

당신의 취향에 딱 맞는 맛집과 명소를 추천합니다!

TEAM URS

IT 융합학부 박영준 20201730

IT 융합학부 박상원 20172097

산업경영공학부 김창영 20181794

Contents.

1 프로젝트 배경 및 목적

2 소프트웨어 개발

3 추천시스템

4 페이지 순서도

5 시연 영상

6 기대 효과

7 향후 계획

Part 1, 프로젝트 배경 및 목적

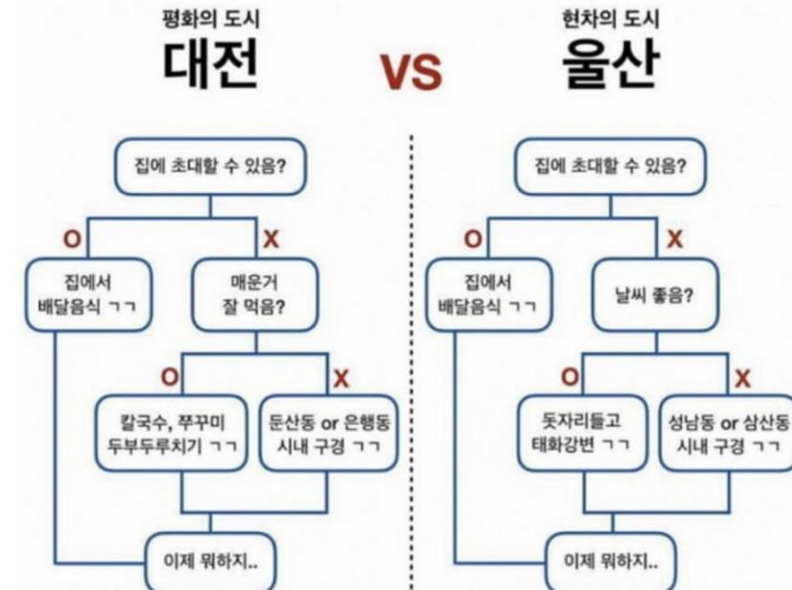


18

부족한 문화생활과 여가 시설, 지하철이 없는 불편한 대중교통, 공기가 나쁜 산업 도시라는 인식이 청년들이 울산을 떠나게 하는 원인

출처 : 울산제일일보(<http://www.ujeil.com>)

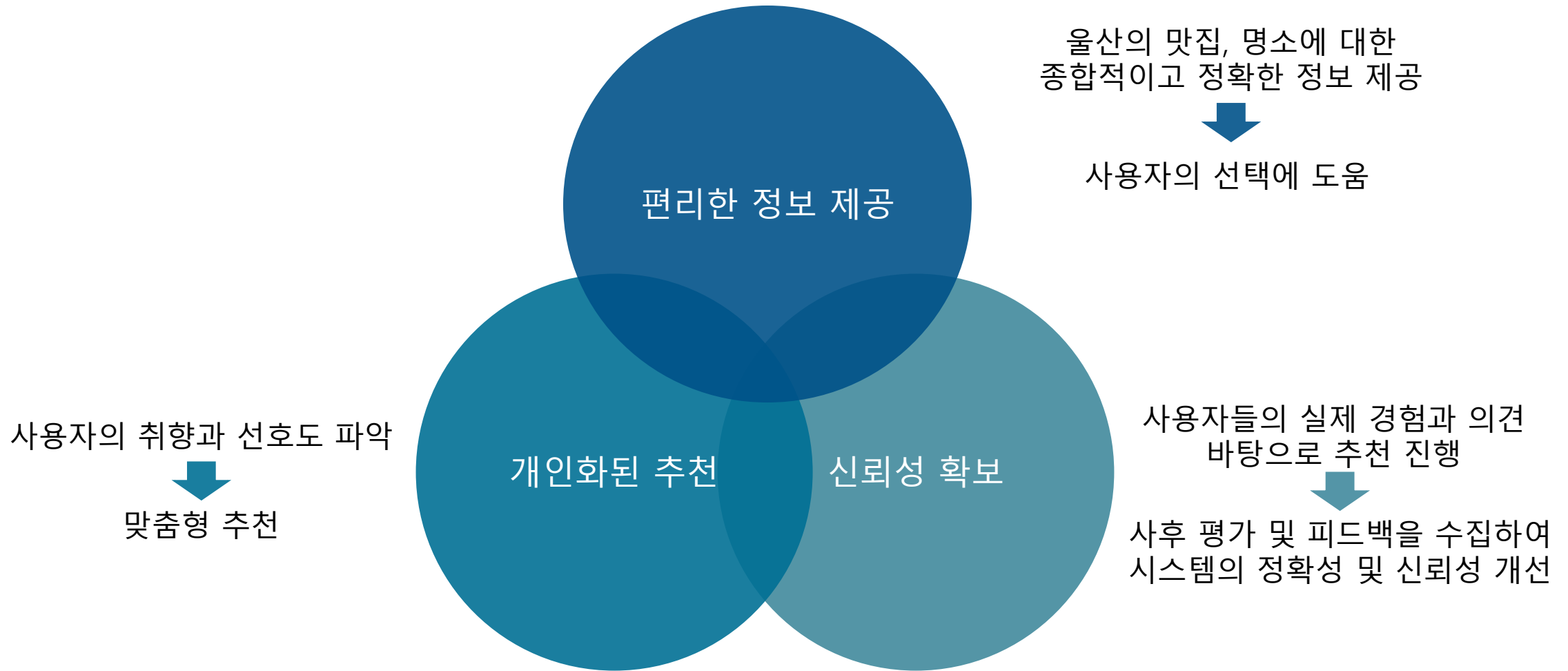
노잼도시 양대산맥 알고리즘



실제로 울산이 관광도시로 발돋움하지 못하고 있다는 전제에 홍보 부족을 문제로 삼는 의견이 지배적이다. 영남알프스, 태화강 등 천혜의 관광자원을 품고도 홍보가 부족해 울산의 이미지가 산업 수도, 노잼 도시로만 그친다는 것이 안타까울 따름이다.

