

텍스트 마이닝,
Word 2 Vec,
K-means 클러스터링을 활용한
공단 내 문화생활 활성화 방안과
정보 제공 시스템 구축 방안

국민대학교 경영학부
빅데이터경영통계전공
경영분석학회 D&A

이세환
이유찬
신민용
경재준

공단?

工團?

공단!

共團!

개요

일과 삶의 균형을 이루는 것은
개인의 삶에서도, 기업의 입장에서든 필수적인 요소가 되었다.

또한, 4차산업혁명 이후에 다양한 이론과 기술이 생겨나고 있고
산업의 변화 속도는 그 어느 때보다 빠르다,

기업들은 이러한 시대의 흐름에 발맞추어 가기 위해
다양한 서비스를 제공해야 한다.

1. 공단 내 문화 활동 활성화

1. 배경설명

-산업단지 내 문화시설 부족

워라밸? → 일과 삶의 균형!
(Work and Life Balance)

(1) 배경 설명 : 워라밸

52시간 근무제



소확행

(소소하지만 확실한 행복)



Work and Life Balance

의식과 정책적인 면 모두에서 사회 흐름의 변화가 일어나고 있다.

(2) 배경 설명 : 워라벨과 현실

직장인 85.7%가 한국사회 직장인의 근무시간은 과한 편이라고 느꼈다.

꿈 같은(37.8%) 일이라는 생각도 많이 했으며, 가치 있고(34.9%), 건강한(30.6%) 삶의 방식으로 바라보는 한편으로 부럽고(28.2%), 이뤄지기 힘들며(23.9%), 비현실적이라는(20.3%) 생각도 많이 하는 모습이었다.

전체 79.6%는 삶의 질 개선을 위해서는 노동시간 단축이 반드시 필요하다고 답했다.

하지만 직장인의 76.3%는 한국은 일과 삶의 균형을 기대하기 어렵다고 응답했다.

워라벨이 어려운 이유로 일을 중시하는 문화와 경제적 여유 부족, '과도한 노동시간을 주로 꼽았다.

시장조사전문기업 엠브레인 트렌드모니터가 전국 만 19~59세 직장인 남녀 1000명을 대상으로 '워라벨' 관련 인식조사를 실시한 결과, 대부분의 직장인들은 한국사회의 노동시간이 너무 많고, 사람들이 일에만 매몰되어 사는 것 같다고 느끼고 있었으며, 이로 인해 일과 삶의 균형을 뜻하는 워라벨이 잘 실현되지 못한다고 생각하는 것으로 조사되었다.

꿈 같은(37.8%) 일이라는 생각도 많이 했으며, 가치 있고(34.9%), 건강한(30.6%) 삶의 방식으로 바라보는 한편으로 부럽고(28.2%), 이뤄지기 힘들며(23.9%), 비현실적이라는(20.3%) 생각도 많이 하는 모습이었다.

20대 = 워라벨 세대

문화지원시설에 대한 선호도 및 중요도가 높은 것으로 제시되어,
근로자의 근본적 욕구 및 기대에 부응할 수 있는
이용자 특성을 고려한 지원시설 구성방안이 요구되어진다.

도시내 첨단 산업단지 이용자 특성을 고려한 지원시설 분석 연구. 최형구, 김원필, 서울과학기술대학교 2016

(3) 산업단지 내 문화시설 구축 노력

산업단지에 첨단 편의, 문화시설 유치

산단공, 민간사업자 공모

한국산업단지공단(이사장 황규연)은 낡은 산업단지에 첨단 업무시설 및 주거·편의·문화·복지시설 등을 유치하는 ‘산업단지 구조고도화사업’에 참여할 민간대행사업자를 24일부터 공모한다.

산단공은 낡은 공장으로는 4차 산업혁명 시대에 대응하는 데 한계가 있다고 보고 부지용도 변경 등 파격적인 인센티브를 제공해 단지 내 부족한 편의시설 및 첨단 복합시설을 유치·조성할 계획이다.

... 한국경제, 17.04.23

근로자들의 문화 생활
요구 파악 및 커뮤니티 형성!

