

맛집 빅데이터 분석 웹서비스

번호: 26

팀 명: 금정산 독수리

분과: B

팀원: 최경섭 (201724599), 장희승 (201724561)

지도교수: 홍봉희 교수님

목차

1. 과제 배경

- A. 국내 맛집 빅데이터 현황
- B. 기존 문제점

2. 대상 문제 및 요구조건 분석

- A. 기존 맛집 데이터 제공 시스템
- B. 제안할 맛집 데이터 제공 시스템

3. 설계문서(설계)

- A. 데이터 수집
- B. 개발환경
 - 1. 빅데이터 분석
 - 2. 웹서비스 구축

4. 현실적 제약 사항 및 대책(제약 사항)

- A. 데이터 수집 제약 사항
- B. 소프트웨어 제약 사항

5. 개발일정 및 역할 분담 (개발일정)

- A. 일정
- B. 역할분담

1. 과제 배경

A. 국내 맛집 빅데이터 현황

최근 평점 시스템의 활성화 및 관광과 맛집의 연계로 맛집 빅데이터 제공 시스템이 늘어나고 있다. 제공 주체는 기업체부터 관광 활성화를 원하는 지자체까지 다양하며 그중 네이버, 카카오, 구글과 같은 대형 기업체가 제공하는 시스템을 사람들이 많이 사용한다.

1) 지자체 시스템 현황

지자체의 시스템은 관광 연계를 목표로 구축되어 있다. 해당 지역의 업체를 소개하는 형식으로 정보를 제공하며 관광 코스를 연계하여 제안하기도 한다. 하지만 시스템의 완성도 부족 및 홍보 부족으로 인해 사용자 수는 매우 적다. 이는 악순환의 결과로 이어져 정보 업데이트가 느려지고, 관리 부족으로 이어져 시스템의 침체가 가속화되고 있다.

2) 기업체 시스템 현황

기업체의 시스템은 전국을 대상으로 구축되어 있다. 초창기 시스템은 단순 평점 및 사용자들의 리뷰 취합 형태로 이루어졌지만, 현재는 신뢰성을 높이기 위해 해당 업체의 객수 데이터를 시간대별로 나누어 제공하고 있다. 평점에 대한 필터링 기능을 가지고 있으며 일정 평점 이상의 가게만 모아서 볼 수 있는 기능이 있다.

B. 기존 문제점

현재 맛집 빅데이터 제공 시스템은 모든 사용자의 평점 및 평가를 취합하여 제공하는 형태로 구축되어 있다. 많은 정보를 통해 객관적인 평가가 이루어질 수 있지만, 광고 및 악의적 평점 테러와 같은 문제점이 존재한다.

1) 시스템 신뢰성

광고 및 악의적 평점 테러의 문제는 시스템의 신뢰성에 문제를 일으키고 있다. 특히 광고 글의 빈도는 지역주민을 주 고객층으로 상대하는 상권보다 타 지역 사람을 상대하는 관광지 주변 상권에서 높게 나타나며 이는 맛집 소개 시스템에 대한 불신으로 이어지고 있다.

2) 평점 필터링 시스템의 부재

앞서 언급되었던 평점 시스템의 완성도를 높이기 위해 광고와 평점 테러를 필터링하는 시스템이 필요하지만, 현재 제공되고 있는 시스템에는 해당 기능이 제대로 작동하지 않고 있다. 사용자 개인이 경험을 통해 직접 필터링하고 있으며, 이는 사용자의 이용 편의성에 악영향을 끼치고 있다.

3) 데이터 분류의 부족함

또다른 문제점은 '데이터 분류의 부족함'이다. 기존 시스템의 경우 평점, 영업시간을 중심으로 데이터 분류를 진행하고 있다. 하지만 음식점을 이용하는 사람들의 목적은 다양하다. 사용자가 음식점 사용 목적 및 성향에 맞는 가게를 찾는 것에 많은 시간이 소요되며, 스마트기기 사용에 익숙한 중장년층은 잘못된 정보의 위험에 더 많이 노출된다.

2. 대상 문제 및 요구조건 분석

A. 기존 맛집 데이터 제공 시스템

먹방 및 TV등 미디어의 발달로 사람들은 맛집에 대한 관심이 커졌고 그로 인한 맛집 검색 및 후기 등 맛집을 찾는 사람이 늘게 되었다. 그러나 현재 맛집 데이터 제공 시스템은 광고 혹은 대가성 리뷰 등, 소비자가 정확한 판단을 하기 힘들어졌다.

B. 제안할 맛집 데이터 제공 시스템

단순히 맛집 DB만을 보여주거나, 광고 또는 대가성 리뷰가 아닌 빅데이터 관점으로 접근하여 서비스를 제공한다. 빅데이터 기술을 통하여 신뢰할 수 있는 맛집 랭킹과 다양한 카테고리, 개인의 입맛을 고려한 리뷰 정보를 제공한다.

