

길JOBUI

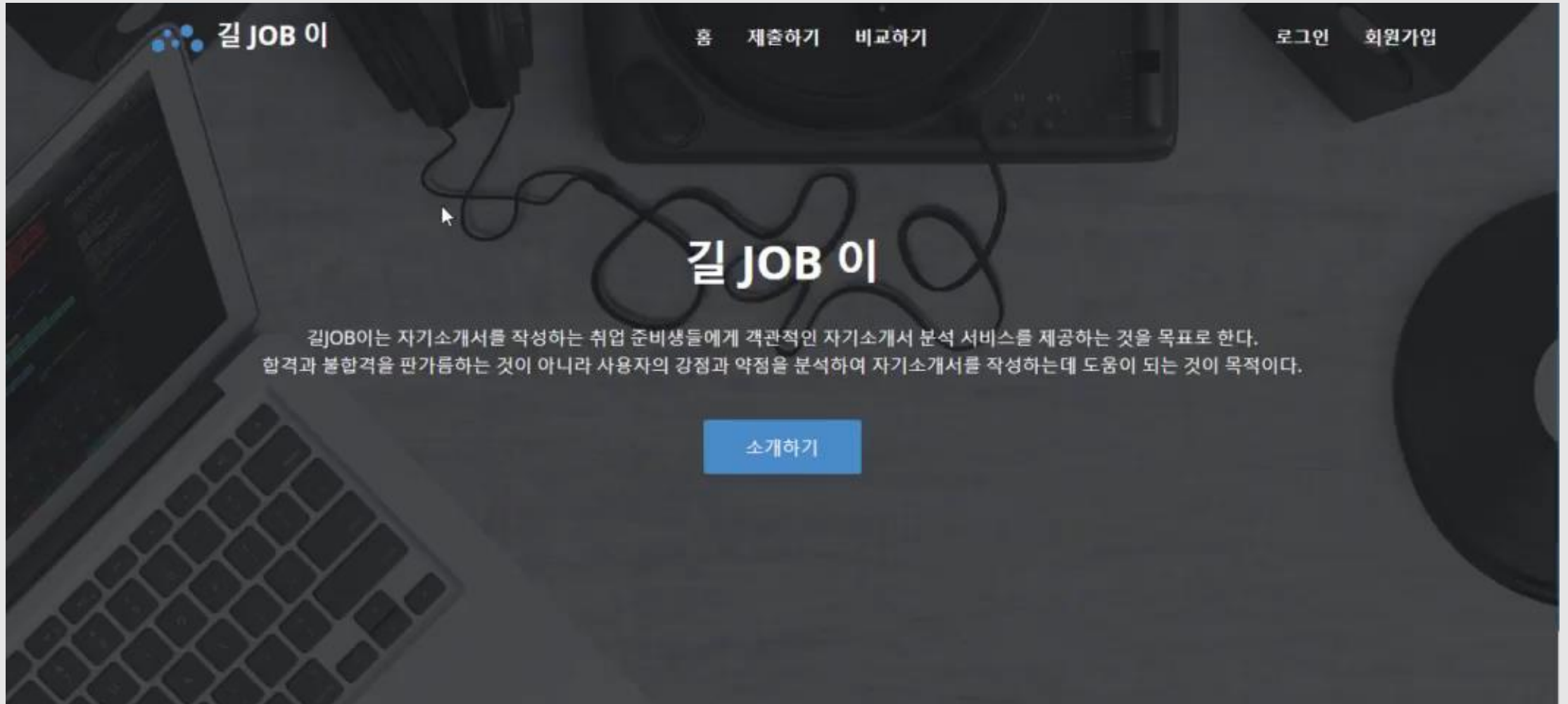
By Team 4726

소프트웨어학부 고현경
김혜인
김희주
이선흥
이수민

프로젝트 목표

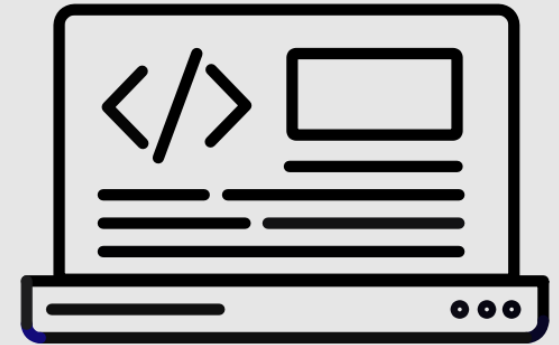
자연어 처리를 활용한 자기소개서 분석

시나리오 동영상



수행내용 서버

Bootstrap을 이용한 view 완성



B Bootstrap



AWS EC2를 활용한 웹서버 환경 구축

수행내용 서버



Spring security를 이용한 로그인기능 구현

spring boot JPA를 이용하여 DB 구축



수행내용 자연어처리 - 데이터수집



BeautifulSoup



기업별



직무별



NEWS



BLOG



핵심역량과 관련된 텍스트 집합(10개)



창의



도전

...

수행내용 자연어처리 - 전처리



토큰화(형태소 분석)