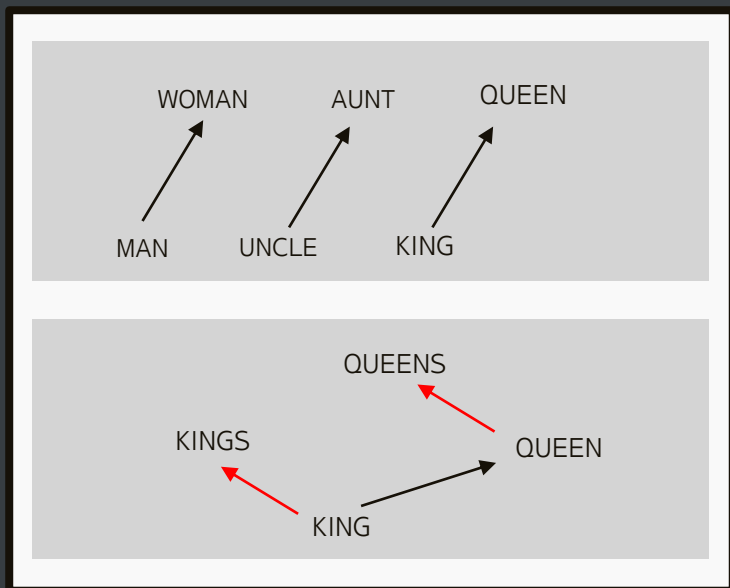
2. 아이디어 소개

-취미활동 공유 플랫폼: 원리 설명

1. Word2vec

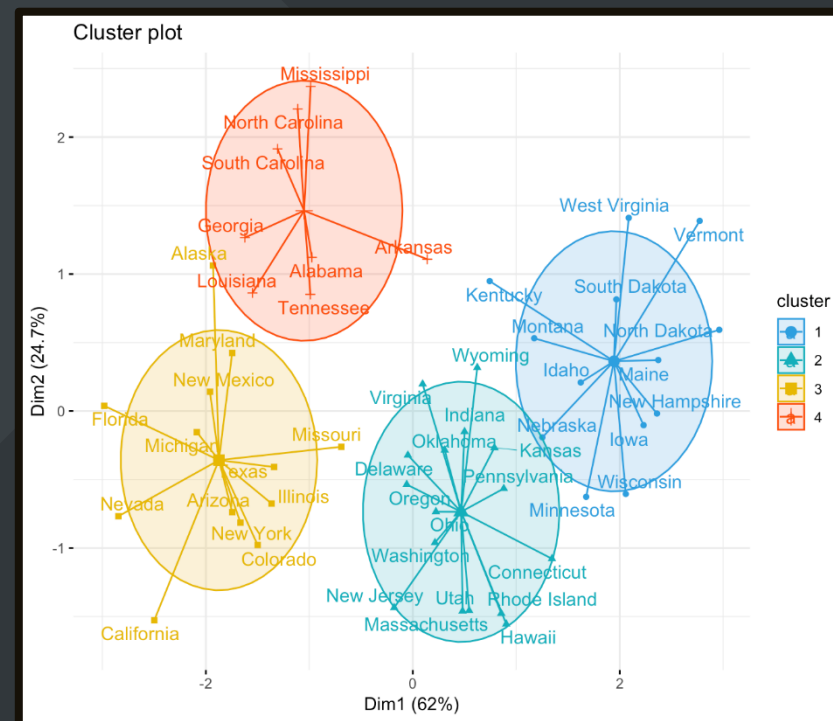
단어를 벡터공간에 할당하는 개념이다.

가까운 거리에 있는 단어일수록 의미상 유사도가 높다.
(skip-gram)

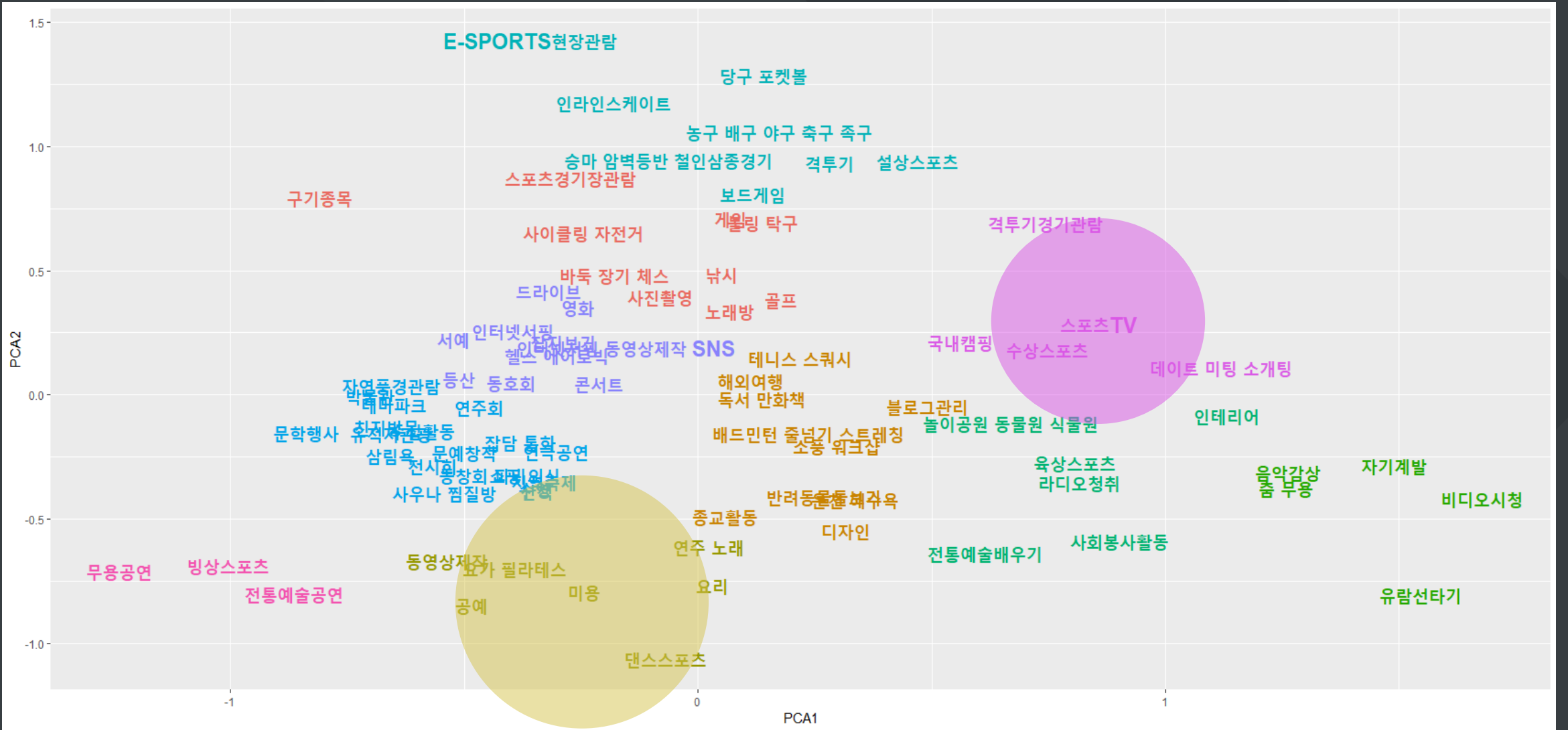


2. K-means Clustering

1. k 개의 중심점을 사전에 정하고
2. 거리 기준으로 가까운 중심에 각 요소를 할당 시킨 뒤
3. 요소들의 평균으로 중심점을 변경하고
4. 중심점이 바뀌지 않을 때까지 이를 반복한다.