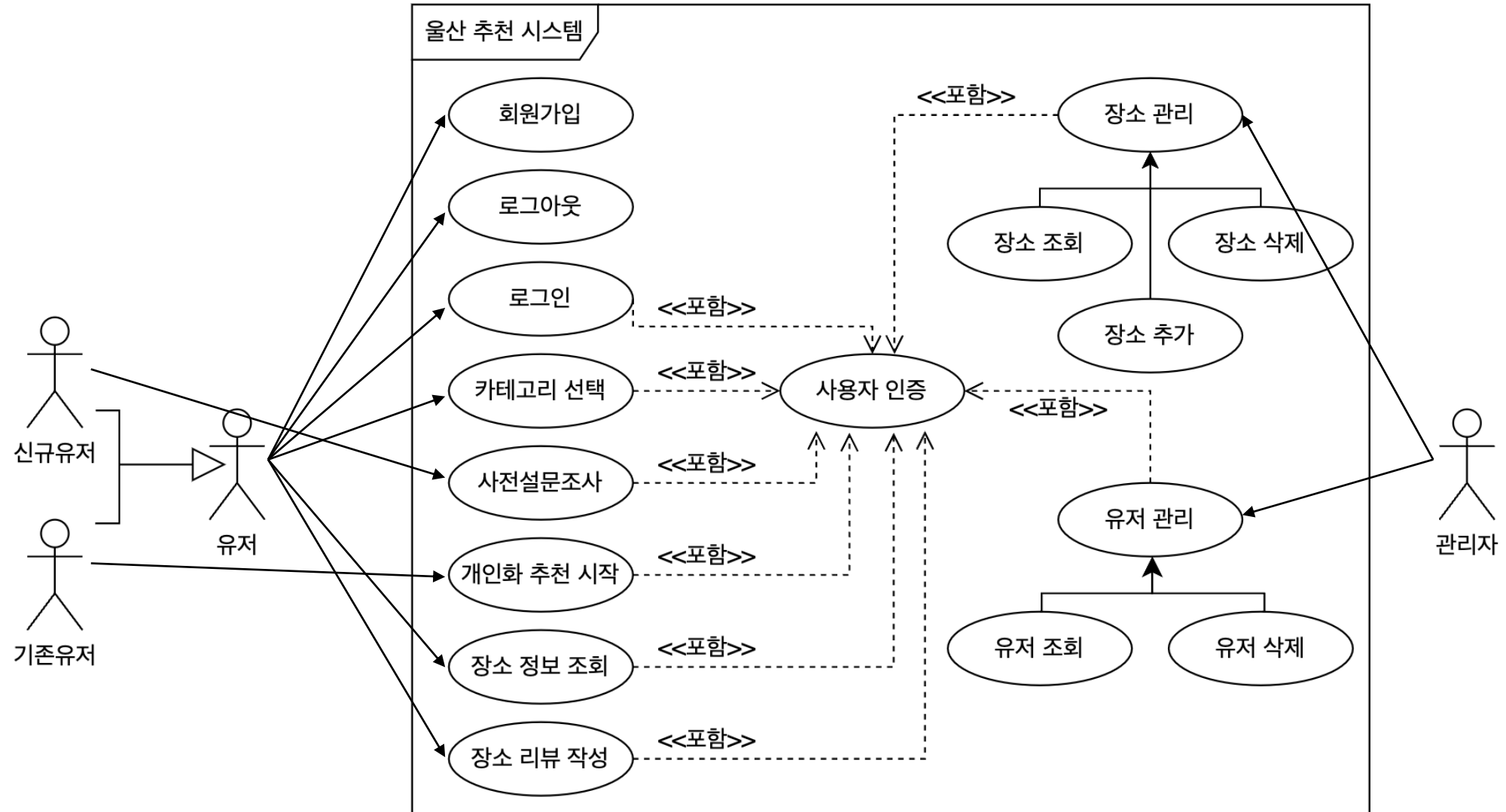
-이종필, "노잼 도시 울산? 유잼 도시 울산을 꿈꾸다!", 울산경제신문, 2023.02.14

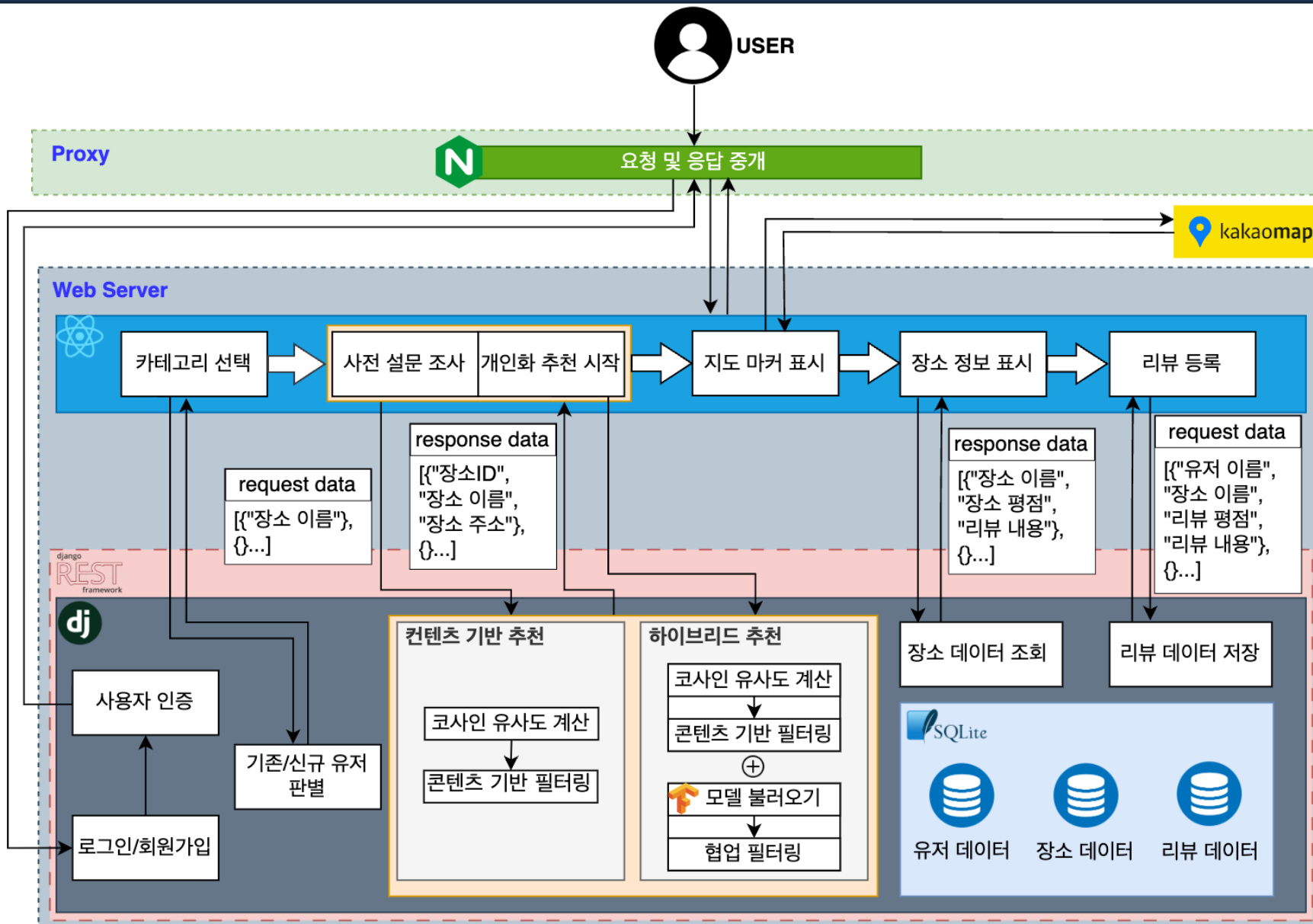
“ 울산의 다양한 음식과 관광지 경험에 도움 ”



