

제 3회 KUIAI 해커톤

데이터 기반 패션 인플루언서 추천 알고리즘

안암어스(Anam-Us)

산업경영공학부 2020170815 김경호

산업경영공학부 2020170823 윤화평

산업경영공학부 2020170812 임정섭

산업경영공학부 2020170810 최대원



Contents

001 문제 접근

- 데이터 관찰
- 시스템 모델링

002 알고리즘 설계

- 데이터 전처리
- 추천 데이터

003 알고리즘 구현

- 파이프라인 구조
- 전처리 단계
- 추천 단계
- 알고리즘 결과

004 플랫폼 서비스 기획

- 알고리즘 분석
- 기획안

001

문제 접근



데이터 관찰 – Input(상품)

상품(무신사)

상품1

텍스트
(상품 설명)

상품n

텍스트
(상품 설명)

상품(오픈 데이터셋)

상품1

텍스트
(이미지
메타정보)

데이터 관찰 – Input(소비자)

소비자

소비자1

피드1-1



텍스트
(게시글 본문,
해시태그)

피드1-30



텍스트
(게시글 본문,
해시태그)

데이터 관찰 – Output(인플루언서)

인플루언서

인플루언서1

피드1-1



텍스트
(게시글 본문,
해시태그)

피드1-30



텍스트
(게시글 본문,
해시태그)

인플루언서100

피드100-1



텍스트
(게시글 본문,
해시태그)

피드100-30



텍스트
(게시글 본문,
해시태그)