(4) 유사도 기반의 군집 시각화



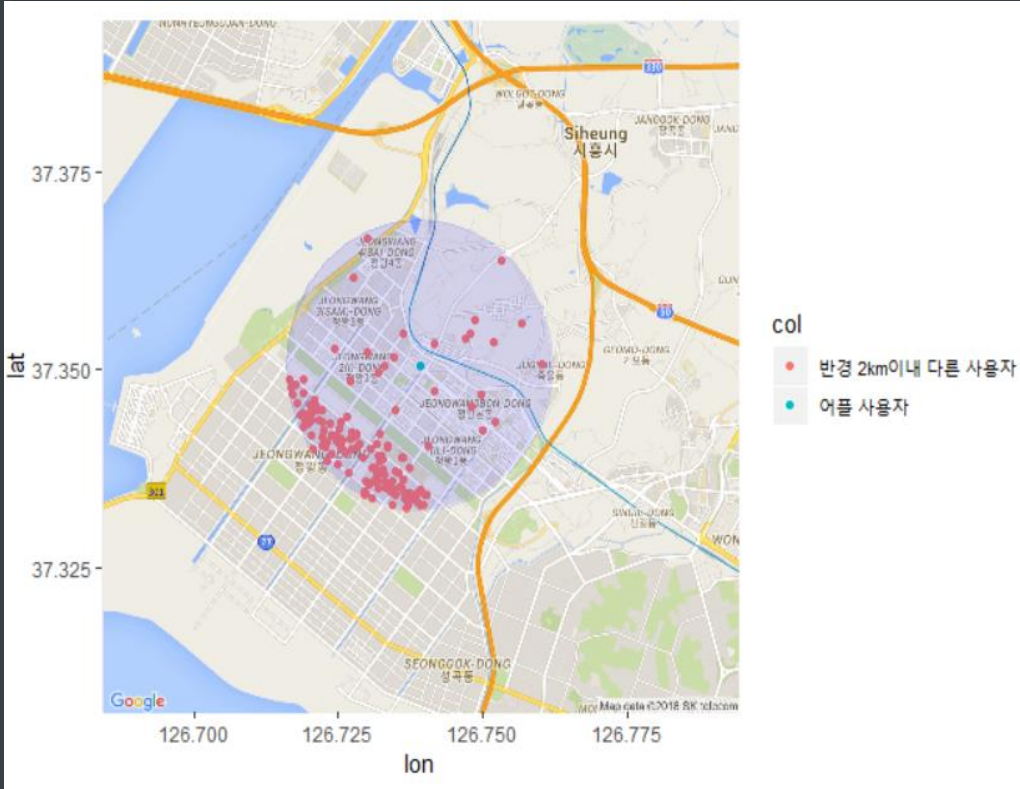
<10개 군집 시각화>

(5)k-means 기반 대표 여가 추출

군집	대표 여가	군집 된 여가
군집 1	바둑 장기 체스	게임, 낚시, 노래방, 스포츠경기장관람, 바둑 장기 체스 ...
군집 2	소풍 워크샵	종교활동, 독서 만화책, 소풍 워크샵 , 온천 해수욕, 블로그관리...
군집 3	미용	미용 , 요리, 요가 필라테스, 공예, 댄스스포츠...
군집 4	자기계발	음악감상, 자기계발 , 비디오시청, 유람선타기, 춤 무용...
군집 5	라디오청취	라디오청취 , 육상스포츠, 사회봉사활동, 인테리어, 전통예술배우기...
군집 6	농구 배구 야구 축구 족구	농구 배구 야구 축구 족구 , 당구 포켓볼, 격투기, 보드게임, 인라인스케이트...
군집 7	문예창작	테마파크, 수집활동, 문예창작 , 문학행사, 연주회, 박물관...
군집 8	헬스 에어로빅	동호회, 영화, 등산, 헬스 에어로빅 , 드라이브...
군집 9	스포츠TV	스포츠TV ,데이트 미팅 소개팅, 국내캠핑, 수상스포츠, 격투기경기관람...
군집 10	빙상스포츠	전통예술공연, 무용공연, 빙상스포츠 ...

(6)GeoCode를 이용한 위, 경도 변환

이름	주소	Latitude	Longitude
주)더원	경기도 시흥시 서해안로 242, 510호,근생2호 (정왕동, 시화하이테크 아파트형공장)	37.4405553	126.7839189
BMT KOREA	경기도 시흥시 마유로10번길 17 (정왕동, 중앙정밀)	37.3219855	126.7146126
주)샤인테크놀	경기도 안산시 단원구 정왕천동로94번길 10, 4마 103호 (성곡동)	37.320475	126.7315335
태현시스템	경기도 시흥시 마유로32번길 88, 3바 606 (정왕동)	37.3191214	126.7222147
주)영광정밀	경기도 시흥시 마유로132번길 52, 3라 707 (정왕동)	37.3282674	126.7257574
주)레이저센타	경기도 시흥시 공단1대로 284, 정왕동 1276-9 시화공단 3다 201-1호 (정왕동)	37.3473226	126.7416459
주)대일일렉트	경기도 시흥시 공단1대로379번안길 59, 4나 302호 (정왕동)	37.3280543	126.7471408
부영테크	경기도 시흥시 옥구천서로237번길 78, (정왕동 1242-7, 시화단지 1나308) (정왕동)	37.3499798	126.7113511
대성테크	경기도 안산시 단원구 진흥로102번길 19, 5라 605 (성곡동)	37.315919	126.7440074
에스엔이엔지	경기도 시흥시 공단1대로 13, 4층 418,419호 삼양스마트테크노(지) (정왕동)	37.3381123	126.7248191
서울산업기계	경기도 시흥시 마유로32번길 79, 3바 315호 (정왕동)	37.3201309	126.7217486
신영특장엔씨	경기도 시흥시 마유로10번길 76, 시화단지 3바 1002호 (정왕동)	37.3183537	126.7194603
월드와이어컷팅	경기도 시흥시 군자천로185번길 25, (정왕동 1264-2 2다 603) (정왕동)	37.3370283	126.7217433
삼성E&C	경기도 시흥시 경기과기대로 219, 4층 419 (정왕동)	37.3269131	126.7259926
동서엔지니어링	경기도 안산시 단원구 변영로15번길 43, 4바 862호 (성곡동)	37.3154496	126.7260596



사용자의 공단간 거리를 기반으로 취미활동 그룹을 할당함.
거리 계산을 위해 공단의 주소들을 위, 경도로 변환.

