

세종대 캡스톤 디자인 2조

발화 패턴 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
- 시스템 구성도
- 기능 구성
- 개발 환경

4. 향후 일정

- 개발 일정
- 팀원 역할
- 팀 & 개인 목표

문제 인식

개발 목적

01.

습관적으로 내뱉는 말, 단어에서
평소 언어습관을 분석함



정치인, 셀럽들의 언어습관을 파악해
순간적인 모습으로 남을 판단 짓는
일을 줄인다

02.

기존 RNN,LSTM을 이용해 모델링시
연산과정이 오래 걸림



분석 모델의 연산량은 줄이고,
정확도는 높인다.

01. 발화 유형 진단

언어 습관을 어떻게 판단할 수 있을까?

문장과 단어를 분석

유튜브, 트위터 등에서
인물이 말하고 작성한 내용들 수집

긍정적 단어, 부정적 단어, 기타로 분류



긍정비율과 부정비율 계산

전체 발화 내용 중 긍정/부정적 단어, 문장
빈도를 계산

긍정적 비율 - 부정적 비율의 절대값 계산

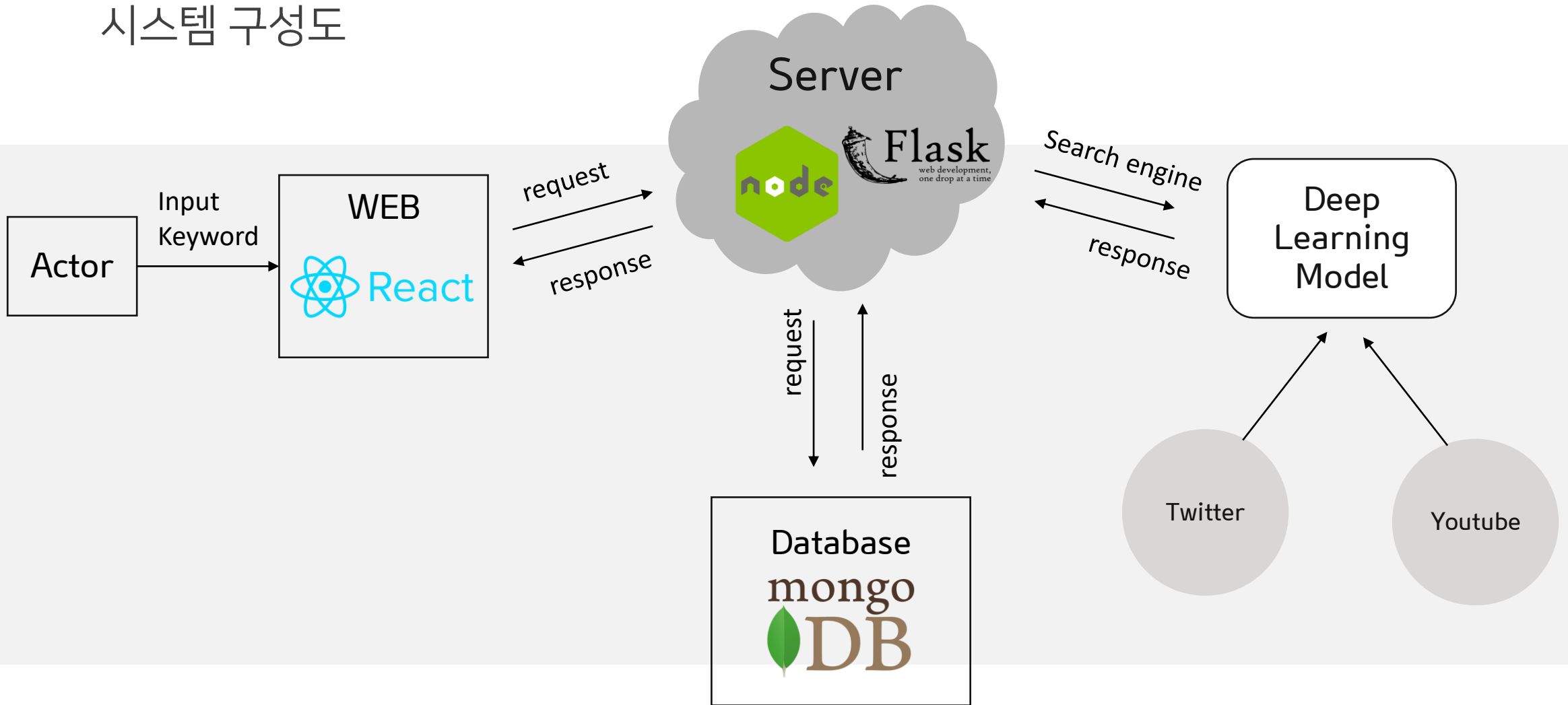


5가지 발화 유형으로 분류

초긍정/긍정/보통
/부정/매우 부정 으로 분류

문제 해결

시스템 구성도



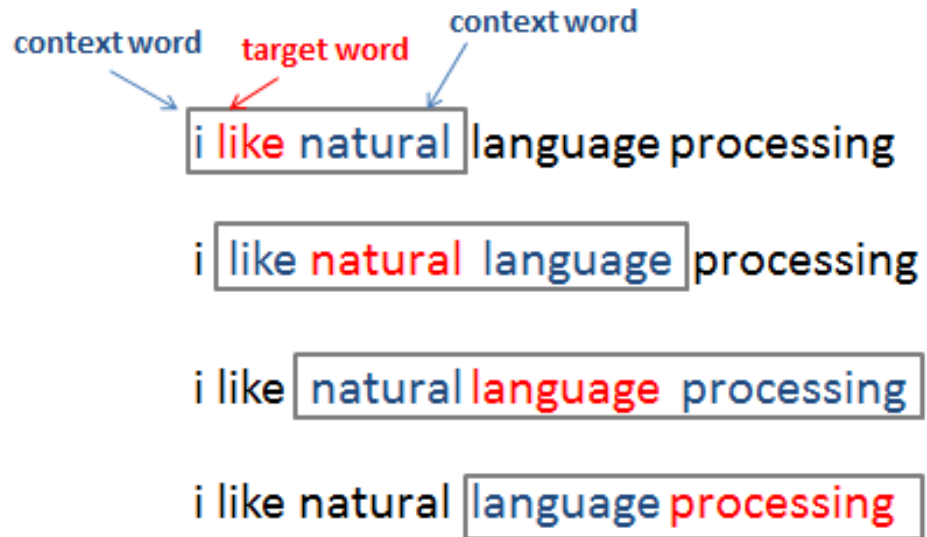
문제 인식

문제 해결

활용 및 시장성

향후 일정

01. 문장, 단어 분석



Word2vec모델 연산 과정에서 nzb 방식을 적용

기존 모델보다 메모리 접근 횟수가 줄어들고

인덱스의 사이즈를 감소

02. 웹 서비스

발화 유형		
인식 시작	분석 결과	이용 안내

'도널드 트럼프'의 발화 유형은?



헤이트 스피치의 달인



자세한 분석 결과



분석 결과

- 발화 유형
- 부정적 단어 비율
- 긍정적 단어 비율
- 자주 사용하는 단어
- 언어습관

문제 해결

개발 환경

01. 데이터 셋 구축

유튜브, 트위터 등 웹상에서
관련 자료 수집

02. 모델링

IDE | 파이토치(PyTorch)

