

2017년 봄학기 TIM61101 빅데이터와 신제품 개발

Progress Presentation

걸러서 봐

목차

1. 프로젝트 주제 변경
2. 프로젝트 진행 상황

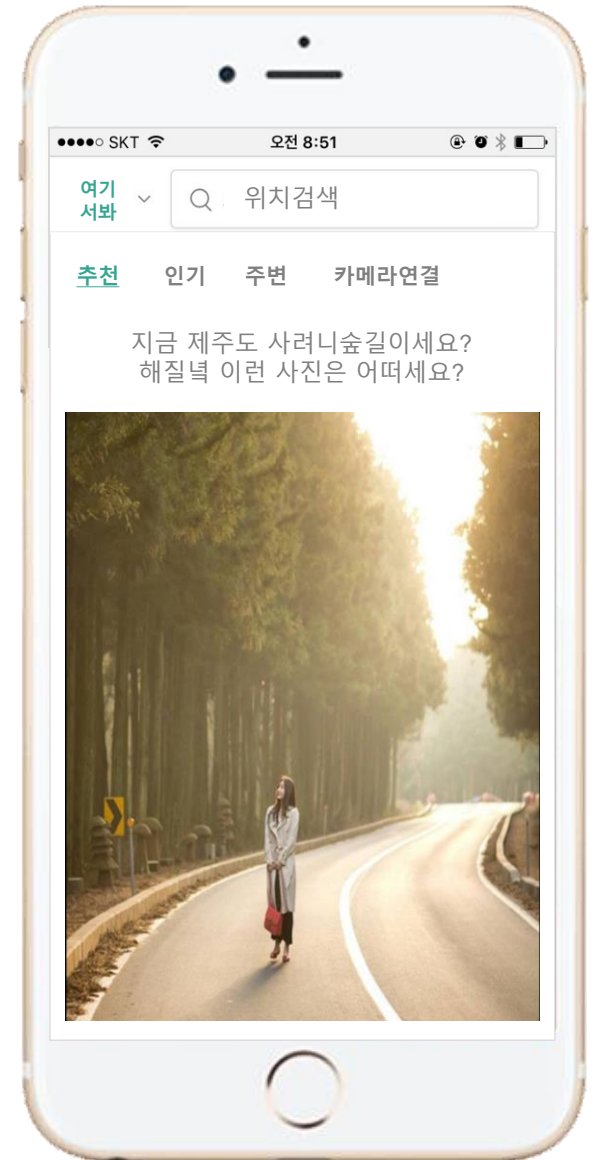
프로젝트 주제 변경

여기 '서' 봐

- 나도 몰랐던 나의 사진 취향 분석
- 여행지에서 나의 취향에 맞는 사진 추천
- 내 맘에 쏙 드는 사진을 찍도록 도와주는 서비스

❖ 주제 진행의 한계점

- 구도를 결정할 수 있는 요소
- 사진 촬영 장소(spot)
- Instagram data



프로젝트 주제 변경

‘걸러서’ 봐



- 인스타그램 내 검색에서 광고를 필터링
 - 사용자가 원하는 정보만을 얻을 수 있도록 도움
 - 추가적인 서비스 확장이 가능
-
- 작성자가 사용하는 텍스트 기반 필터링
 - 특정 키워드 중심으로 범위 축소
 - 위치 정보 등을 이용한 추가 정보 제공



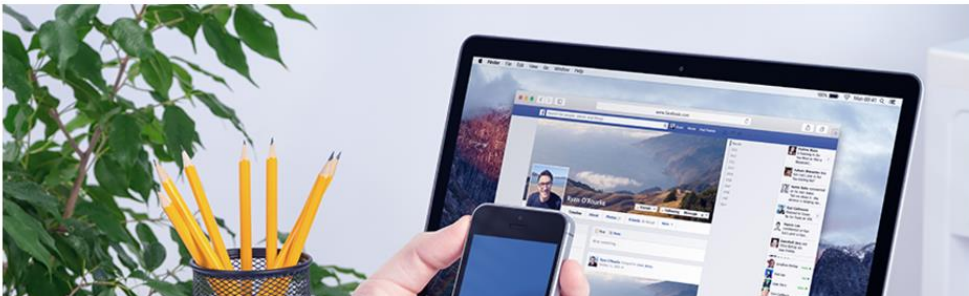
걸러서 봐

Problem Statement

페이스북, 돈은 벌지만 불안 불안

호실적에도 광고 의존도 심화로 주가 하락...TV 콘텐츠 제작 등 주목

승인 2017.05.08 11:28:03(월) | 고재석 기자 jayko@sisajournal-e.com



“사용자들은 광고에 지쳤을 것”

“Instagram 역시 광고 천국 될 것”



'광고포화' 페이스북 대신 인스타로

더피알 (회원가입) - 2016. 11. 21.