도메인

객체1

그룹1-1



텍스트

그룹1-n



텍스트

객체m

그룹m-1

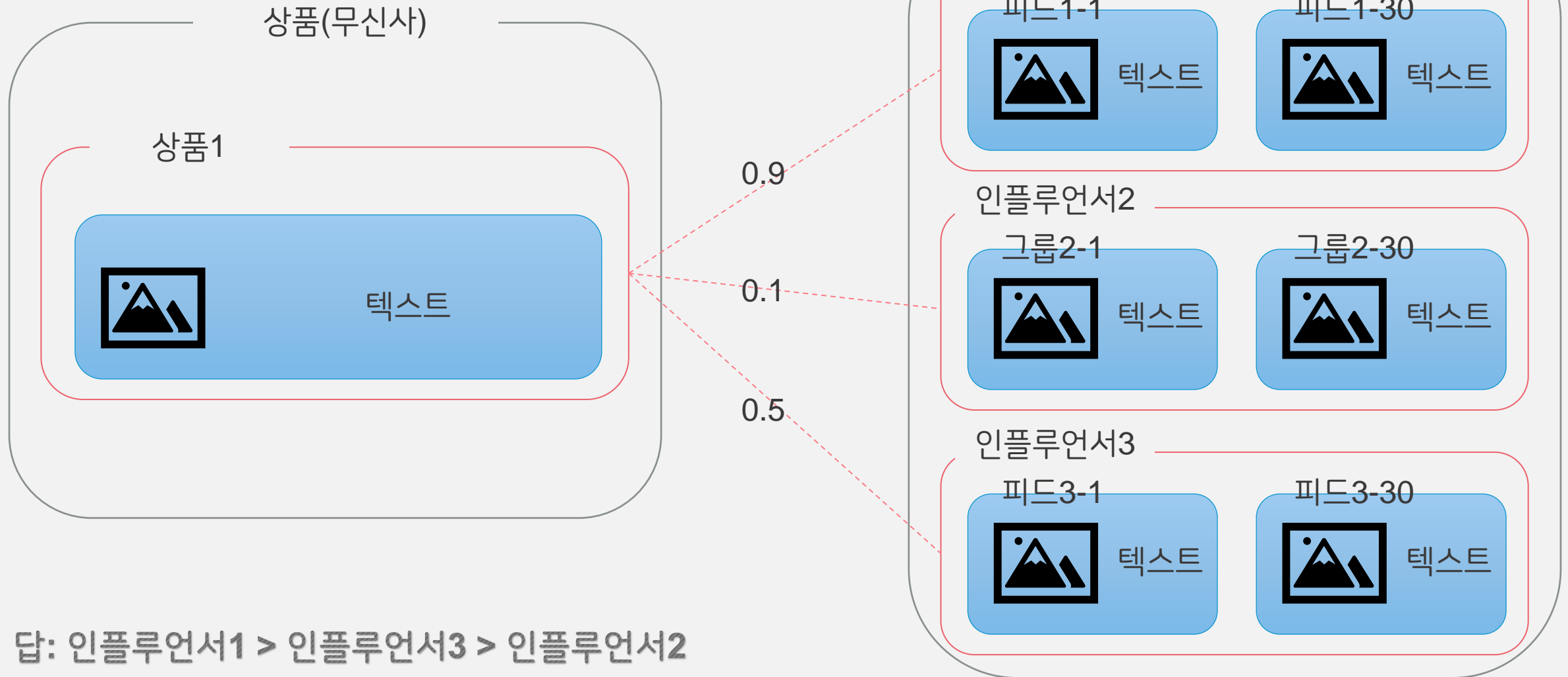


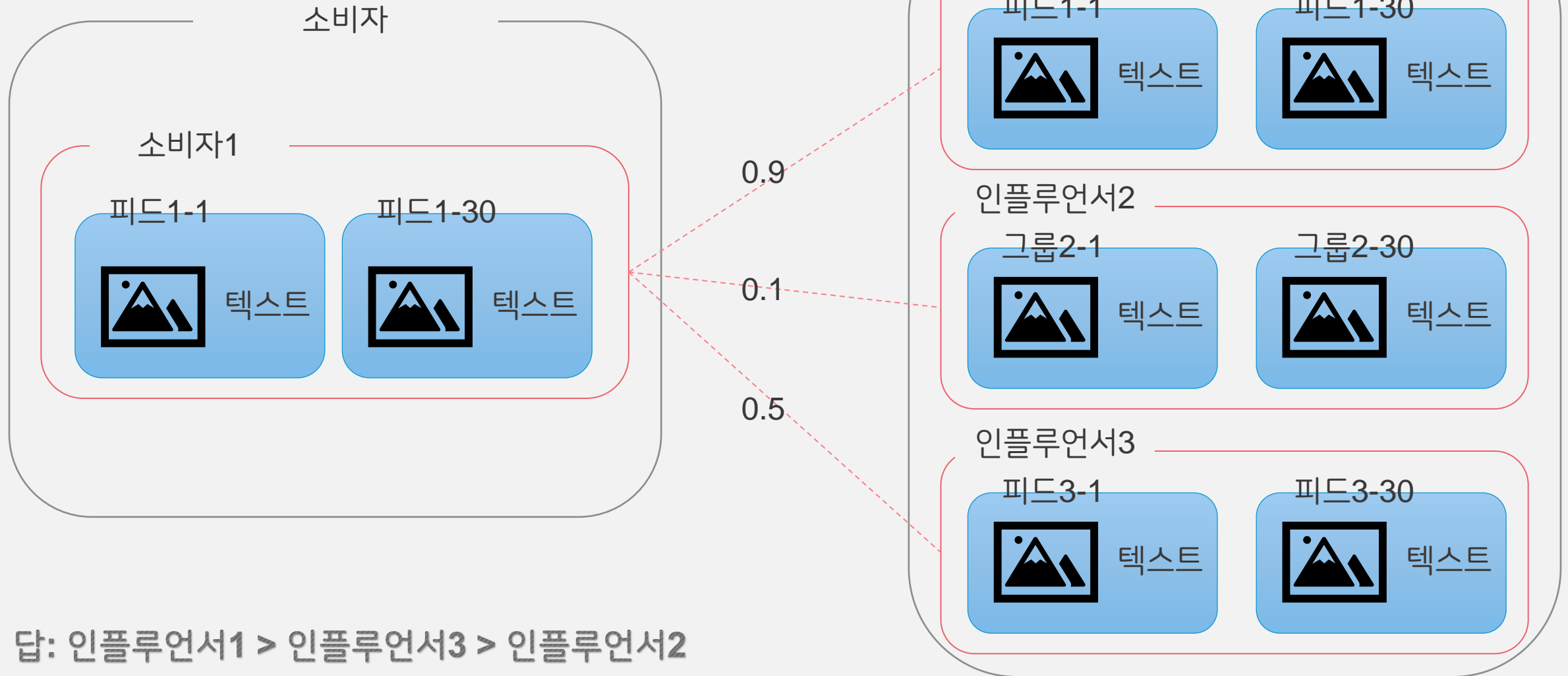
텍스트

그룹m-n



텍스트





002

알고리즘 설계



상품(무신사): $m=1$
소비자: $m=30$

상품(무신사) / 소비자

객체1

그룹1-1

텍스트'1-1

그룹1-m