어플 사용자를 중심으로 2KM 반경에 있는 사용자를 시각화 한 것.
이처럼 사용자의 공단을 중심으로 가까운 사람을 우선적으로 매칭

구동화면 예시



회원가입

KMU_DNA
아이디를 입력해주세요

이름, 한글 또는 영문만 입력 가능

생년월일
1995 년 11 월 23 일

직장 :

직종 : 생산직(통신)

관심 분야 (선택 사항)

통신 ▼ 케이בל ▼ 3분야 ▼

입력받는 정보 1 : 구직 여부, 직종
구직중인 경우 구인,구직 관련 정보 제공.

입력받는 정보 2 : 관심 분야
관심분야를 입력하는 경우
관련 이슈 알림 서비스 제공.

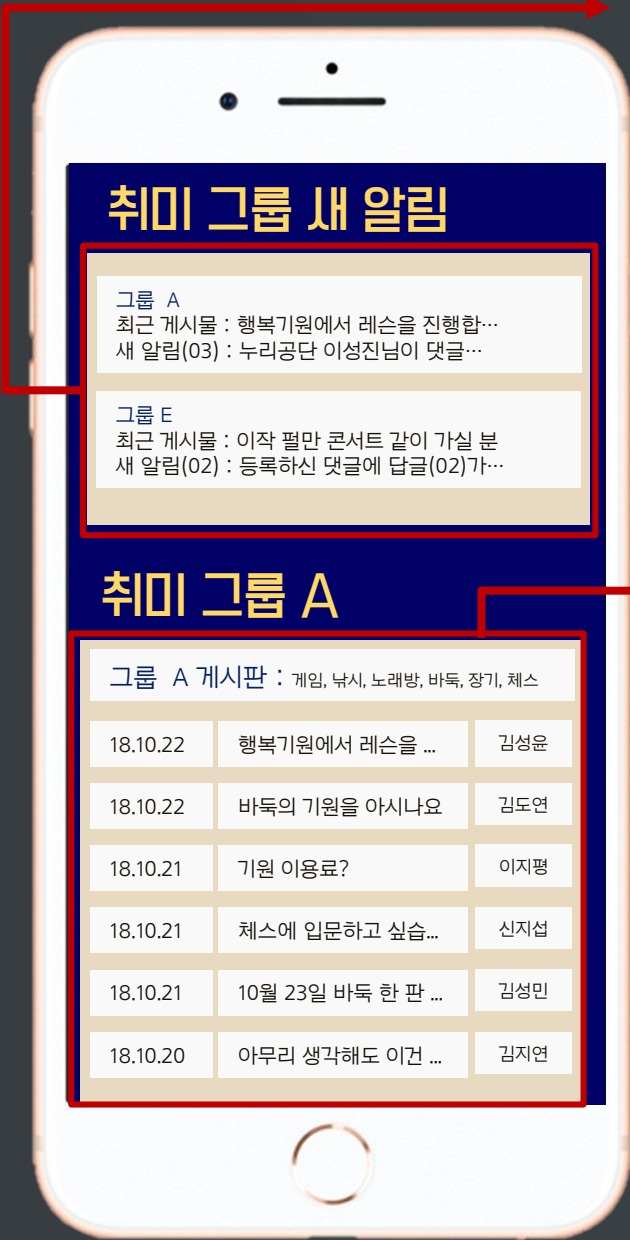
취미 그룹을 대표할 수 있는 활동과
소속된 취미들이 함께 표시된다.

사용자가 취미 그룹을 최대 3개까지 설정할 수 있다.

취미 설정

관심있는 취미 그룹을 선택해 주세요(최대 3개)

A	여가활동 - 1 대표 여가활동 : 바둑, 장기, 체스 게임, 낚시, 노래방, 바둑, 장기, 체스 등
B	여가활동 - 2 대표 여가활동 : 소풍, 워크샵 소풍, 워크샵, 온천, 해수욕, 종교활동 등
C	여가활동 - 3 대표 여가활동 : 축구, 농구, 족구, 야구 농구, 배구, 야구, 축구, 족구, 당구, 격투기 등
D	여가활동 - 4 대표 여가활동 : 미용, 요리, 요가 미용, 요리, 요가, 필라테스, 댄스스포츠 등
E	여가활동 - 6 대표 여가활동 : 음악감상, 비디오시청 음악감상, 자기계발, 비디오시청, 무용 등



소속된 취미 그룹 내 새로운 알림이 있으면 알려준다.

소속된 취미 그룹 내 게시물이 게시판 형태로 보여진다.

→ 여가 활동 및 취미가 맞는 사람들끼리 매칭을 시키는 시스템 제작.

→ 여러 사람들과 취미활동을 같이하며 삶의 질을 높이고 소속감을 고취시켜 일의 능률을 높일 수 있도록 함

→ 국민 여가 실태 조사에서 선호도 상위 5개항목을 뽑은 뒤 W2V와 K-MEANS 클러스터를 통해 클러스터를 나누어 회원가입시 자신이 관심 있는 취미를 선택하게 함

→ 해당 군집을 선택한 회원끼리 매칭하여

서로의 취미를 공유하고 같이 활동할 수 있게 하는 서비스를 제공

→ 청년들에게 공단의 긍정적 이미지를 심고

친화적 공단 구성에 기여할 것으로 기대

➡ 이용자들의 여가와 문화에 대한 니즈를 파악 可!