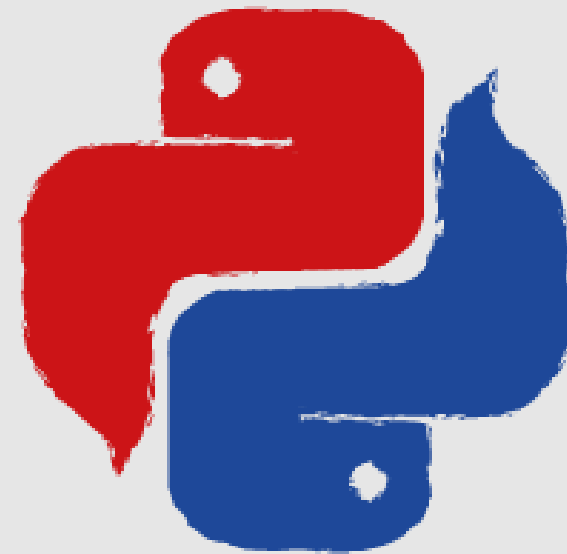
정규화



어근화

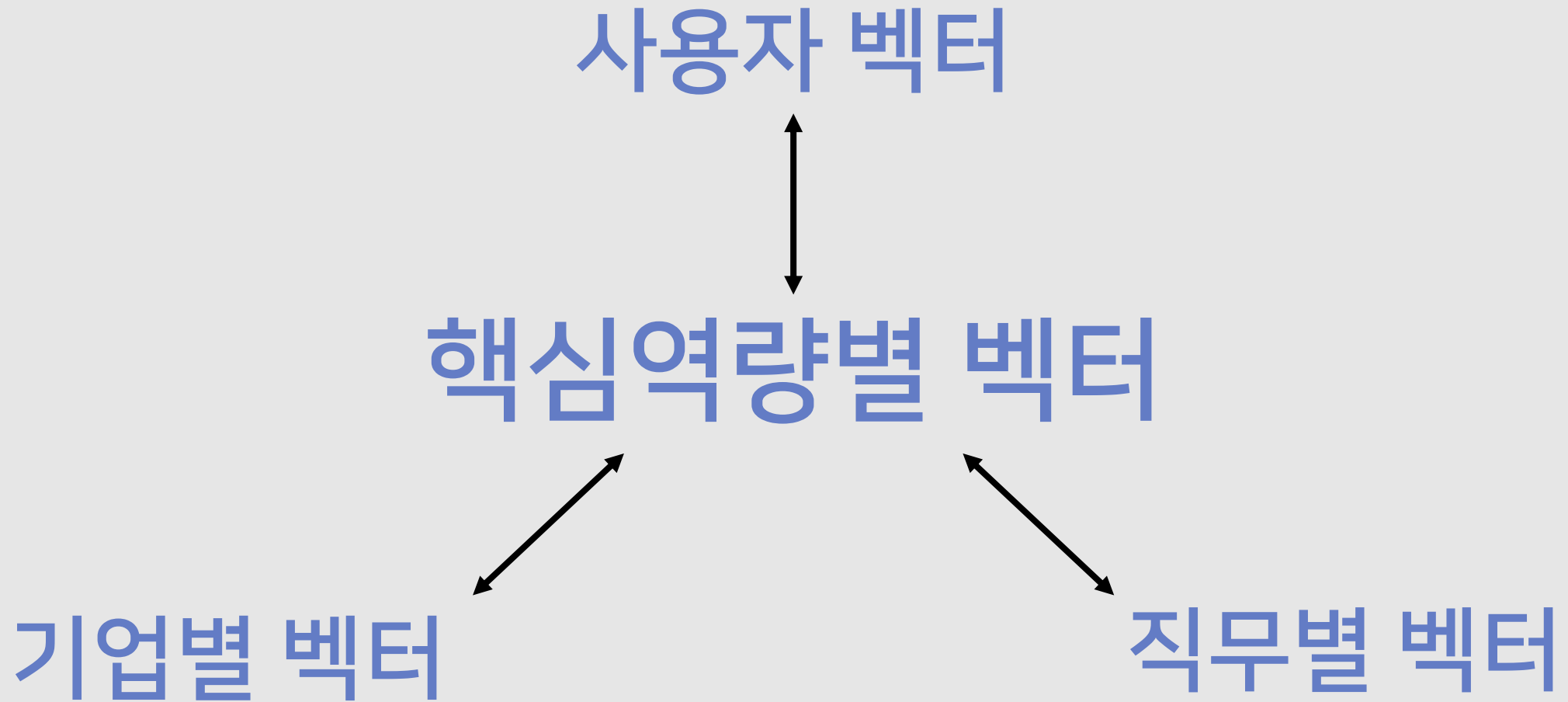