텍스트'1-m

게시물(이미지+텍스트)를 글로 가공하여
두 객체 간 유사도 계산!

인플루언서

인플루언서1

피드1-1

텍스트'1-1

피드1-30

텍스트'1-30

인플루언서2

그룹2-1

텍스트'2-1

그룹2-30

텍스트'2-30

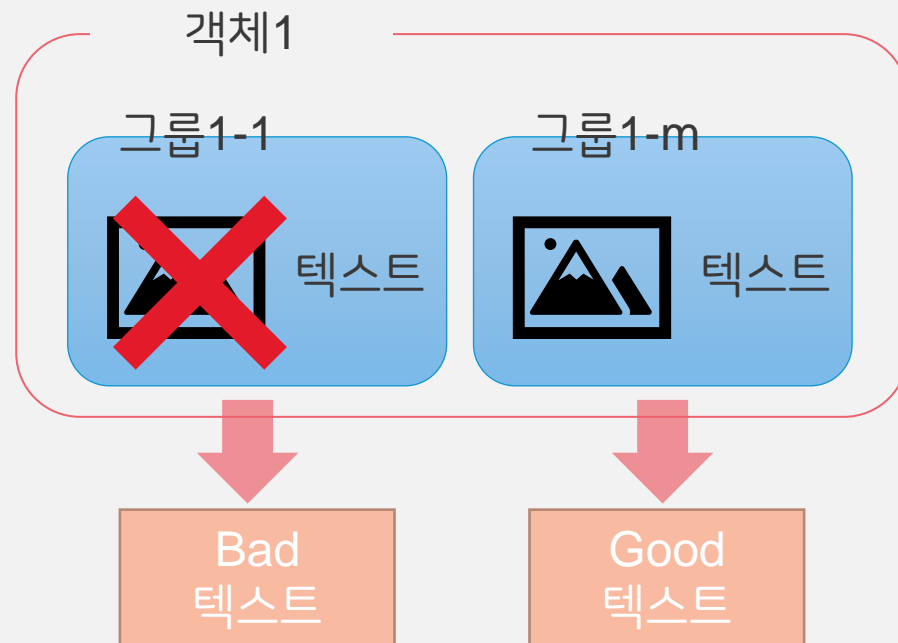
인플루언서3

피드3-1

텍스트'3-1

피드3-30

텍스트'3-30



글만으로는 부족하여 이미지도 텍스트 구할 때 활용!