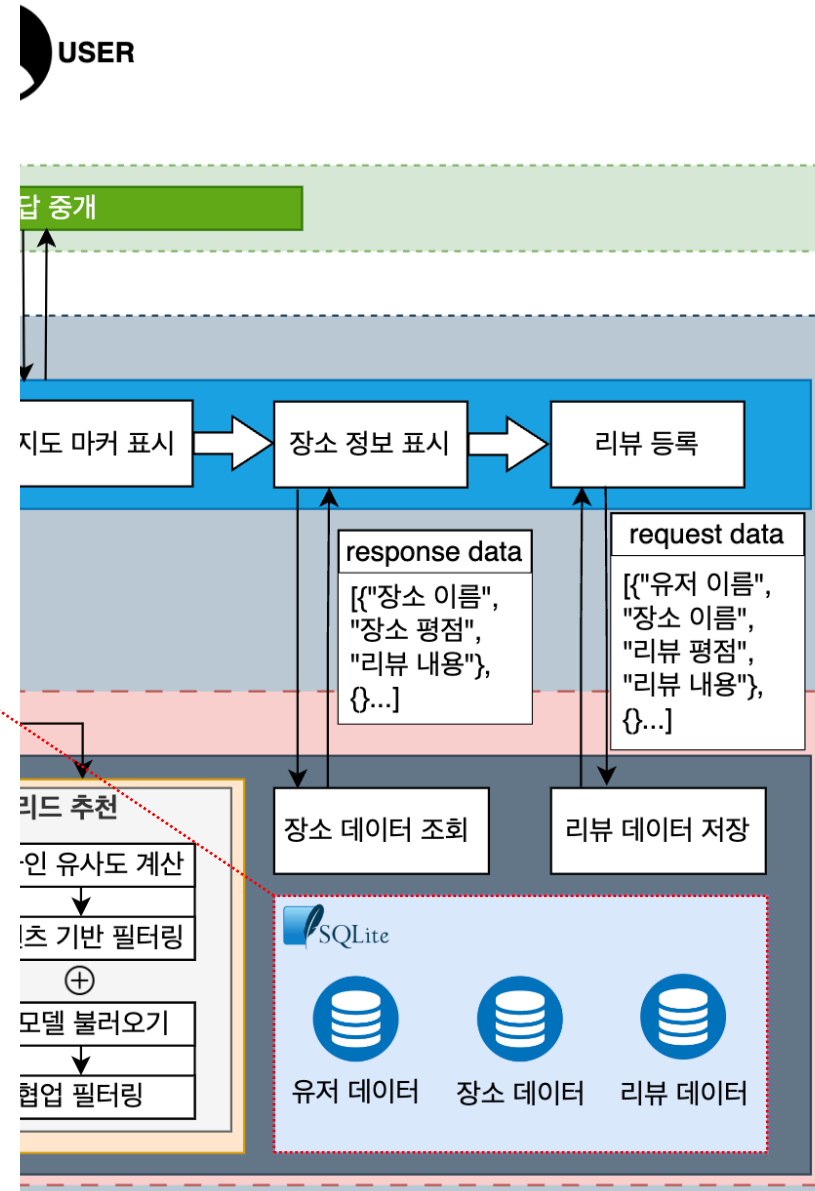
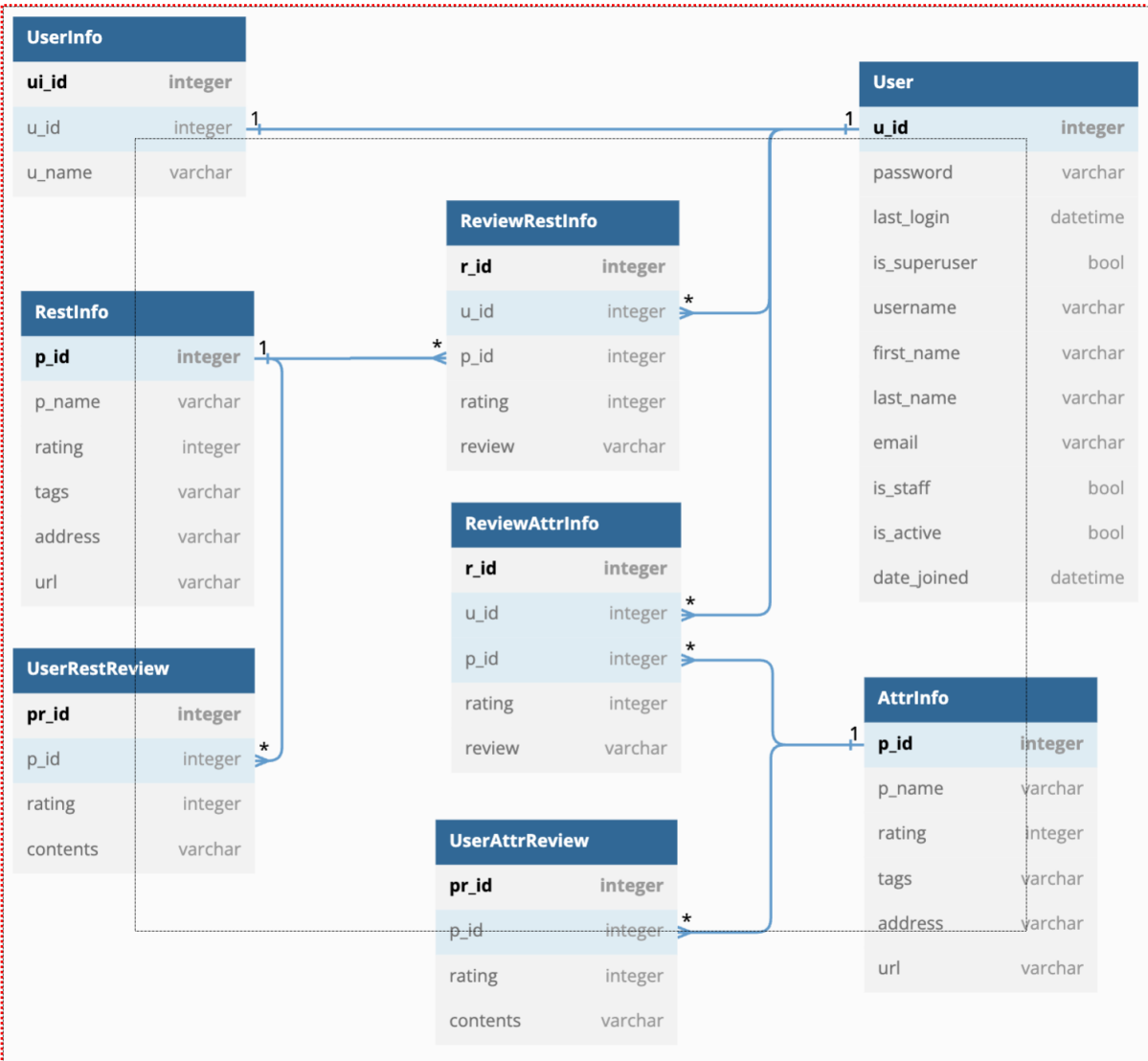
Part 2,