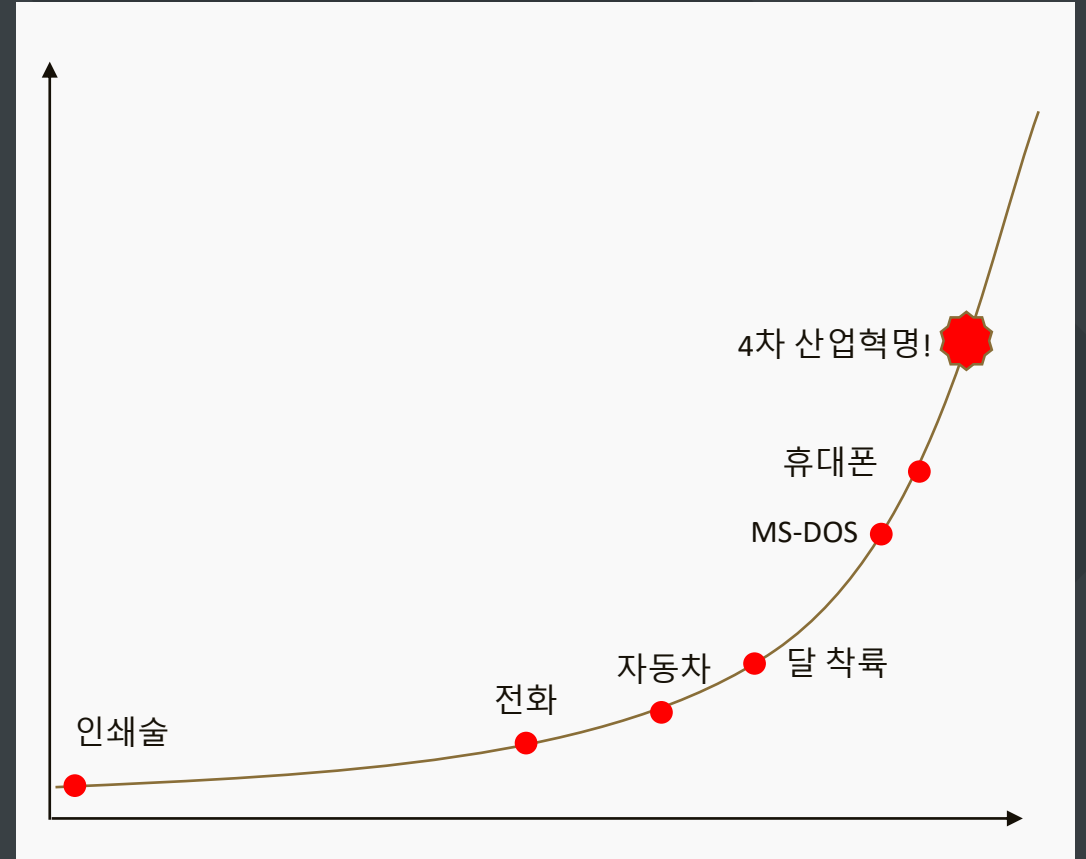
2. 4차 산업혁명에 맞는 효과적인 정보 제공 시스템

1. 배경설명

-4차 산업혁명과 그에 따른 산업환경의 변화

4차 산업혁명?

→ 인공지능, 사물 인터넷, 빅데이터. 등
첨단 정보통신기술이 경제·사회 전반에 융합되어
혁신적인 변화가 나타나는 차세대 산업혁명.
→ 기술의 발전 속도는 점점 빨라진다.



(1) 배경설명 : 4차 산업혁명

"4차 산업혁명 대비하자"...산업단지에도 스마트 바람

산단공, 5대 과제 선정

"4차 산업혁명 잘 몰라"...입주기업 68%가 '무대응'

1 미국, '이란 원

2 삼성, 3분기

3 제너럴 모터

산업단지 입주 기업의
4차 산업혁명 준비 정도 (단위:%)



※조사:한국산업단지공단 입주업체 244개 사대상
조사 기간:3월23일~4월21일

4차 산업혁명 관련 기술 도입 시
애로사항 (단위:%)



※복수응답

(2) 산업단지 내 교육과정 및 내부 노력



지역 특화산업 육성과 경제 활성화를 위한 국가산업단지 후보지 선정 및 발표 - 미래 지향적 스마트 산업단지로 조성 추진 -

- 국토교통부(장관 : 김현미)는 8월 31일(금) 산업입지정책심의회(위원장 : 국토교통부 1차관)를 개최하여 국가산업단지* 후보지 7개소**를 선정하였다고 밝혔다.

* 과학기술산업 육성, 낙후지역 개발 등을 위해 국토부장관이 지정하여 공장 등 산업시설용지, 지원·주거·상업시설용지 등을 조성(전국 44개, 787km², 18.6월말 기준)

** 강원 원주, 충북 충주, 경북 영주, 충북 청주, 세종, 충남 논산, 전남 나주

수송혁명, 스마트 이동체
#온라인신청 #신청순마감 #스마트이동체 #드론 #자율주행차 #현대모비스 #엠피넥스 #영우원

강사진 소개

시간	강사명	프로필
13:00 ~ 14:00	"자율주행자동차 단계별 개발동향"	<ul style="list-style-type: none"> 현) 현대모비스 ICT연구소 책임연구원 전) 삼육대학교 카메카로닉스학과 강사 전) 삼성동력(미래형자동차) 윤곽프로그램 실습강사 전) 현대기아자동차 연구원 NGV교육 실습강사
14:05 ~ 15:35	"스마트이동체 핵심부품기업 초청특강"	<ul style="list-style-type: none"> 현) (주)엠피넥스 대표이사 현) (사)글로벌 선도기업 협회 회장 현) 한국산업단지공단 글로벌 선도기업 리더스클럽 회장 현) 동국대학교 산업기술연구원 겸임교수 전) 팬택엔지니어링 선임연구원 전) 현대전자 휴대폰 개발 연구원
15:40 ~ 17:10	"4차 산업혁명시대에서 국내외 드론의 산업적 활용"	<ul style="list-style-type: none"> 현) 한국항공우주연구원 선임연구원