데이터 전처리 – 이미지 활용법



새, 날다

1. 이미지를 설명하는 텍스트를 생성

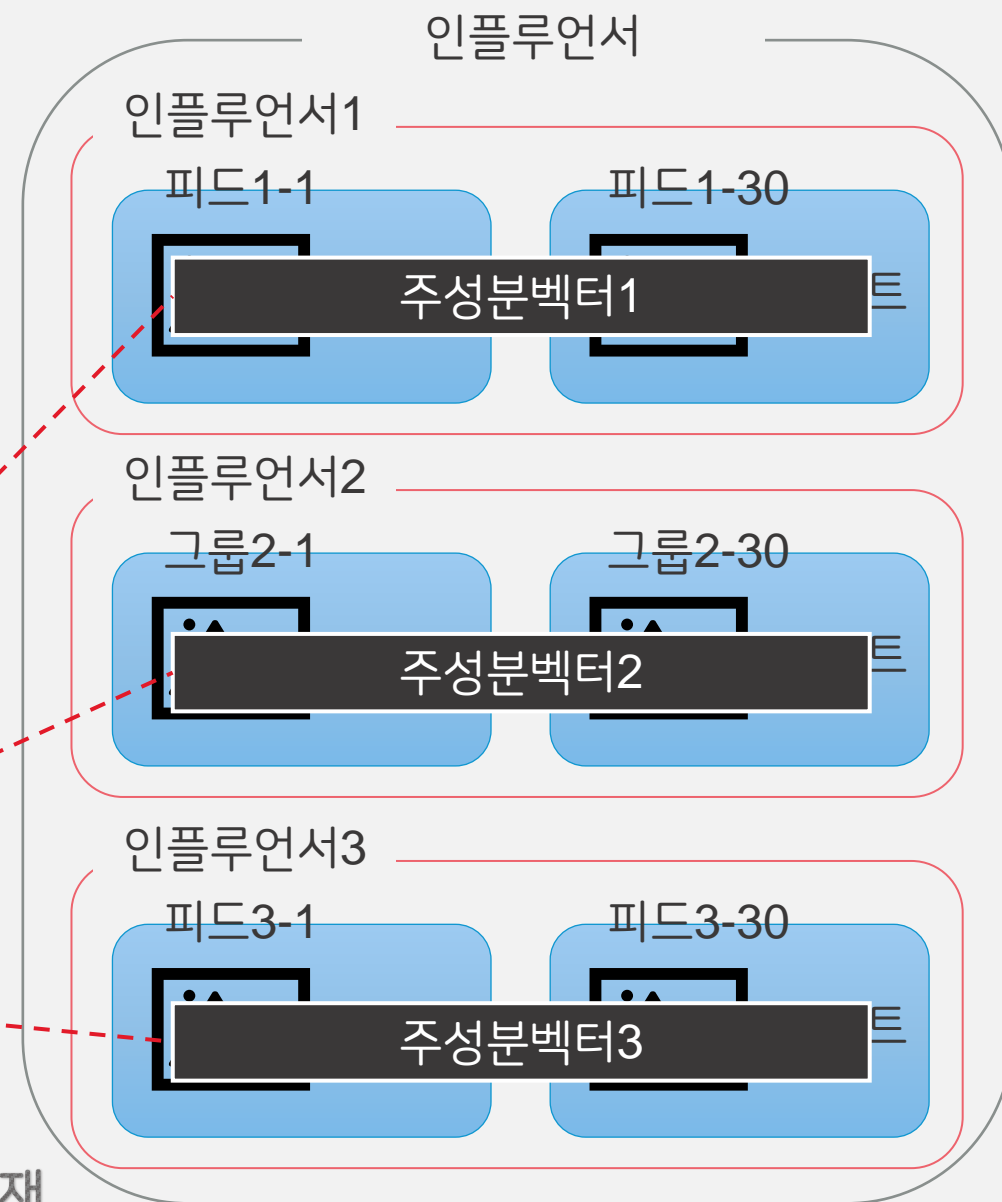