불용어 제거



KoNLPy

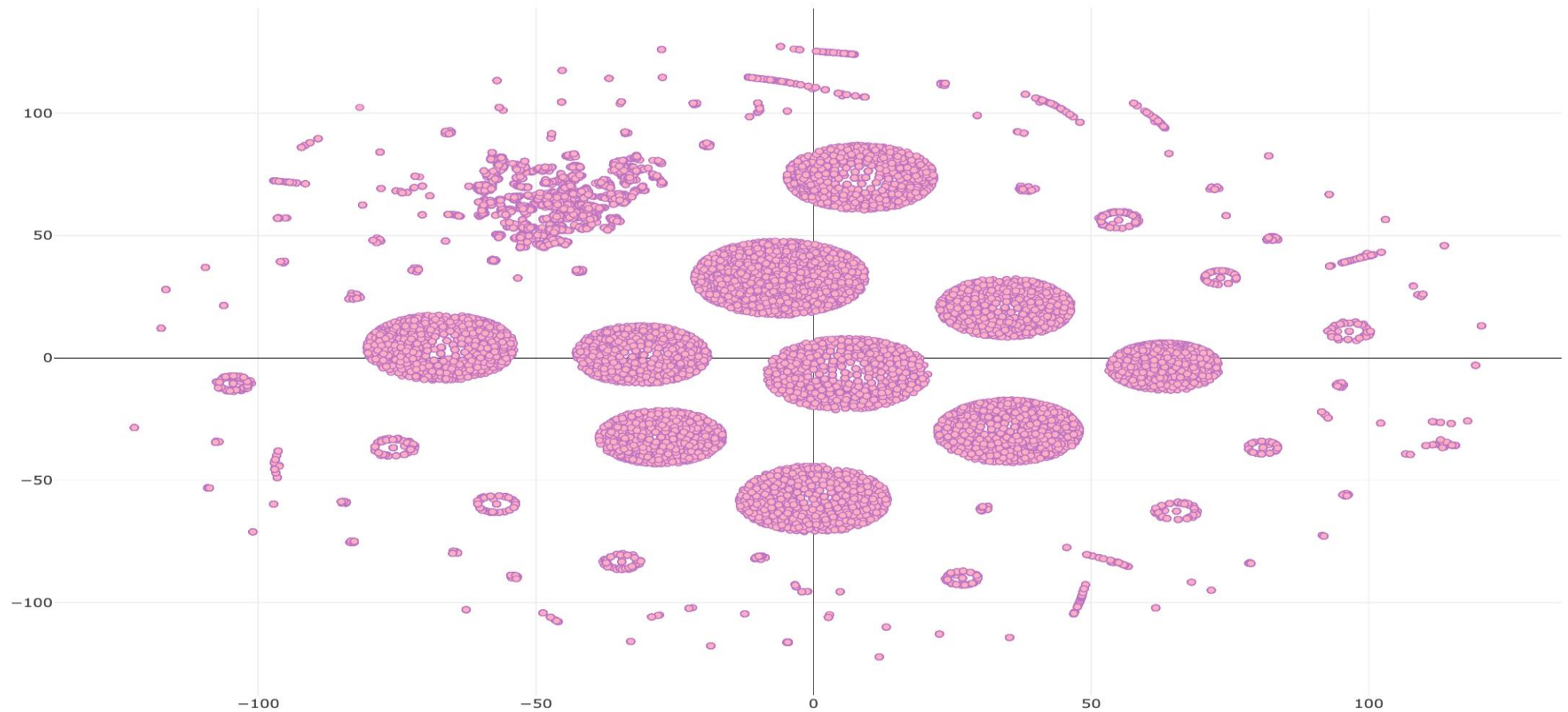
수행내용 자연어처리 - 벡터화(Scikit-learn 사용)