1) 카테고리 분류

기존 시스템의 평점, 영업시간 등 가게의 정보와 개인의 입맛 또는 연령대 및, 음식점을 찾는 목적, 분위기 등 다양한 카테고리를 제공한다. 해당 서비스를 통해서 사용자의 선택지를 넓힐 수 있다. 한식, 일식, 중식 등 판매하는 음식에 대한 카테고리, 혼밥 하기 좋은 곳, 데이트하기 좋은 곳 등 사용자 개인의 상황에 맞춘 카테고리 또는 메뉴의 맵기 같은 사용자의 입맛에 따라 선택할 수 있는 카테고리를 만들 수 있다.

2) 신뢰성

시스템의 신뢰를 위해 광고성 또는 대가성 리뷰를 반복 작성하는 사용자에게 대해 평점을 깎거나 이용을 제한하고, 다른 사용자에게 유용한 정보를 제공하거나 객관적인 리뷰를 작성하는 사용자에게는 높은 평점을 줌으로써 다른 사용자들이 시스템을 신뢰할 수 있게 한다. 이러한 방법을 통해 광고 및 악의적 평점테러 등 객관적이지 못한 정보를 사용자들이 필터링 할 수 있다.

3) 정보 최신화

가게의 영업 여부의 인한 문제점 또한 발생할 수 있다. 주말 및 공휴일 또는 주인의 개인 사정으로 인한 가게 휴무일일 경우 또는 가게의 폐업 등 운영하지 않는 가게를 빠르게 파악하여 가게의 상태를 업데이트 하는 최신화가 필요하다. 최신화를 통해 현재 가게가 영업을 하는지, 주말 또는 공휴일에 영업을 하는지, 가게가 폐업하지 않았는지 등 가게의 사정을 바로 확인할 수 있게 하여 사용자의 편의성을 증대한다.

3. 설계문서(설계)

A. 데이터 수집

초기 데이터: google Maps API의 데이터 활용

서비스 운용 중 데이터: 추적 코드를 활용한 데이터 수집, crawling

B. 개발환경

개발언어: PYTHON, HTML, Java Script

개발도구: PyCharm, Notepad ++, JavaScript IDE

실행화면: WINDOW(x64)

데이터베이스: MYSQL

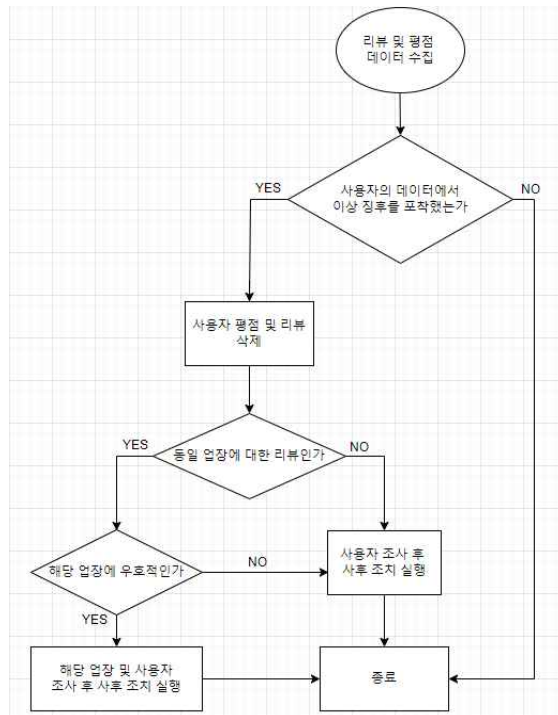
API: google Maps API

C. 프로세스 (신뢰성, 카테고리 분류, 최신화)

1) 카테고리 분류

사용자들의 리뷰 데이터 취합하여 키워드 분석을 진행한다. 시스템에서 구현하고자 하는 카테고리는 음식의 특징, 연령, 음식점 방문 목적이다. 시스템 구축 과정에서 키워드 분류를 진행하고, 해당 분류에 맞는 단어들을 리뷰에서 탐색한다. 음식점 검색 단계에서 카테고리 선택 단계를 추가하여 선별적인 맛집 정보를 탐색할 수 있다. 이를 통해 사용자의 맛집 탐색 소요시간 단축을 기대할 수 있다.

