 환경



문제 해결

개발 환경

01. 언어

Python, C, JavaScript

02. 프레임워크

Flask

03. 협업툴

Glthub

04. 플랫폼

Anaconda, Kaggle, Node.js

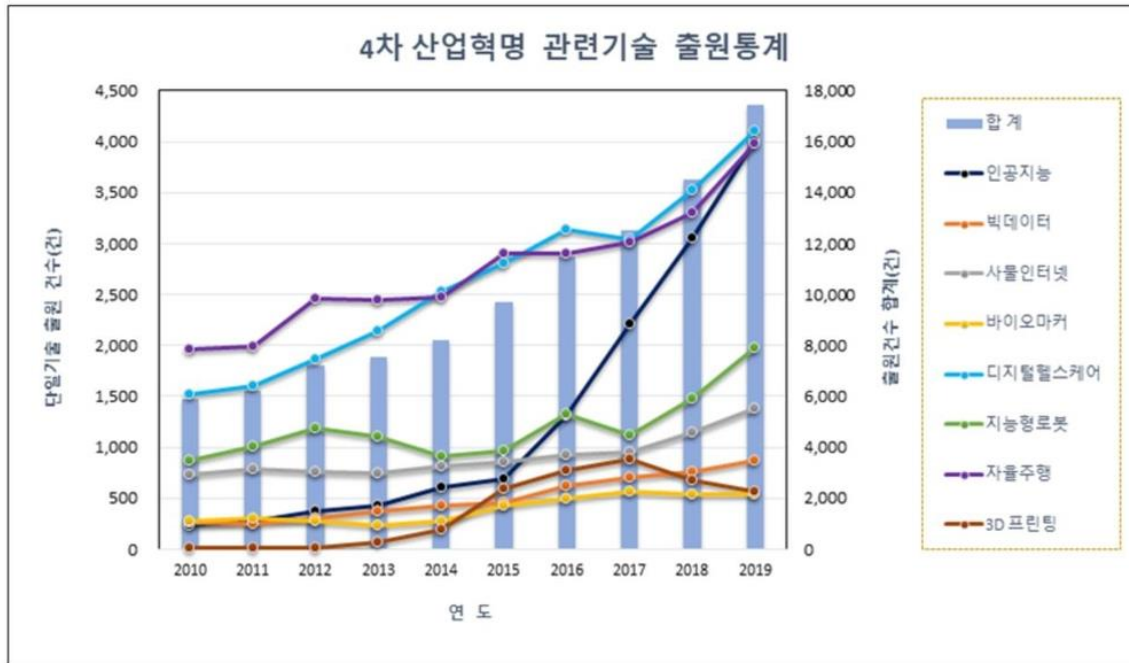
05. 라이브러리

React.js, Pytorch, Scikit learn,
TensorFlow

06. IDE

Pycharm, Jupyter Notebook,
Visual Studio Code

01. 인공 지능 분야 동향



최근 10년간 4차 산업혁명 관련 기술 출원통계, 특허청 제공, 2020-09-17

인공지능 분야 특허 꾸준히 증가 추세

인공 지능 소프트웨어 시장 확대 전망

데이터 처리비용(메모리 등)은 적게, 처리 속도는 빠르게

활용 및 시장성

시장 조사

02. 유사 사이트 비교

DeepMoji has learned to understand emotions and sarcasm based on millions of emojis. Here's a [video](#) explaining a bit more. Type a sentence to see what our AI algorithm thinks.

You love hurting me, huh?

Words are highlighted based on emotional impact. Click a word to turn it on/off.

you love hurting me , huh ?

💔 😭 😞 😠 😡

high confidence

DeepMoji 문서 감정 분석 결과

대부분 감정 분석에 초점이 맞춰져 있음

문장을 직접 복사, 붙여넣기 해야함

분석 가능한 텍스트 길이가 제한 됨

문제 인식

문제 해결

활용 및 시장성

향후 일정

활용 및 시장성

활용성

어떻게 활용할지에 대해

미국 선거철 다가옴

관심 있는 정치인들의 평소 단어 사용에 대해
알아보고 판단할 수 있음.

관심있는 아이돌의 언어 습관 알 수 있음.



문제 인식

문제 해결

활용 및 시장성

향후 일정

향후 일정

개발 일정

일정	9월			10월				11월				12월
	2	3	4	1	2	3	4	1	2	3	4	1
프로젝트 기획												
요구사항 분석 & 개발 환경 구축												
데이터 수집 및 전처리												
텍스트 분석												
기존 모델 검증												
모델 NZB 기술 적용												
웹 설계 및 구축												
모델 학습												
웹 UI 구현 및 데이터베이스 구축												
모델 검증 및 통계 제출												
시각화												
최종 문서 작성												
최종 발표												

문제 인식

문제 해결

활용 및 시장성

향후 일정

향후 일정

팀원 역할

01. 수민		02. 주희	03. 아현
PM 및 문서 작성		회의록 작성	텍스트 분석 모델
데이터 수집 및 전처리		데이터 수집 및 전처리	NZB 인덱싱 기법 적용
텍스트 분석 모델		텍스트 분석 모델	모델 학습
NZB 인덱싱 기법 적용		NZB 인덱싱 기법 적용	모델 검증 및 통계제출
모델 검증 및 통계제출		모델 학습	
			04. 승주
			PPT 제작 및 대본 작성
			웹서버 설계 및 구축
			웹 UI 설계 및 구현
			시각화
문제 인식		문제 해결	활용 및 시장성
			향후 일정