교육안내

- 교육일시 : 2018. 5. 23(수) 13:00 ~
- 교육비 : 무료
- 교육인원 : 70명
- 강의장소 : 한국산업단지공단 서울자여본부 키콕스벤처센터 3층 대회의실
- 신청방법 : www.kicox.or.kr → 교육신청 → 일정 및 신청
- 문의전화 : 070-8895-7372, 7374

2018년도 강의 일정

2/21(수)	3/28(수)	4/25(수)	5/23(수)	6/27(수)	9/12(수)	10/24(수)
지능화 인프라 : DNA (Data, Network, AI)	스마트이동체	융합서비스	산업기반			
Big Data 활용 (개강특강)	완료	완료	자율주행차 드론(무인기)	스마트시티 맞춤형 헬스케어	AR/VR 지능형로봇	지능형반도체 첨단소재 혁신산업 신재생에너지

한국산업단지공단

어떠한 서비스로
이를 보조할 수 있을까?

이용자들이 자신에게 맞는 최신의
정보에 빠르게 접근할 수 있게 만들자!

2. 아이디어 소개

-정보 공유 플랫폼: 원리 설명

1. konlpy

- 한국어 형태소 분석기인 konlpy를 활용하여 업종 중 명사만 파싱(parsing)

2. Word 2 VEC

- 파싱된 명사를 벡터공간에 할당

3. K-means Clustering

- K개의 군집으로 군집화

	x	y
발효	-3.742416	1.283807
디지털	-3.091872	-0.879565
램프	-5.753968	-1.628704
구조	-7.229112	1.836086
출력	-4.613565	-0.641172
실장	-6.231019	-0.865565
담배	-1.967661	-0.381467
진단	-4.619846	-1.785745
육상	-7.238537	1.857673
분말	-4.006332	2.384303
구류	-5.654897	2.260339

['배전반',
'전기',
'자동',
'제어',
'반',
'제조업',
'기기',
'자동',
'측정',
'제어',
'장치',
'제조업',
'에너지 저장',
'장치',
'제조업',
'기타',
'전기',
'변환',
'장치',
'제조업',

업종 클러스터링 결과 예시

```
Cluster 4
['저장', '포', '수산', '동물', '장육', '제외', '육류', '육', '가금']

Cluster 5
['송풍기', '냉장', '냉동', '조화', '공기', '배기']

Cluster 6
['신품', '부품', '제조', '조', '향', '자동차']

Cluster 7
['취급', '컨베이어', '산업', '물품', '밸브']

Cluster 8
['제조업', '장비', '외', '장치']

Cluster 9
['보조', '무인 항공기', '전투', '무인', '우주선', '항공', '유인', '비행장', '항공기']
```

생산품 클러스터링 예시

```
Cluster 0
['전차', '기술', '외관', '우주선', '원자력', '란스', '로터', '성능', '오락', '반도체소자', '다기', '포탄', '베드', '케미칼', '세라믹스', '라디오', '전자통신', '실습', '프로펠러', '레스', '직교좌표', '피니언', '냉향', '역', '토목공사']

Cluster 1
['제제', '서적', '교재', '카트리지', '토너', '고분자', '농약', '합기', '크림', '캡슐']

Cluster 2
['식생', '매트']

Cluster 3
['변압기']

Cluster 4
['비용']

Cluster 5
['금']

Cluster 6
['제책', '의약', '지관', '양면', '두발', '종이컵', '편광', '완구']
```

→ 단어 수가 적어 군집화가 성공적으로 이루어지지 않았다.

이후 각 공장과 클러스터의 유사도를 구하여
각 공장을 가장 유사도가 높은 클러스터로 분류한다.

각 공장의 업종코드를
세세분류(5) 세분류(4) 소분류(3) 중분류(2)로 구분한다.

	name	five	four	three	two
1	은성산업(주)	25112	2511	251	25
2	은성산업(주)	35114	3511	351	35
3	은성산업(주)	35119	3511	351	35
4	(주)신한정기	25912	2591	259	25
5	삼지금속공업(주)	24213	2421	242	24
6	삼지금속공업(주)	24229	2422	242	24
7	양덕ENG	29280	2928	292	29
8	(주)이노화인	25923	2592	259	25
9	썬스프링	25943	2594	259	25
10	에스엘테크	24191	2419	241	24
11	신영기계	25924	2592	259	25
12	(주)형신기계제작소	29199	2919	291	29
13	(주)형신기계제작소	29176	2917	291	29
14	(주)재원정밀	22241	2224	222	22
15	(주)재원정밀	22249	2224	222	22
16	(주)재원정밀	29294	2929	292	29
17	광선산업	29294	2929	292	29
18	(주)신성크린텍	29175	2917	291	29

Factor화(Factorize) 하면
cluster로서 이용할 수 있다.

```
> str(n2)
'data.frame': 11352 obs. of
 $ name : chr "은성산업(주)" "
 $ five : Factor w/ 400 levels
 $ four : Factor w/ 170 levels
 $ three: Factor w/ 82 levels "
 $ two : Factor w/ 27 levels "
```