소프트웨어 개발

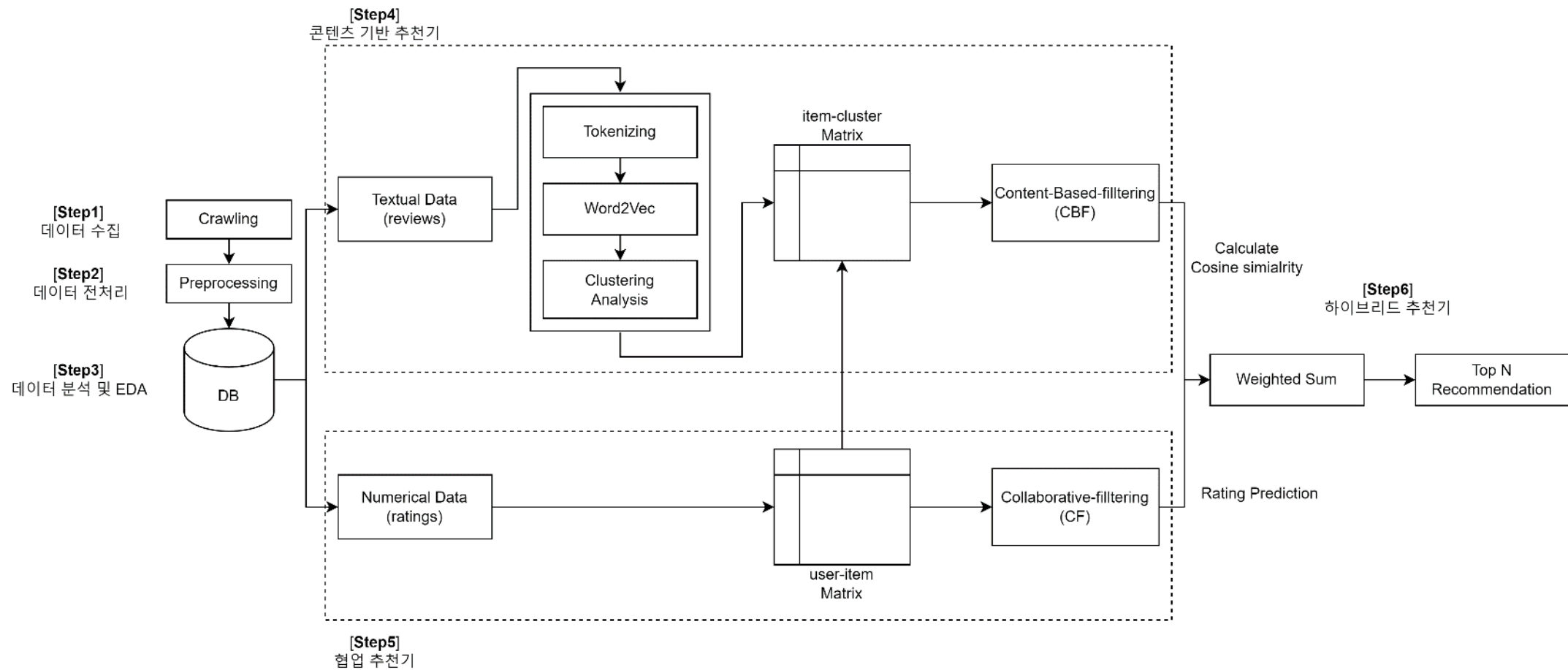




Part 2, ER 다이어그램



Part 3, 추천 시스템



[1단계] 데이터 수집

[Table 1] 맛집 데이터 개요

	사용자 수	장소 수	평점 수	리뷰 수
계	24,545명	492곳	54,500개	40,225개

[Table 2] 명소 데이터 개요

	사용자 수	장소 수	평점 수	리뷰 수
계	9,153명	495곳	15,405개	10,548개

- Kakao map과 google map에서 울산 음식점/울산 명소를 검색한 결과를 기반으로 각 장소에 대한 기본 정보, 평균평점, url정보와 사용자별 리뷰데이터를 수집

1) 데이터 클렌징(Data Cleansing)