[더피알=이윤주 기자] **인스타그램**이 새로운 **광고** 플랫폼으로 빠르게 부상하고 있다. 페이스북이 사실상 '**광고포화**' 상태에 이른 만큼 대체제로 각광 ...



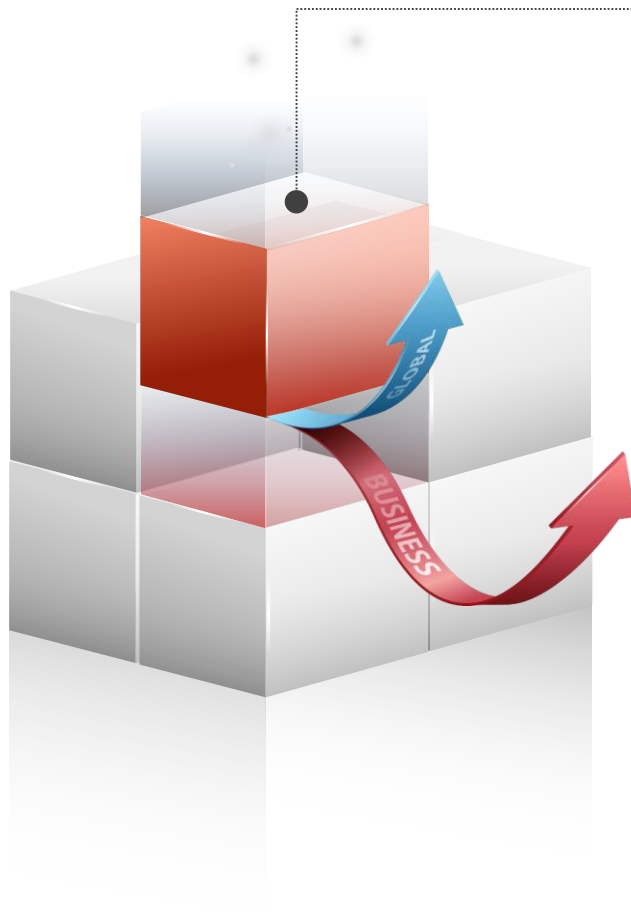
인스타그램 팔로워 늘리기가 썩쓸한 이유

플래툼 - 2017. 3. 16.

사진에 감성적인 필터를 입혀 정사각형의 이미지로 나만의 전시공간을 만들 수 있는 인스타그램은 블로그나 페이스북의 **광고**에 지친 젊은 층들에게 ...