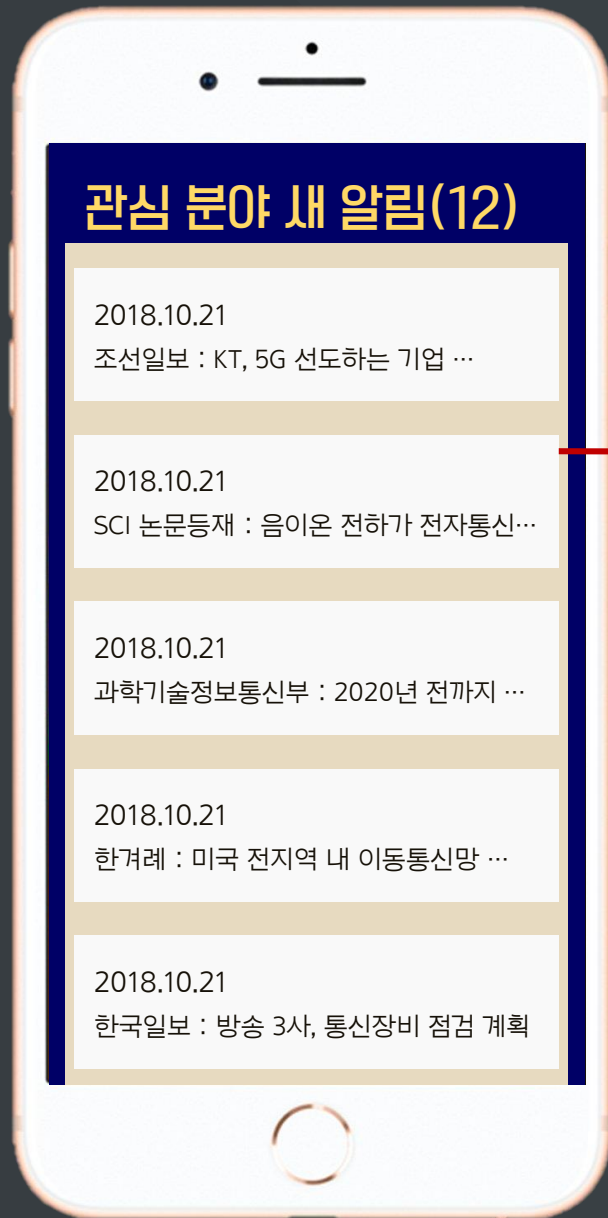

산업 / 기술 분야 최근 소식, 뉴스



데이터 크롤링



업종별 클러스터를 활용하여
유사도 높은 클러스터 內 사용자에게 전달!



관심 분야 새 알림(12)

2018.10.21

조선일보 : KT, 5G 선도하는 기업 ...

2018.10.21

SCI 논문등재 : 음이온 전하가 전자통신...

2018.10.21

과학기술정보통신부 : 2020년 전까지 ...

2018.10.21

한겨레 : 미국 전지역 내 이동통신망 ...

2018.10.21

한국일보 : 방송 3사, 통신장비 점검 계획

입력한 관심사와 관련이 깊은 소식이 있으면 전달받을 수 있다.

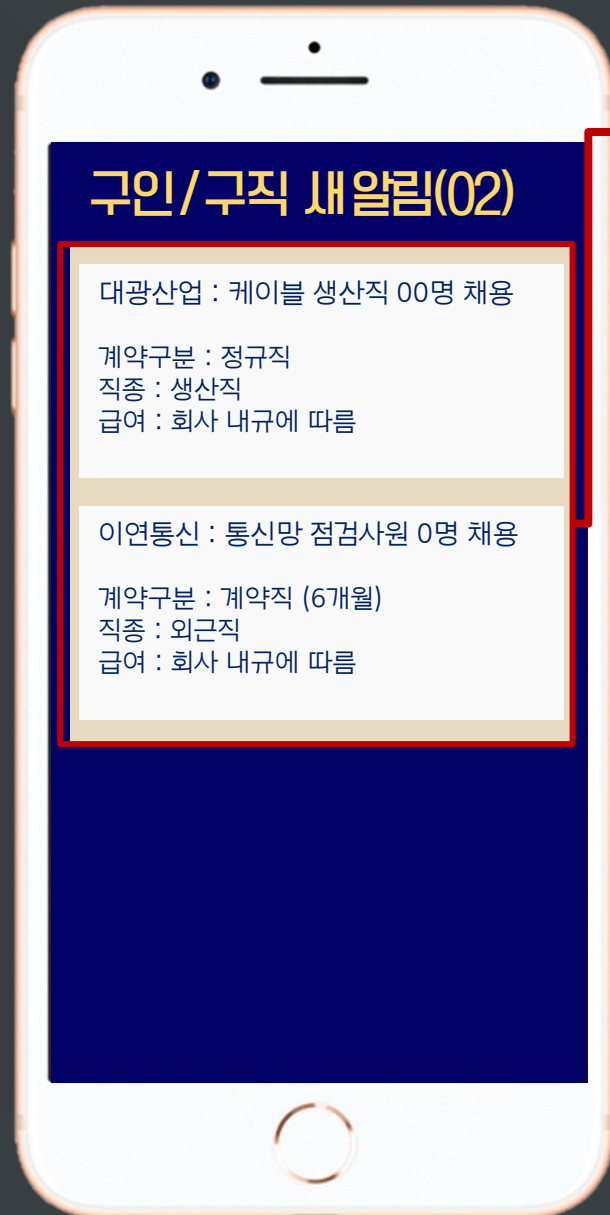
→회원가입 정보를 토대로 근로자의 개인 성향과 일치하는 정보들만 선택적으로 보여줌

→발빠르게 변화하고 있는 기술들에 대한 정보를 쉽게 습득할 수 있음

→기존 기술의 숙련도 향상과 신기술 습득을 용이하게 함으로서 생산성 향상을 기대

2. 추가 활용 방안

-구인/구직 플랫폼



구인, 구직중인 경우 입력한 정보에 따라
구인, 구직 관련 정보가 나타난다.

취직, 이직시 자신의 특기, 보유 기술을 필요로 하는 회사를 직접 찾는 것은
쉽지 않음.

그러나 공단은 많은 기술 기반 산업이 집약된 곳이므로
각 공장이 어떠한 기술을 활용하고 있는지를 알고 있다면 이를 활용할 수
있음.