수집한 리뷰데이터에는 이모티콘 특수문자들이 존재, 텍스트 분석에 부적합한 형태이기 때문에 제거

2) 불용어 제거

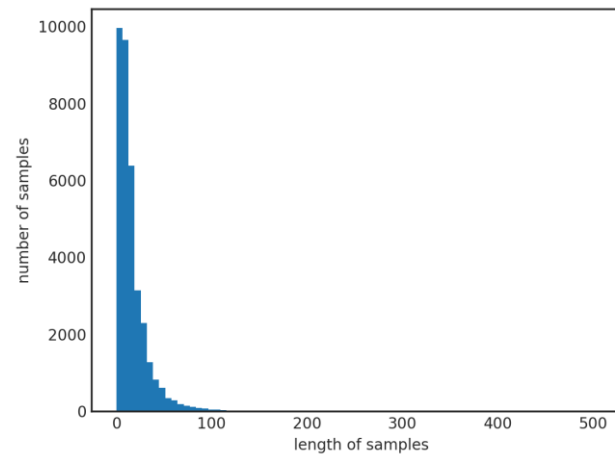
불용어는 "아", "휴 "어" 와 같이 문자의 의미를 파악하는 데 기여하지 않는 단어를 의미

불용어 제거 시 얻을 수 있는 효과

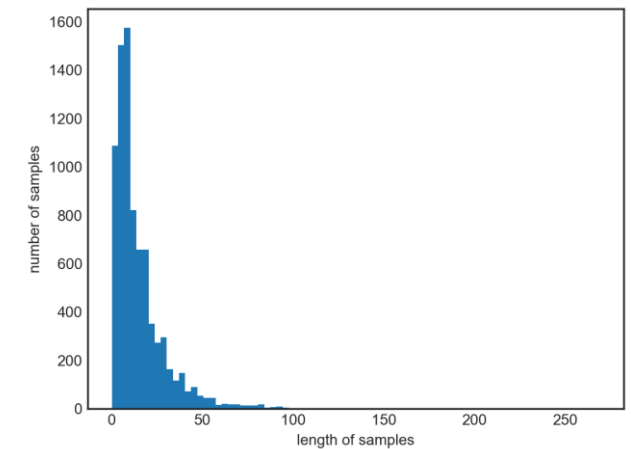
- 문서 크기 감소
- 메모리 절약
- 노이즈 제거
- 분석 정확성 향상