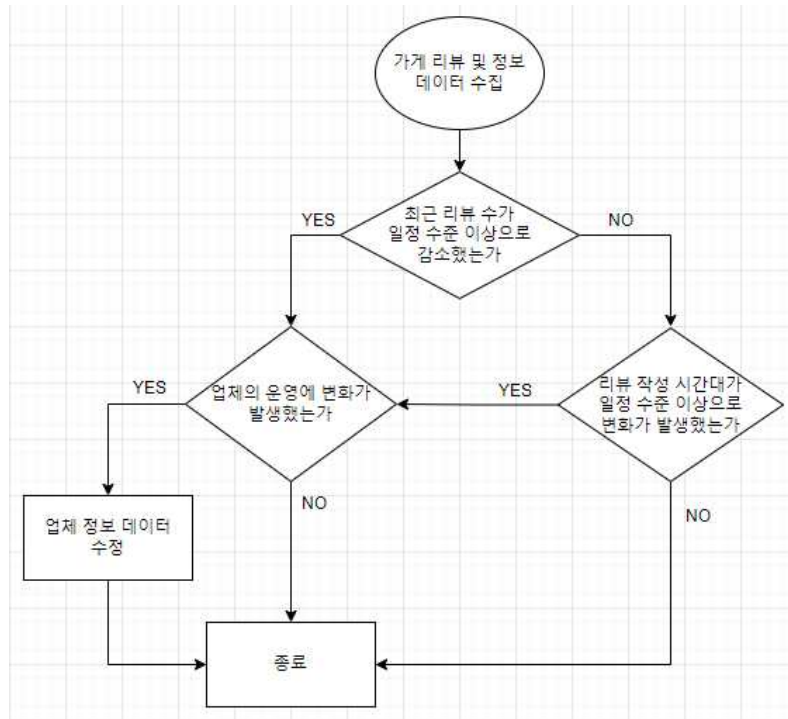
2) 신뢰성



사용자들의 리뷰 및 평점 데이터를 수집한다. 사용자의 데이터에서 이상 징후가 포착 되면 검사를 진행한다. 이때 이상 징후의 예시로는 '단기간에 과도한 리뷰 작성', '특정 업체의 갑작스러운 리뷰 증가', '유사 업체에 대한 평점 테러'가 있다. 이상 징후가 포착되면 우선적으로 해당 사용자의 평점과 리뷰를 일시적으로 삭제한다. 이후 동일 업체에 대한 리뷰인지 파악한다. 동일 업체에 대한 리뷰일 때, 해당 업체에 우호적이라면 업체와 사용자를 동시 조사 후 사후 조치를 실행한다. 우호적이지 않다면 평점 테러에 무게를 두고 사용자를 조사를 실시한다. 동일 업체에 대한 리뷰가 아니고, 여러 업체에 대해 낮은 평점과 악평을 작성한다면 사용자에게 대한 조사를 실시한다. 문제가 없다면 이전에 삭제했던 사용자의 평점과 리뷰를 복구하고, 문제가 있다면 사후 조치를 진행한다. 해당 작업을 통해 서비스의 신뢰성을 높일 수 있다.

3) 최신화

가게 리뷰 및 정보 데이터를 수집한다. 최근 리뷰 수가 일정 수준 이상으로 감소했는지 여부와 리뷰 속 사용자들의 방문 시간대가 일정 수준 이상의 변화가 감지된다면 업체의 운영 변화 여부를 확인한다. 운영 변화에는 휴업, 폐업 등의 영업 여부와 영업 시간의 변화가 있을 수 있다. 업체 운영에 변화가 발생했다면 변경된 사항을 페이지에서 수정한 뒤 해당 작업을 종료한다. 이를 통해 기존 시스템에서 문제가 있었던 정보 업데이트를 조금 더 빠르게 진행하여 사용자에게 조금 더 완성도 높은 서비스를 제공할 수 있다.



4. 현실적 제약 사항 및 대책(제약 사항)

A. 데이터 수집 제약 사항

1) 초기 데이터의 신뢰성 문제

Google Maps API의 초기 데이터의 광고 및 대가성 리뷰를 모두 확인 불가능하므로 신뢰성에 문제가 없다고 가정한다.

2) 데이터의 방대함에 대한 문제

전국 맛집 데이터의 범위가 방대하므로 불필요한 데이터들을 최대한 제거한 뒤 초기 데이터로 사용한다.

B. 소프트웨어 제약 사항

1) 최신화에 대한 문제

업체 각각의 기존 고객 수 및 유동인구를 고려하여 최신화에 대한 다른 기준을 제시해야 하지만, 해당 기술을 구현하기에 많은 시간이 걸릴 것으로 예상되므로 업체의 마지막 리뷰가 작성된 이후 한달이 지났을 때 가게의 운영에 변화가 생겼다 판단하고 데이터를 수정한다.

5. 개발일정 및 역할 분담 (개발 일정)

A. 일정

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							
						데이터 분석 청년인재양성 프로그램 수강 웹 서비스 및 데이터 베이스 구축											
													최종 보고서 작성 및 심사 준비				
														발표 준비 및 마무리 작업			

B. 역할분담

이름	작업 목록
최경섭	카테고리 분류 알고리즘 설계 및 구현 신뢰성 알고리즘 설계 및 구현 데이터 베이스 구축
장희승	Sample data 수집 최신화 알고리즘 설계 및 구현 API 데이터 필터링 및 적용
공통	웹 서비스 구축 보고서 작성 발표 준비 서비스 테스트 고도화 작업