걸러서 봐

Gap Analysis



걸러서 봐

인스타그램 텍스트 기반 광고 필터링

INSTAGRAM

- 무분별한 광고 유입 방지
- 양질의 콘텐츠 보유
- 사용자 이탈 방지 효과

END USER

- SNS의 불편함을 개선
- 원하는 정보만을 선별적 확보

걸러서 봐

Flow Chart

❖ 자료수집

- Python
- Raw data 수집

❖ 전처리

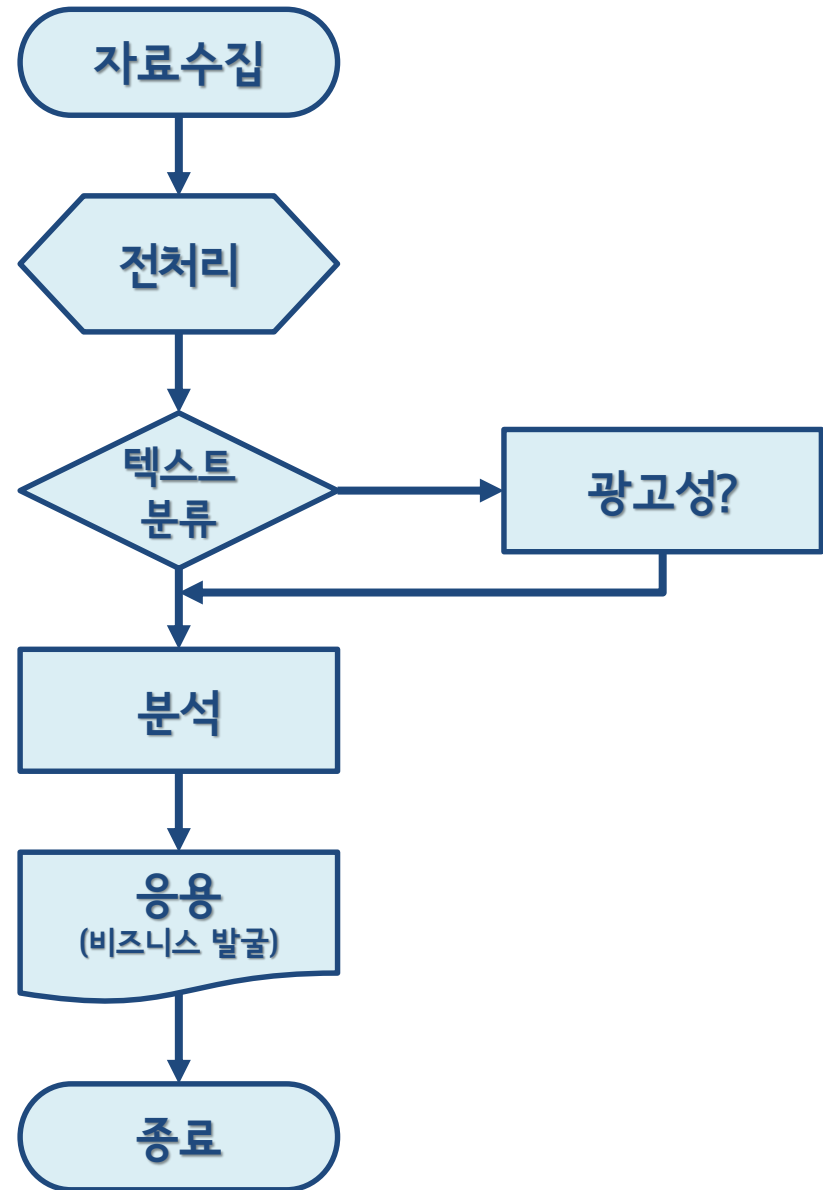
- R
- 한글 제외 모두 삭제 (특수문자, 영어, 이모티콘 등)

❖ 텍스트 분류

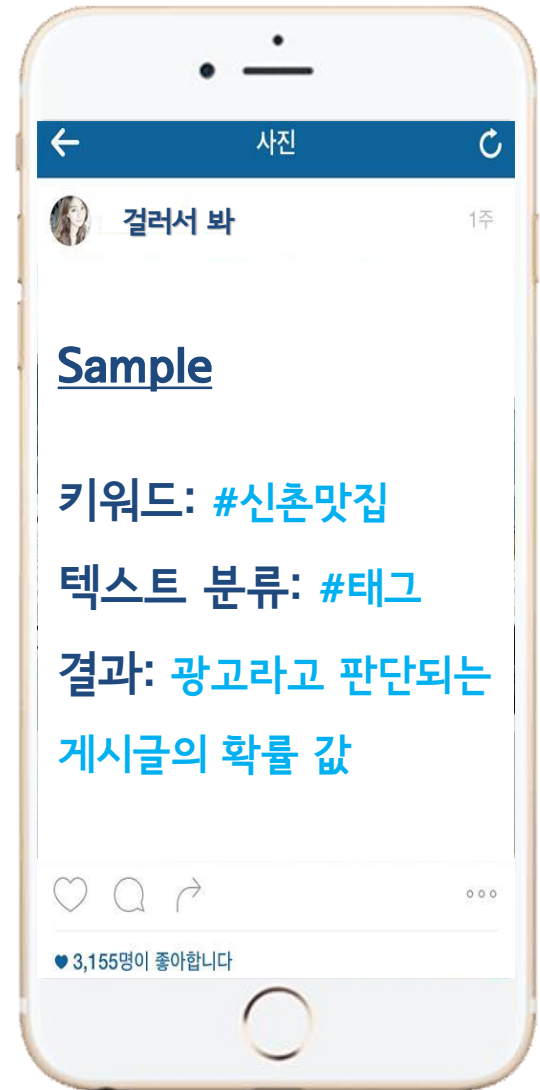
- 광고성 텍스트 분류
- 태그 / 게시글 / 프로필 정보

❖ 분석

- Bayesian Spam Filtering
- Multiple-word Feature
- Markovian Filtering



현재 진행 상황

[illegible]

걸러서 봐

향후 진행 방향

- 광고라고 분류되어지는 텍스트를 어떻게 학습시킬 것인가?
- 어떤 알고리즘을 가지고 분석을 해야 할지?
- 텍스트 정보 이외의 것들을 어떻게 활용할지?

감사합니다