676개의 불용어 사전 정의하여 불용어 제거



[Fig 2] 맛집 리뷰 길이



[Fig 3] 명소 리뷰 길이

1) 형태소 분석

한국어 정보 처리를 위한 패키지인 KoNLPy의 Mecab을 사용

2) 키워드 분석

- 맛집 데이터

전체 단어 수 : 18,321개

빈도 2개 이하인 희귀 단어 수 : 10,849개

희귀 단어 비율 : 59.22%

- 명소 데이터

전체 단어 수 : 10,849개

빈도 2개 이하인 희귀 단어 수 : 6,342개

희귀 단어 비율 : 64.47%



[Fig 4] 맛집 워드 클라우드



[Fig 5] 명소 워드 클라우드

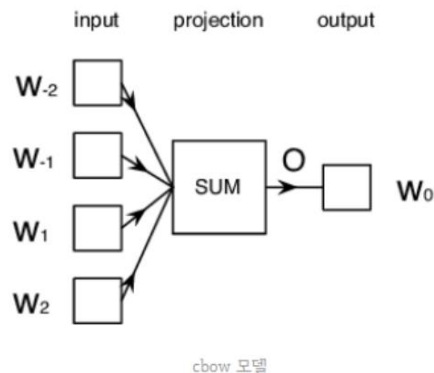
1) 토큰화

한국어 정보 처리를 위한 패키지인 KoNLPy의 Mecab을 사용

2) 워드 임베딩

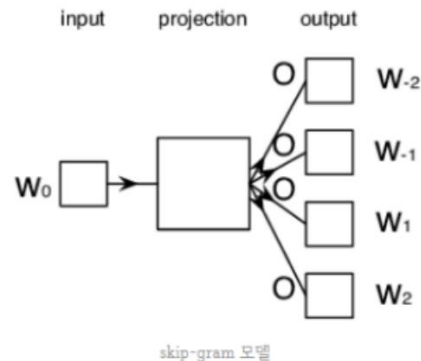
단어를 벡터 형태로 변환하는 작업

- CBOW



중심의 단어로 주변에 있는 단어들을 예측하는 방법

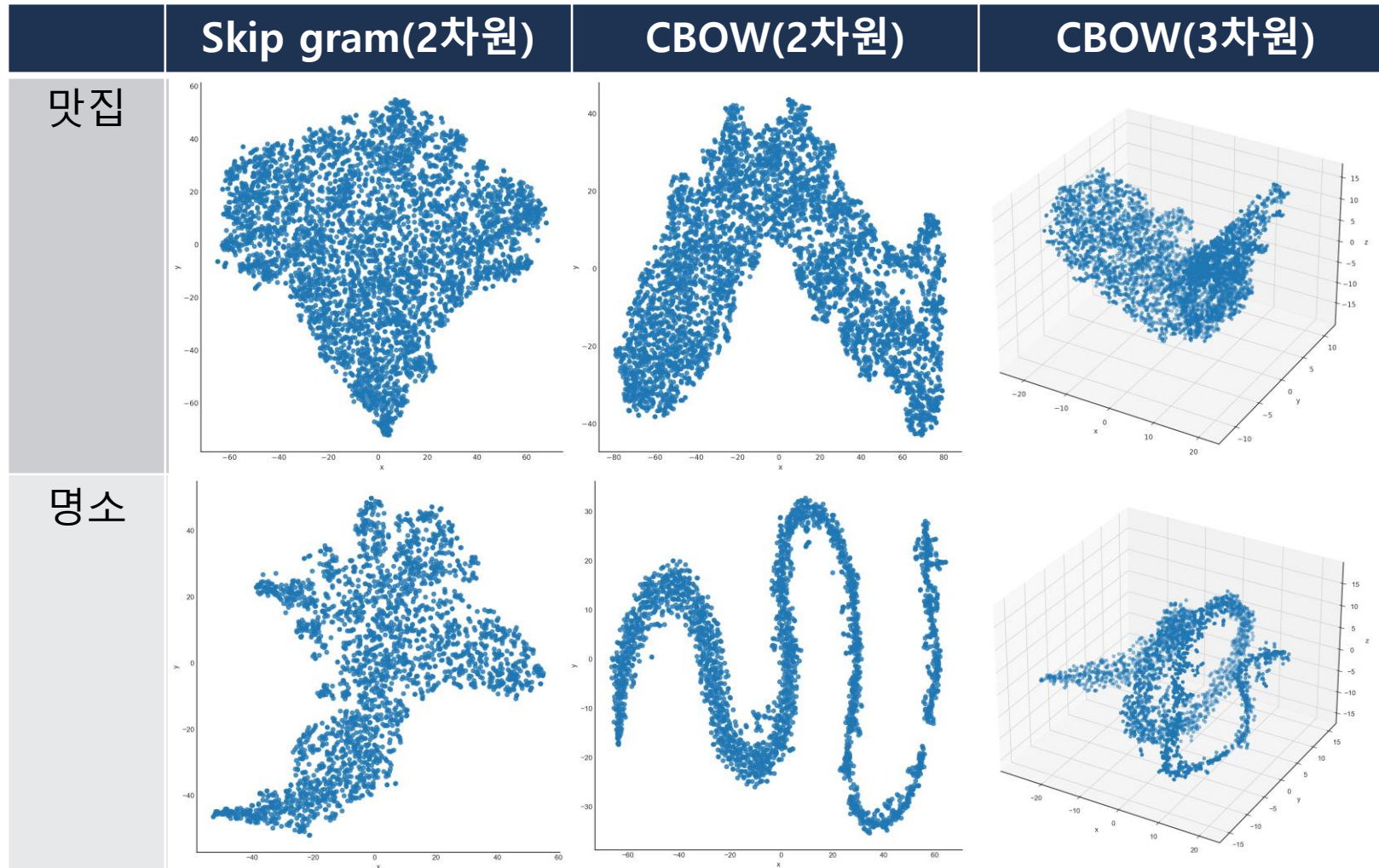
- Skip-gram



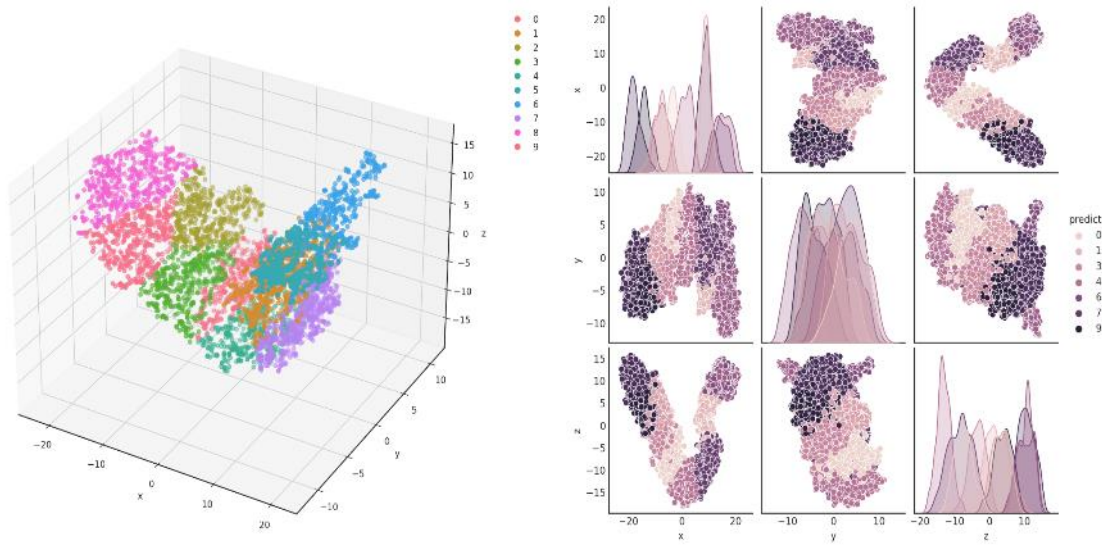
주변에 있는 단어들을 가지고, 중간에 있는 단어들을 예측하는 방법

[4단계] 콘텐츠 기반 추천

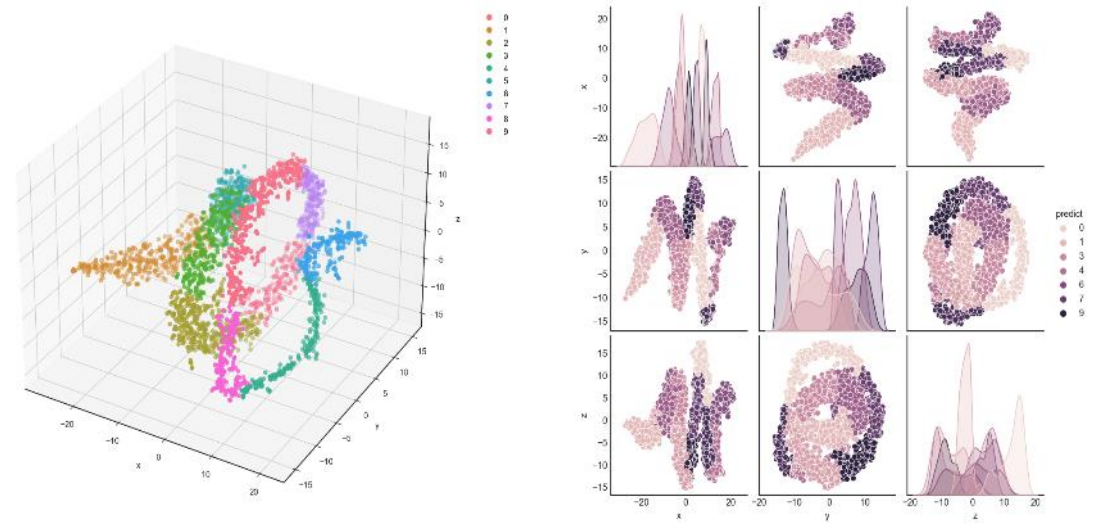
[Table 3] t-SNE를 활용한 시각화



3) 군집 분석



[Fig 6] 맛집 데이터 K-means 군집화



[Fig 7] 명소 데이터 WARD 군집화

4) 콘텐츠 기반 필터링 (CBF)

- Item-Cluster Matrix

	c_0	c_1	c_2	...	c_9
i_0	0	0	0	...	1
i_1	0	1	...		4
i_2	1	16	...	1	0
i_3	...				
...					
i_n					

[Fig 8] Item-Cluster Matrix

아이템별 리뷰들을 형태소별로 분리
토큰들을 군집별로 평균을 집계

- 코사인 유사도

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}},$$

아이템 간의 코사인 유사도를 비교
유사도 기준 상위 K개 추천

1) 협업 필터링 (CF)

