

# 맛집 빅데이터 분석 웹서비스

번호: 26

팀 명: 금정산 독수리

분과: B

팀원: 최경섭 (201724599), 장희승 (201724561)

지도교수: 홍봉희 교수님

# 목차

- 1. 요구조건 및 제약사항 분석에 대한 수정사항**
  - 1.1 요구조건
  - 1.2 제약 사항 분석에 대한 수정사항
    - 1.2.1 기존 제약사항에 대한 수정사항
    - 1.2.2 추가 제약사항 및 대책
- 2. 설계 상세화 및 변경 내역**
  - 2.1 크롤링 서버
  - 2.2 카테고리 분류 알고리즘
  - 2.3 최적화 알고리즘
  - 2.4 최신화 알고리즘
- 3. 갱신된 과제 추진 계획**
- 4. 구성원 별 진척도**
- 5. 보고 시점까지의 과제 수행 내용 및 향후 계획**
  - 5.1 google maps 가게 크롤링
  - 5.2 각 알고리즘에 따른 결과
    - 5.2.1 카테고리 분류
    - 5.2.2 신뢰성
    - 5.2.3 최신화
  - 5.3 데이터베이스 설계 계획

## 1. 요구조건 및 제약사항 분석에 대한 수정사항

### 1.1 요구조건

빅데이터 기술을 통하여 신뢰할 수 있는 맛집 랭킹과 다양한 카테고리 및 개인의 입맛을 고려한 리뷰 정보를 제공하는 웹 서비스 구축한다.

- 데이터 수집을 위한 크롤링 기능 구현  
(google maps를 통한 데이터 수집)
- 사용자의 편의성을 위한 카테고리 분류  
(알고리즘을 통한 카테고리 분류)
- 정보의 신뢰성을 위한 신뢰성 및 최신화 알고리즘 구현  
(신뢰성 및 최신화 알고리즘 설계)

### 1.2 제약 사항 분석에 대한 수정사항

#### 1.2.1 기존 제약사항에 대한 수정사항

- 초기 데이터의 신뢰성 및 데이터의 방대함 문제  
→ 리뷰가 없는 가게 제외 및 관련성 높은 순 리뷰 20개 내외 수집

#### 1.2.2 추가 제약사항 및 대책

- 방대한 리뷰의 양으로 인한 최신화 문제  
→ 관련성 높은 순 정렬 후 리뷰 수집
- 관련성 높은 순 정렬 후 리뷰 수집에 따른 최신화 알고리즘의 무력화  
→ 초기 데이터 이후 추가되는 리뷰에 한정하여 최신화 알고리즘 적용
- 음식점 리스트와 리뷰 크롤링이 개별적으로 이루어져야 함을 인지  
→ 2개의 모듈을 설계하여 개별적으로 크롤링 진행. 이후 알고리즘을 활용하여 하나의 csv 파일로 통합

## 2. 설계 상세화 및 변경 내역

### 2.1 크롤링 서버

- 구글맵 부산광역시 맛집 가게들을 크롤링하여 데이터베이스 저장
- 크롤링을 통한 데이터를 분류하여 데이터베이스 저장
- 리뷰 작성 시간이 일정하지 않으므로 실시간으로 새롭게 작성된 리뷰를 업데이트

### 2.2 카테고리 분류

- 크롤링을 통한 가게 정보를 카테고리별로 분류

### 2.3 신뢰성 및 최신화 알고리즘

- 알고리즘을 통해 기존에 작성되었던 리뷰 신뢰성 판단 및 최신화
- 새롭게 작성된 리뷰에 대한 필터링 기능 수행

## 3. 갱신된 과제 추진 계획

5월		6월				7월				8월				9월			
3주	4주	1주	2주	3주	4주	1주	2주	3주	4주	1주	2주	3주	4주	1주	2주	3주	4주
Sample data 수집 및 카테고리 분류 알고리즘 설계																	
				신뢰성 알고리즘 설계													
							최신화 알고리즘 설계										
								중간 보고서 작성 및 발표									
										고도화 및 예외 처리 작업							
										웹 서비스 및 데이터 베이스 구축							
														최종 보고서 작성 및 심사 준비			
														발표 준비 및 마무리 작업			

#### 4. 구성원 별 진척도

이름	작업 목록
최경섭	최신화 알고리즘 설계 및 구현 완료 신뢰성 알고리즘 설계 및 구현 완료 데이터 베이스 구축 진행 중
장희승	Sample data 수집 완료 카테고리 분류 알고리즘 설계 및 구현 완료 API 데이터 필터링 및 적용 진행 중

