## **2월 2주차 회의록 (21.02.13)**

4조 김태희, 안세훈, 정영훈, 고준호

**도메인 변경 크롤링 테스트**

**크롤링 툴 재구성**  
- 바뀐 도메인으로 크롤링 구조 변경 필요  
- 크롤링 툴 재구성 후 지역별 크롤링 테스트  
- 네이버 리뷰 정책상 별점이 없어짐에 따라 별점 크롤링 제거

**테스트 결과**  
- 매장의 수와 리뷰의 수가 급격히 많아짐에 따라 시간이 많이 걸림  
- 최근 리뷰 데이터를 많이 확보할 수 있음  
- 현재까지 긴 리뷰(펼치기가 필요한 리뷰)는 크롤링 불가  
- 데이터를 많이 확보하기 위해 크롤링 시간 개선 필요

**모델링 개선**

- 속성값 변경  
커피 맛/ 음료 / 디저트 / 뷰 / 주차/ 메뉴 / 인테리어 / 사진/ 청결/ 친절 / 좌석 / 화장실 / 집중 / 가성비 / 대화   
- 별점 예측 기능 추가  
- 긍부정 분석 재테스트

**2월 계획**  
- 변경된 도메인으로 크롤링 툴 완성 & 크롤링 (완료)  
- 기존 모델에 적용 (3주차 완료 예정)  
- 데이터 태깅 (4주차 시작 예정)  
- 파인튜닝 (개강 후 적용 예정)  
- 개선 모델 적용 (개강 후 적용 예정)

**2월 2주차 진행사항**

**[도메인 변경에 따른 크롤링 재구성]**

텍스트이(가) 표시된 사진

자동 생성된 설명

리뷰에서 더 이상 별점을 지원하지 않음 -> DB에서 별점을 뺄 예정  
Django 구조 수정 필요

카페와 식당 도메인의 경우 페이지 구조가 같아서 크롤링 하나로 두 도메인 동시에 크롤링 가능  
스터디카페에 비해 리뷰 수가 훨씬 많아 학습시키기에 유리함  
(리뷰가 적은 곳이여도 기본 수십개, 많을 경우 1000개 이상)

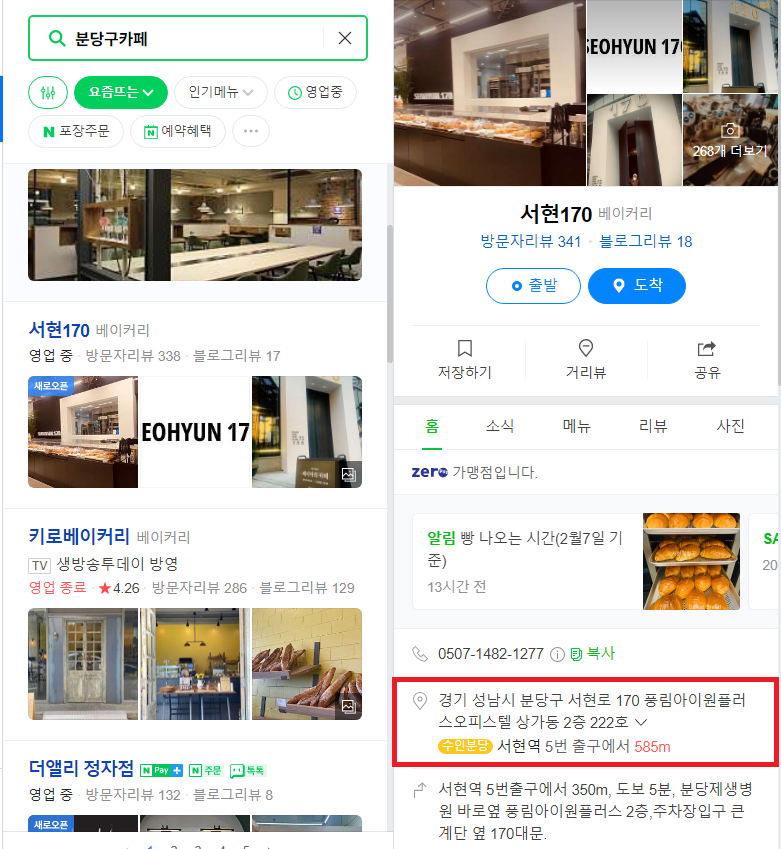
텍스트이(가) 표시된 사진

자동 생성된 설명

리뷰 엑셀파일 크롤링 결과 분당구 카페 기준 카페 12곳 리뷰 5천개 이상

텍스트이(가) 표시된 사진

자동 생성된 설명



스터디카페의 경우 첫번째 프레임에서 지역이 나와있으나  
식당 & 카페의 경우는 상세 정보에 들어가야만 장소가 나와있음

-> 크롤링 코드 수정 약간 필요

스터디카페의 경우에는 자세한 위치 외에도 지역 정보만 크롤링이 가능했으나  
카페&식당은 자세한 위치만 크롤링 가능

텍스트이(가) 표시된 사진

자동 생성된 설명

카페&식당 도메인의 경우 네이버에서 어떤 점이 좋은지 속성을 제공하므로  
리뷰 분석 후 비교 가능

크롤링 수정해야할 사항  
장문 리뷰의 경우 리뷰 펼치기가 눌러지지 않아 크롤링하지 못했으므로 다른 해결방안 필요

도메인이 변경되고 리뷰가 많아진 만큼 속도 개선 추가 필요  
  
  
Django 수정해야할 사항  
크롤링을 할 지역을 넓혀야하므로 구 별로 동적 URL 생성 필요  
웹 CORS 에러 해결 필요  
CORS 에러 해결 후 서버 배포