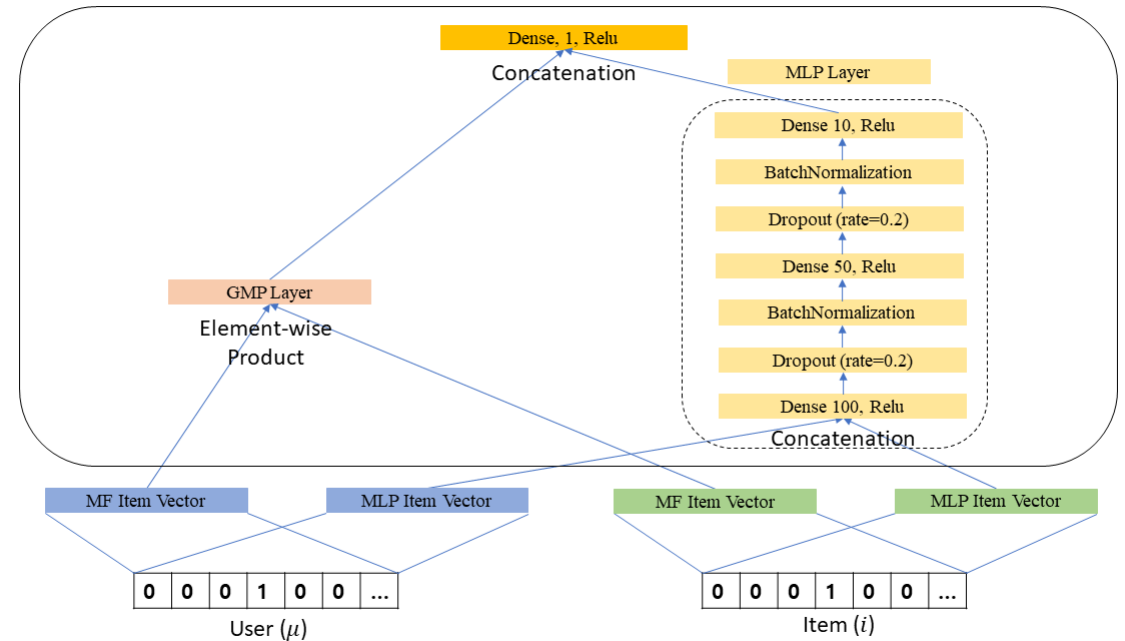
- User-Item Matrix

	i_0	i_1	i_2	...	i_n
u_0	4		3	...	
u_1		1.5			
u_2	1	2	...		
u_4	...				
...					
u_k				...	

[Fig 9] User-Item Matrix

사용자와 아이템 간의 평점을 나타낸 행렬

- 모델



[Fig 10] NCF 모델 아키텍처

1) 하이브리드 추천의 필요성

콘텐츠 기반 필터링의 단점

- 아이템 자체의 정보에 의존
- 사용자의 동적인 취향 변화와 같이 개인화 요소를 고려하기 어려움

협업 필터링의 단점

- Cold Start Problem -> 새로 추가된 아이템에 대해서 충분한 상호 작용 불가
- Long Tail Problem -> 희소한 아이템에 추천 제한



하이브리드(CBF + CF) 추천 시스템 도입

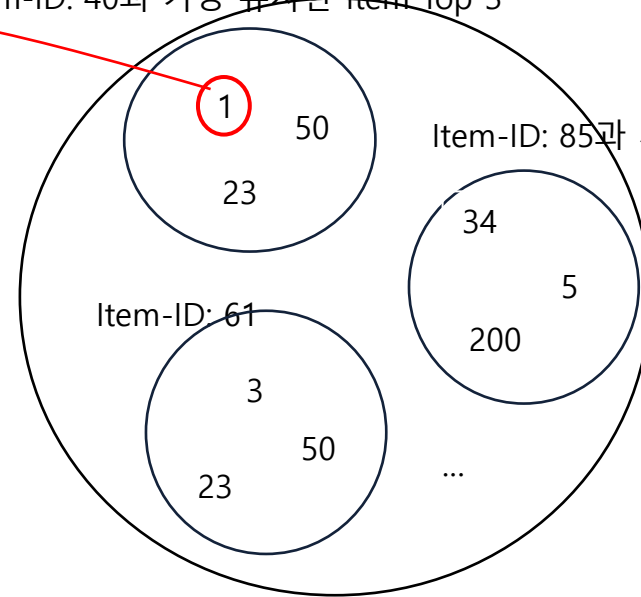
2) 하이브리드 추천 방식

Collaborative-Filtering
(CF)

Item ID	Prediction rating
0	1
1	4.1
2	2.5
3	3.2
....	
n-1	4
n	4.2

User가 선호하는 Item ID
ex) 40, 85, 61, 78, 90, ...Contents-Based Filtering
(CBF)

Item-ID: 40과 가장 유사한 Item Top 3

CBF로 추천된 Item에는
CF로 예측된 평점에 가중치를 부여함

1) 평가 방법

- RMSE

$$RMSE = \sqrt{\frac{1}{N} \sum_i^N (pred_i - target_i)^2}$$

예측값과 실제값 간의 오차를 평가하는 지표

- Hit Rate@K

$$HitRate@K = \frac{\text{사용자가 실제로 선호한 아이템 개수}}{\text{추천한 아이템 개수 (K개)}}$$

사용자가 실제로 선호하는 아이템을 몇 개나 정확하게 추천했는지를 평가하는 지표

2) 실험 결과

[Table 6] 맛집 협업 필터링 RMSE 결과

	KNNwith means	SVD	SVD++	NMF	NCF
RMSE	1.144	1.026	1.048	1.120	1.103

[Table 7] 명소 협업 필터링 RMSE 결과

	KNNwith means	SVD	SVD++	NMF	NCF
RMSE	0.975	0.887	0.899	0.971	0.913

1) 실험 결과

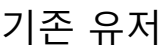
[Table 8] 맛집 Hit Rate 결과

	KNNwith means	SVD	SVD++	NMF	NCF	Hybrid (CBF+NCF)
Hit Rate@10	0.39	0.49	0.48	0.41	0.51	0.53
Hit Rate@15	0.58	0.633	0.627	0.607	0.633	0.633
Hit Rate@20	0.73	0.745	0.76	0.76	0.77	0.775

