**빅데이터프로그래밍 프로젝트 계획서**

60193011 김도연, 60201680 박하나,  
60201689 양채연, 60201697 이지은

**문제 정의**

사람들마다 카페를 가는 목적이 다르지만 기존의 지도 어플에서는 모든 카페가 지역에 따라 통합되어 검색된다. 따라서 사용자가 직접 각각의 카페 리뷰를 찾아보며 자신의 목적에 맞는 카페를 검색해야 한다.

이러한 점을 고려하여 본 프로젝트에서는 지역별로 분류된 카페를 목적 별로 분류하여 사용자가 원하는 장소를 편리하게 찾을 수 있도록 한다.

- 예상 동네 분류 (연희동, 망원동, 서교동, 목동, 남가좌동 …)

- 예상 대표 키워드 (혼자, 분위기, 콘센트, 공부, 디저트, 케이크, 맛, 원두, ...)

- 예상 카페 분류 (카공하기 좋은 카페, 디저트 카페, 커피가 맛있는 카페, 분위기 좋은 카페, 혼자가기 좋은 카페, 대형 카페, 애견 동반 카페, 뷰가 좋은 카페, 친환경 카페 …)

**실행 계획**

**데이터 수집 방법:** 네이버 지도에서 지역별 카페(ex. 연희동 카페) 목록을 가져온다. 네이버 블로그 크롤링으로 각 카페에 대한 상세 리뷰 데이터(ex. 공부하기 좋아요, 디저트가 맛있어요~)를 수집한다.

**데이터 분석 방법:** 수집한 카페별 블로그 데이터에서 KoNLpy를 이용해 불용어를 제거, TF-IDF를 통해 중요하지 않은 키워드 조회한다. Mapreduce ,Spark, Pig 등 수업시간에 배운 방식을 활용해 빈도수가 높은 키워드를 워드카운팅 후 빈도수가 높은 키워드로 카페를 분류하여 해당하는 카페가 어떤 분류에 해당하는지를 결과로 도출한다.