03. 협업툴

Glthub

04. 웹 서비스

프론트엔드 | react.js

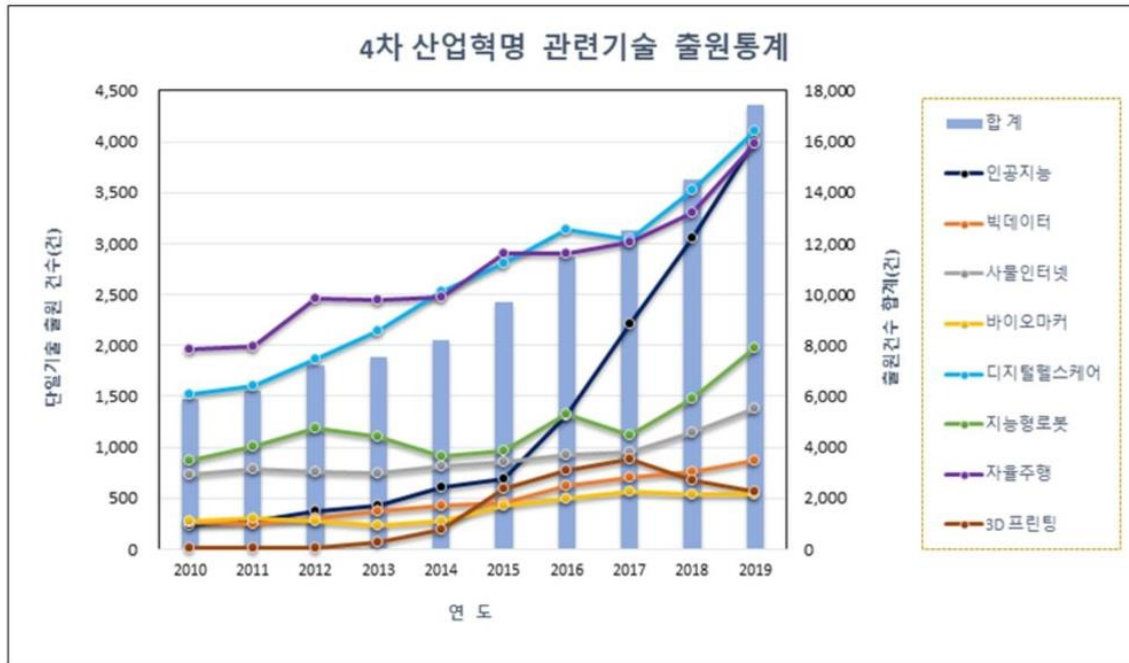
백엔드 서버 | node.js

모델 API 배포 | flask

프로젝트 배포 | AWS EC2

데이터베이스 | mongoDB

01. 인공 지능 분야 동향



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

인공지능 분야 특히 꾸준히 증가 추세

인공지능을 활용한 프로그램들 역시 증가

데이터처리비용(메모리 등)은 적게, 처리 속도는 빠르게

경쟁력을 키우는 기법을 적용한 사례 프로그램

활용 및 시장성

활용성

어떻게 활용할지에 대해

미국 선거철 다가옴

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

관심있는 아이돌의 언어습관과
화법에 대해 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												
웹 디자인 및 설계												
영상에서 텍스트 추출												
텍스트 분석												
데이터 전처리												
모델 NZB 기술 적용												
웹 UI 구현 및 구축												
검증 및 테스트												
단어 사용 통계												
시각화												
최종 문서 작성												
최종 발표												

문제 인식

문제 해결

활용 및 시장성

향후 일정

향후 일정

팀원 역할

01. 수민

PM 및 문서 작성
데이터 전처리 (SUB)
텍스트 분석 모델 설정
NZB 인덱싱 적용
통계 제출 및 시각화

02. 주희

회의록 작성
데이터 수집 및 전처리
영상 텍스트 추출(SUB)
NZB 인덱싱 적용

03. 아현

영상 텍스트 추출
NZB 인덱싱 적용
모델 테스트
통계 제출

04. 승주

PPT 제작 및 대본 구성
웹 설계 및 구축
데이터베이스 구축
UI 설계 및 구현
시각화

문제 인식

문제 해결

활용 및 시장성

향후 일정

향후 일정

팀 & 개인 목표

NZB 인덱싱 기술 적용해 성능 좋은 웹 프로그램 만들기

노수민

문제없이 프로젝트 완성
NZB 기법 이해 및 적용

홍주희

NZB기법을 적용한 딥러닝 프로젝트 완성도있게 만들기

이아현

NZB 인덱싱 기법 이해와 응용

양승주

문제없이 웹 서비스 구현
파이썬을 이용하여 데이터 시각화

문제 인식

문제 해결

활용 및 시장성

향후 일정