수행내용 자연어처리 - 벡터화

	aa	academy	acs	adas	afc	afp	ai	all	ambiente	amforht	...	희생	희생양	히든카드	히스패닉	히어로	히트	힌트	힘내다	힘쓰다	힘합
글로벌영량	0.000000	0.000000	0.014029	0.000000	0.000000	0.014029	0.000000	0.000000	0.000000	0.014029	...	0.019468	0.000000	0.000000	0.000000	0.000000	000000	0.000000	0.000000	0.000000	0.000000
능동	0.000000	0.000000	0.000000	0.015814	0.000000	0.000000	0.037532	0.000000	0.000000	0.000000	...	0.017705	0.000000	0.000000	0.000000	0.000000	000000	0.000000	0.000000	0.000000	0.000000
도전	0.014549	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.014549	0.000000	...	0.031771	0.000000	0.000000	0.000000	0.037963	012368	0.000000	0.030532	0.000000	0.000000
성실	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	000000	0.000000	0.000000	0.011023	0.000000
소통	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	013680	0.000000	0.000000	0.016180	0.000000
인내심	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.02942	0.000000	0.000000	000000	0.014771	0.000000	0.000000	0.000000
정직	0.000000	0.000000	0.000000	0.000000	0.015495	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.010246	0.000000	0.000000	0.000000	0.000000	000000	0.013172	0.000000	0.009201	0.000000
주인의식	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.015356	0.000000	0.015356	0.000000	000000	0.000000	0.000000	0.000000	0.000000
창의	0.000000	0.015422	0.000000	0.000000	0.000000	0.000000	0.022198	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	000000	0.000000	0.000000	0.009158	0.000000
팀워크	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.015639	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000	000000	0.000000	0.000000	0.015724	0.03732

수행내용 자연어처리 - 벡터화



수행내용 자연어처리 - 간단한 수행예시

예시 입력 문장 (자소서 일부)

'저의 이런 능동적인 자세가 조직의 목표달성을 달성하는데 기여할 수 있다고 생각합니다.
이러한 역량을 바탕으로 삼성인이 되어 전기설비 기술 업무에 빛을 발하는 신입사원이 되겠습니다.'