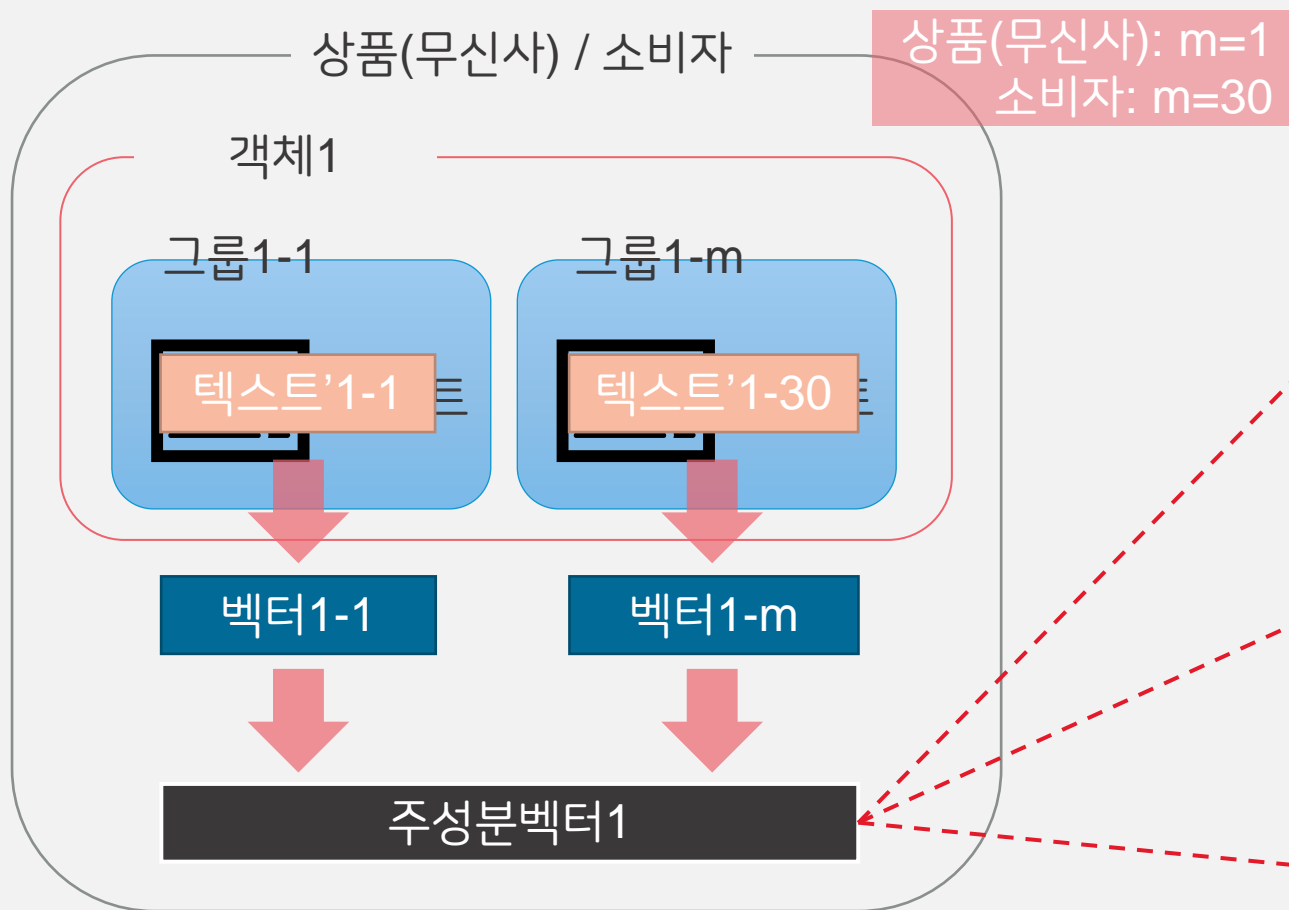
날개, 하늘, **지갑**