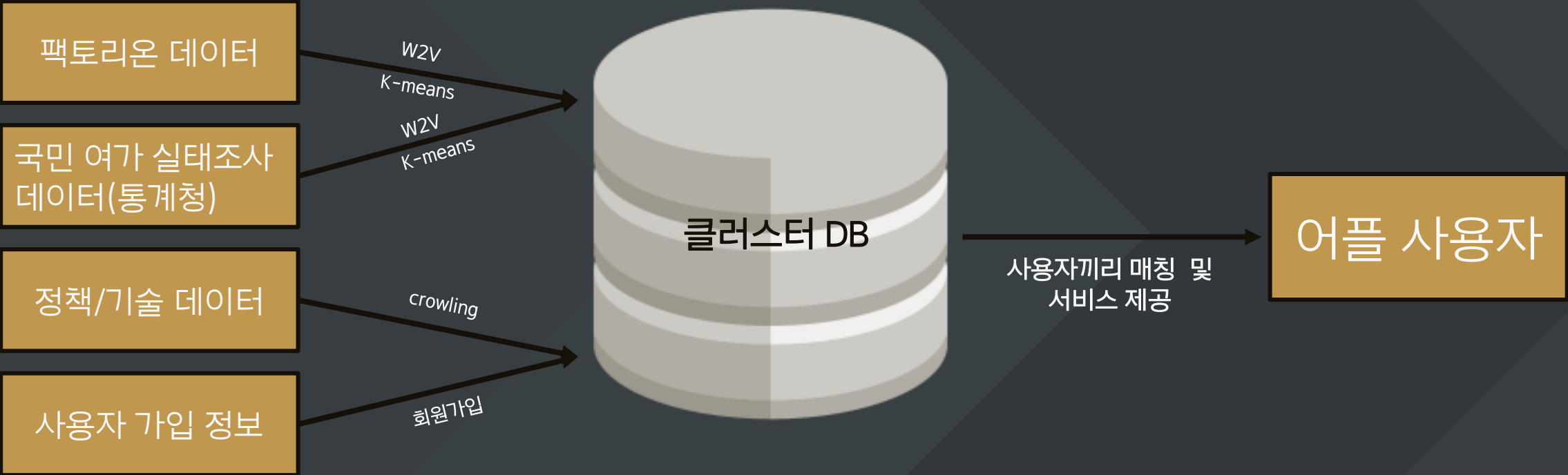
따라서 특기, 보유기술을 기반으로 구인정보를 제공

→ 개인의 특기와 장점에 알맞은 직장을 효율적으로 구할 수 있음.

→기업은 필요에 맞는 인재를 구할 수 있을 것임.

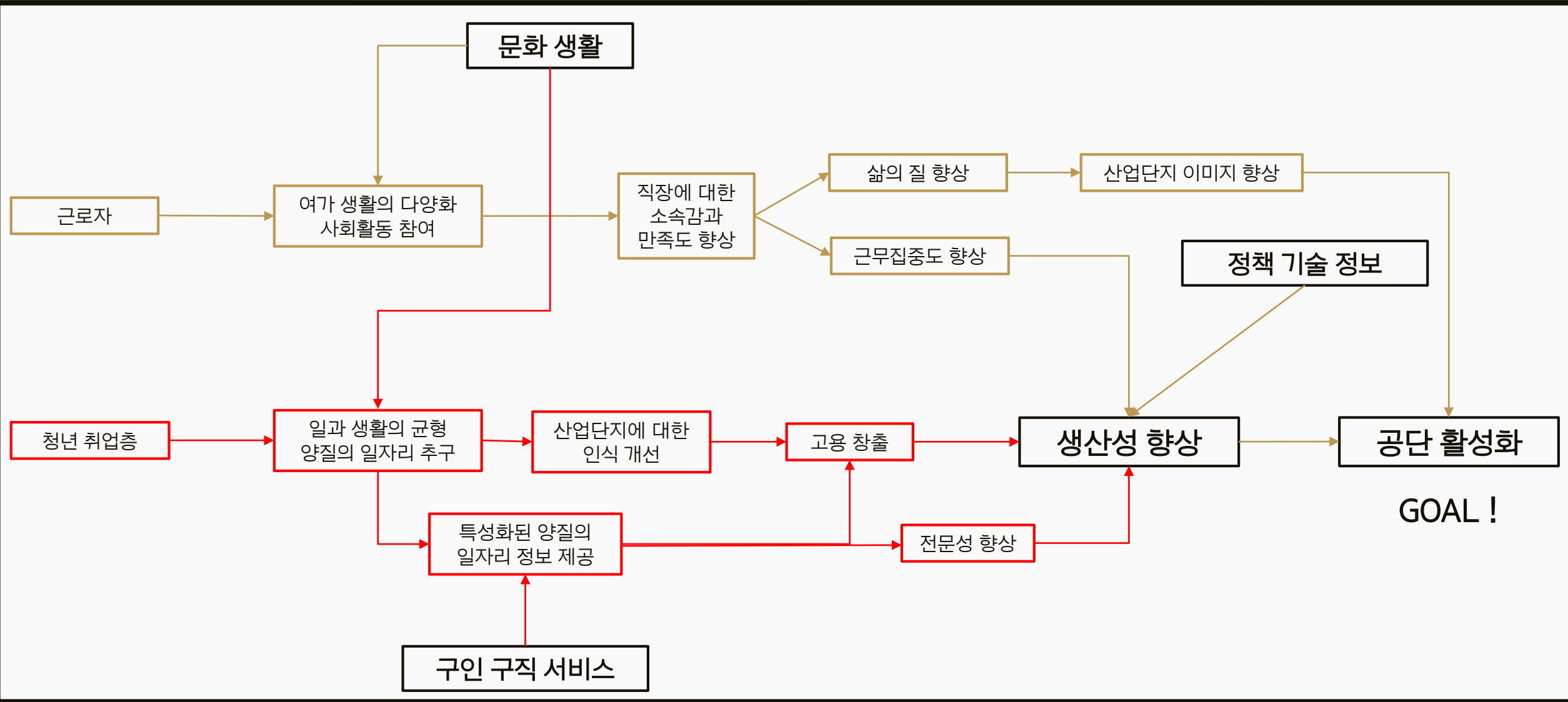
→노동자, 고용주 모두 효과적인 구인,구직이 가능.

데이터 흐름도



기능들을 분리하지 않고
통합 커뮤니티로 만든 이유?

기대효과 흐름도



감사합니다.