(1) 입력 문장을 전처리한 결과로 나온 전체 어휘 (16개)

['능동', '자세', '조직', '목표', '달성', '달성', '기여', '생각', '역량', '바탕', '성인', '전기', '설비', '기술', '업무', '신입사원']

(2) 입력 문장의 어휘중 미리 분석한 역량 데이터와 공통된 어휘 (14개)

['능동', '자세', '조직', '목표', '달성', '달성', '기여', '역량', '바탕', '전기', '설비', '기술', '업무', '신입사원']

수행내용 자연어처리 - 간단한 수행예시

(3) 핵심역량데이터와 공통된 14개의 어휘가 가지는 TF-IDF값 벡터
벡터의 사이즈는 (10 X 14)



	능동	자세	조직	목표	달성	달성	기여	역량	바탕	전기	설비	기술	업무	신입사원
글로벌역량	0.000000	0.000000	0.011177	0.003152	0.013616	0.013616	0.005147	0.009971	0.007075	0.013886	0.006617	0.016684	0.014447	0.000000
능동	0.054282	0.011380	0.011328	0.009411	0.018330	0.018330	0.019182	0.015185	0.009378	0.014851	0.000000	0.018886	0.015158	0.008324
도전	0.000000	0.003888	0.000000	0.006005	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.006583	0.000000	0.000000
성실	0.000000	0.004111	0.000000	0.003750	0.000000	0.000000	0.036293	0.000000	0.000000	0.000000	0.000000	0.000000	0.012385	0.000000
소통	0.000000	0.003773	0.011567	0.003442	0.000000	0.000000	0.000000	0.009576	0.000000	0.000000	0.000000	0.006388	0.011367	0.000000
인내심	0.000000	0.011222	0.000000	0.012543	0.000000	0.000000	0.000000	0.000000	0.008805	0.000000	0.000000	0.004300	0.008233	0.000000
정직	0.000000	0.014709	0.009141	0.006127	0.005306	0.005306	0.000000	0.000000	0.011447	0.000000	0.000000	0.000000	0.011143	0.000000
주인의식	0.000000	0.008108	0.014012	0.000000	0.000000	0.000000	0.000000	0.009805	0.000000	0.000000	0.000000	0.009220	0.014419	0.000000
창의	0.000000	0.000000	0.011404	0.008391	0.004716	0.004716	0.011020	0.017985	0.010173	0.000000	0.017615	0.016701	0.014153	0.000000
팀워크	0.007915	0.015593	0.020676	0.010527	0.005529	0.005529	0.000000	0.019973	0.011928	0.000000	0.000000	0.004133	0.003771	0.000000

수행내용 자연어처리 - 간단한 수행예시

(4) 핵심역량별 TF-IDF값의 합을 통해 자기소개서와 유사한정도 파악

'저의 이런 능동적인 자세가 조직의 목표달성을 달성하는데 기여할 수 있다고 생각합니다.
이러한 역량을 바탕으로 삼성인이 되어 전기설비 기술 업무에 빛을 발하는 신입사원이 되겠습니다.'

핵심역량별 유사도 (높은순)

1	능동	0.224
2	창의	0.116
3	글로벌역량	0.115
4	팀워크	0.105
5	정직	0.063
6	성실	0.056
7	주인의식	0.055
8	소통	0.046
9	인내심	0.045
10	도전	0.016



분석 결과 입력된 문장은
핵심역량 '능동'에 가장 가까운 것을 알 수
있다

변경사항 서버

계획



mongoDB



Amazon
EC2

정확한 기준이 없었음

DB변경

자연어 처리
서버

결과페이지
텍스트

수정

JPA
Java Persistence API



amazon
DynamoDB



Amazon
Lambda

인터뷰, 기업분석자료(NCS)
기반 텍스트 제공

변경사항 자연어처리 계획

각각의 벡터가
다른 기준을 가지고 있음

모든 비교에
코사인 유사도 사용

합격 자기소개서

벡터화
과정

벡터
비교방식

학습
데이터셋

수정

핵심역량 벡터를 기준으로
나머지 벡터를 만듦

사용자-기업 비교에만
코사인 유사도 사용

합격 자기소개서
+
인재상

향후 추진계획 서버

1. 사용자의 분석 기록을 비교해주는 API 개발
2. Web server와 자연어 처리 AWS lambda 간 통신 구현
3. AWS dynamo DB , S3 연동
4. View와 Controller 연동을 위한 Ajax 통신

향후 추진계획 자연어처리

1. 합격 자기소개서 추가 수집
2. 사용자 자기소개서와 비교할 수치 보정
3. 텍스트 분류 단계의 다양한 알고리즘 효율성 비교

THANK YOU !

By Team 4726