#### 5. 보고 시점까지의 과제 수행 내용 및 향후 계획

##### 5.1 google maps 가게 크롤링

A1	A	B	C	D	E	F
1		가게이름	평점	주소	리뷰수	카테고리
2	0	솔밭집	3.8	선동 341-4	55	한식
3	1	참맛집식당	5	남산동 3-9번지 금단로101번길9-1 1층 일부 금정구 부산광역시 KR	1	음식점
4	2	해묵	4.3	구남로24번길 8	1,265	일식
5	3	해운대암소갈비집	3.9	중동2로10번길 32-10	3,050	한국식 BBQ
6	4	동백섬횡집	4	해운대해변로209번나길 17	680	회 전문점
7	5	은행나무집	4.3	선동 430-2	37	한식
8	6	부산모밀	4.2	남산동 17-26	246	소바 전문점
9	7	60년전통할매국밥	4.2	중앙대로533번길 4	2,195	돼지국밥 전문점
10	8	금수복국 해운대본점	4.1	중동1로43번길 23	5,837	복어 요리 전문식당
11	9	손영환 비빔국수 본점	3.9	두구동 601-6	571	국수 전문점
12	10	송정3대국밥	3.9	서면로68번길 33	2,884	돼지국밥 전문점
13	11	개미집 해운대점	3.8	구남로 34	3,435	낙지전문점
14	12	고육	4.2	광남로 6	868	일식
15	13	기장순칼국수	4	서면로 56	1,523	칼국수집
16	14	속씨원한대구탕 본점	4	달맞이길62번길 28	1,587	한식
17	15	나가하마 만게즈 한국본점	4.4	우동1로 57	873	라면
18	16	신발원(新發園)	4.1	대영로243번길 62	2,553	중식
19	17	박옥희 할매집원조복국	4.2	달맞이길62번길 28	1,198	복어 요리 전문식당
20	18	선팀맛집	4.5	재송동 918-4번지 해운대구 부산광역시 KR	2	한식
21	19	장수원돼지국밥	3.6	남산동 91-10번지 장수원돼지국밥 금정구 부산광역시 KR	16	음식점
22	20	쌍둥이돼지국밥 본점	3.9	유엔평화로 35-1	2,166	돼지국밥 전문점
23	21	본전돼지국밥	4.1	중앙대로214번길 3-8	3,846	돼지국밥 전문점
24	22	점레네	3.9	좌동로10번길 72	319	한국식 소고기 전문 음식점
25	23	맛집	3.4	초량중로 11	48	한식
26	24	삼오정	4.2	서면로68번길 11	419	한식
27	25	금하아구찜	4.4	남산동 122-16	62	한식
28	26	원조석대추어탕	4.1	석대동 489-1	419	추어탕집
29	27	밀양순대돼지국밥 해운대점	3.9	구남로 28	2,951	한식
30	28	라호짬뽕	4	우동 마린시티1로 137	193	중식
31	29	神田そば   칸다소바 부산서면점	4.3	동천로 105	809	라면
32	30	청따오포차	4.4	남산동 남산로	48	중식
33	31	맛있는 오리	4.4	구동 신천로35번길 선두	46	음식점