[Table 9] 명소 Hit Rate 결과

	KNNwith means	SVD	SVD++	NMF	NCF	Hybrid (CBF+NCF)
Hit Rate@10	0.68	0.63	0.66	0.62	0.7	0.71
Hit Rate@15	0.846	0.826	0.84	0.84	0.862	0.862

Part 4, 페이지 순서도



초기 화면

로그인

카테고리(맛집/명소)
선택

하이브리드 추천

지도상 마커 표기

리뷰 등록

신규 유저

데이터가 일정 수준 이상
도달하면 기존 유저로 편입

회원가입

사전 설문지 작성

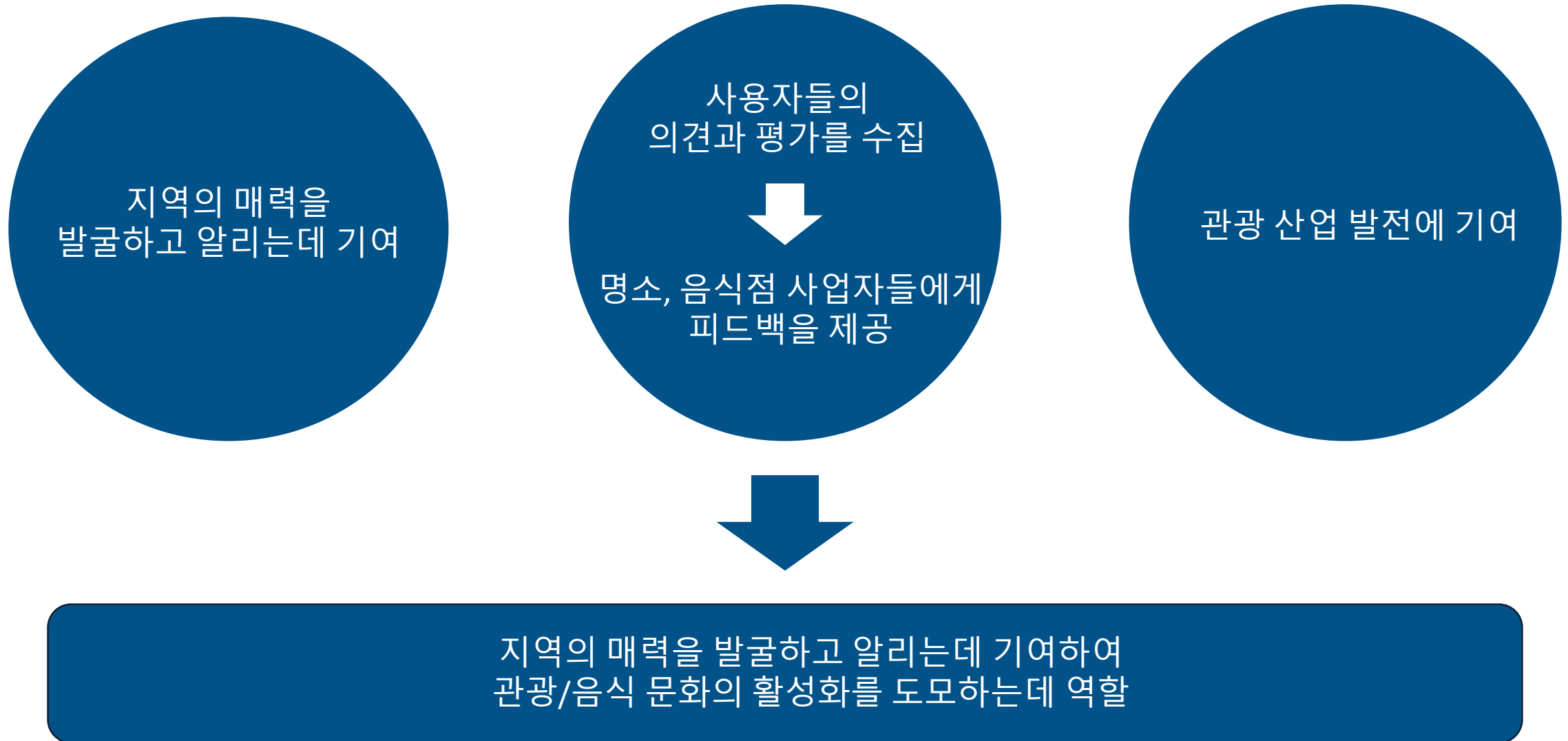
콘텐츠 기반 추천

Part 5,

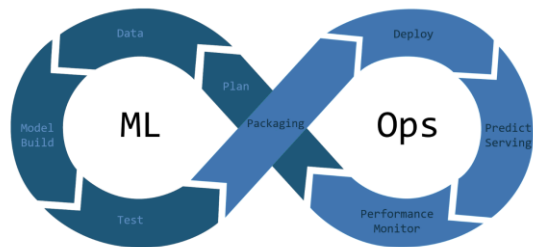
시연 영상

Part 6,

기대 효과

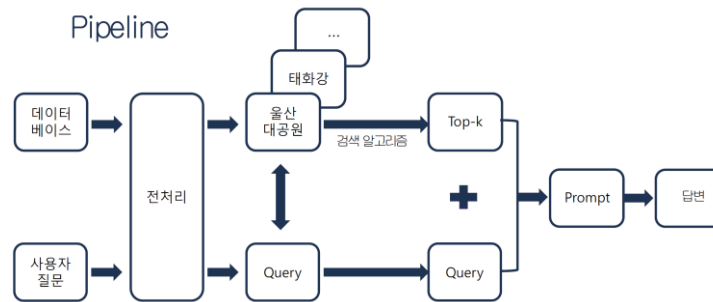


Part 7, 향후 계획



- 초기 데이터 수집과 분석을 통해 시스템을 구축했지만, 새로운 데이터의 수집과 기존 데이터의 업데이트 필요
- MLOps를 도입하여 데이터 수집부터 모델 배포까지의 단계를 체계적으로 관리

다양한 형태의 추천



- ChatGPT를 활용한 챗봇 기능을 해당 플랫폼에 탑재하여 보다 다양한 추천 서비스를 제공

지속적인 데이터 수집과 관리

확장성과 유연성

- 다양한 플랫폼과의 연동을 통해 사용자에게 편의성 제공

「

배운 지식과 역량을 활용한 사회적 가치 실현!

」