busan.py 코드 실행결과 csv파일

## 5.2 각 알고리즘에 따른 결과

- 신뢰성 알고리즘과 최신화 알고리즘의 통합

- 신뢰성 및 최신화 알고리즘

추후 리뷰 크롤링이 완료된 후 실제 csv파일을 활용하여 수행할 계획

<예제 파일을 활용한 결과물>

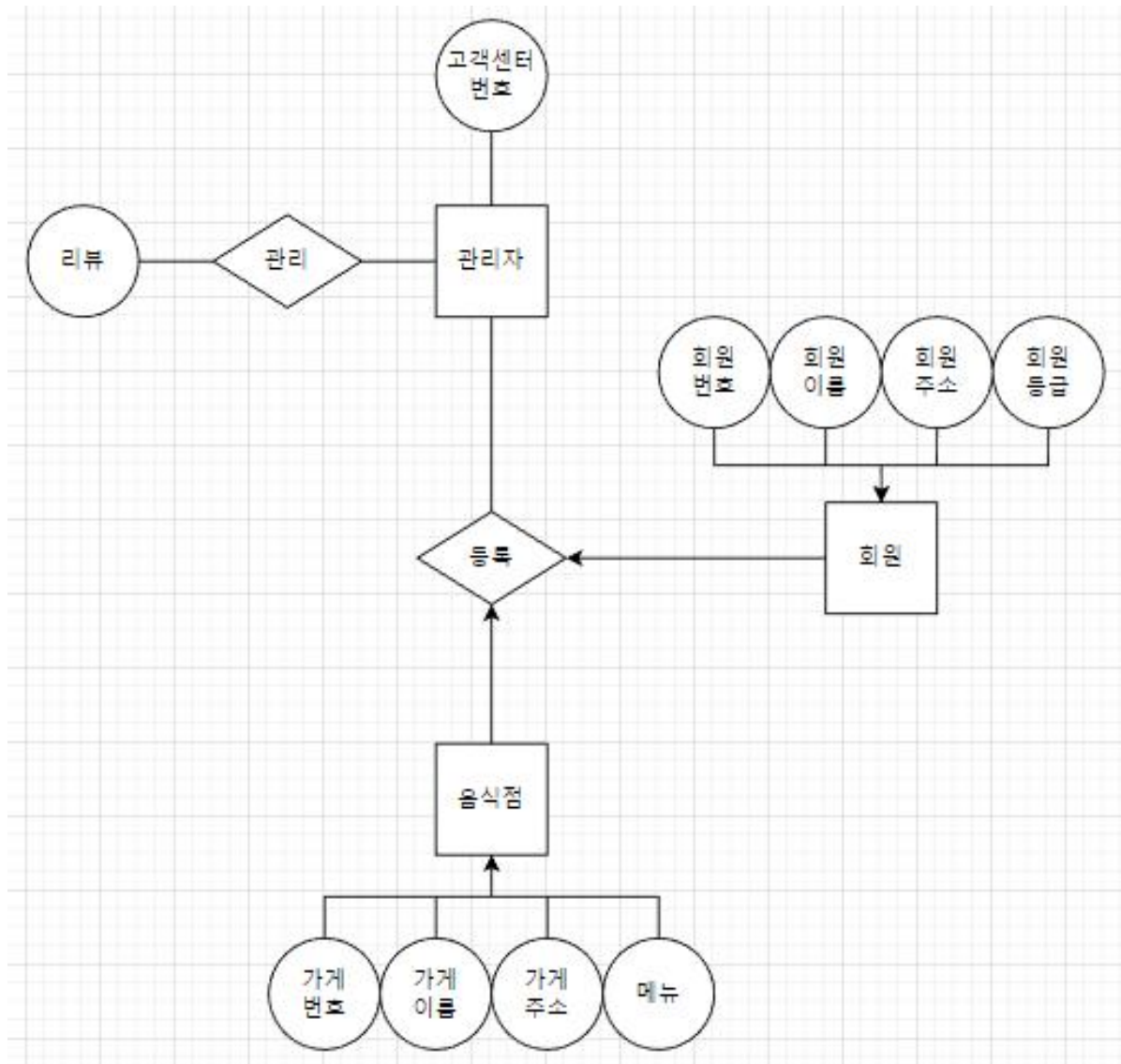
가게이름										
	A	B	C	D	E	F	G	H	I	J
1	가게이름	ID	별점	날짜	리뷰					
2	술발집	jinyoung l		5 20220627	맵지않아서 초등학생도 잘먹었습니다~					
3	술발집	15H9 C		3 20220624	추억의 맛입니다. 양념이 예전에 비해 많이 약해 졌더군요.					
4	술발집	모란모란		4 20220426	그냥 저냥 먹을만 하네요					
5	술발집	busan		4 20220612	아쉽네요.					
6	술발집	busan		3 20220608	soso					
7	참맛집식당	hello		5 20220428	맛있네용					
8	참맛집식당	bye		3 20210527	그저 그렇습니다.					
9	참맛집식당	ㅋㅋ		2 20220617	예전같지 않네요					
10										

리뷰 예제 csv 파일

	A	B	C	D	E	F	G	H	I	J	K
1	가게이름	평점	ID	별점	날짜	리뷰					
2	술발집	3.5	jinyoung l		5 20220627	맵지않아서 초등학생도 잘먹었습니다~					
3			15H9 C		3 20220624	추억의 맛입니다. 양념이 예전에 비해 많이 약해 졌더군요.					
4			모란모란		4 20220426	그냥 저냥 먹을만 하네요					
5			busan		3.5 20220612	아쉽네요.soso					
6	참맛집식당	3.33	hello		5 20220428	맛있네용					
7			bye		3 20210527	그저 그렇습니다.					
8			ㅋㅋ		2 20220617	예전같지 않네요					
9											

신뢰성 알고리즘에 따라 작성된 csv 파일

## 5.3 데이터베이스 설계 계획



E-R다이어그램을 이용한 개념적 설계