날개, 하늘

2. 주어진 텍스트에서 이미지와 관련 있는 것만 추출

추천 데이터 - 텍스트 > 벡터 > 주성분



게시물마다 추출한 텍스트는 소비자/인플루언서마다 여러 개 존재
→ 벡터로 변환 후, 주성분 구하여 객체끼리 비교

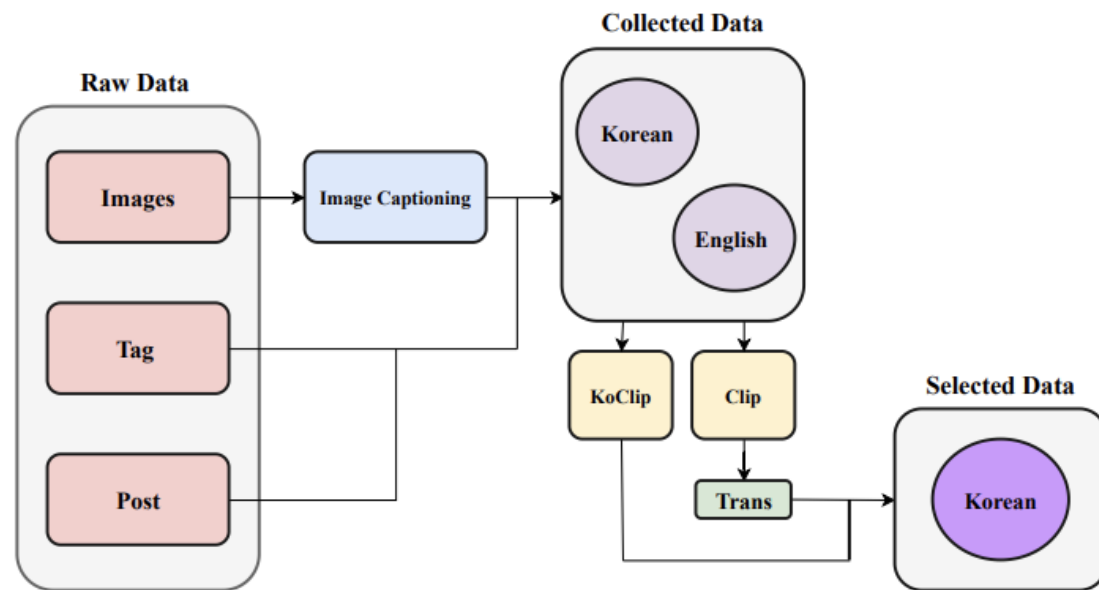
003

알고리즘 구현

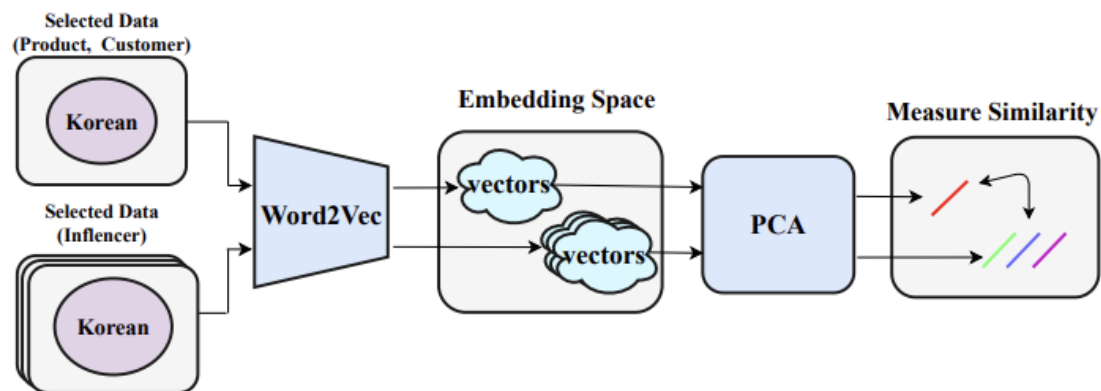


파이프라인 구조

1. Preprocessing Phase



2. Recommending Phase

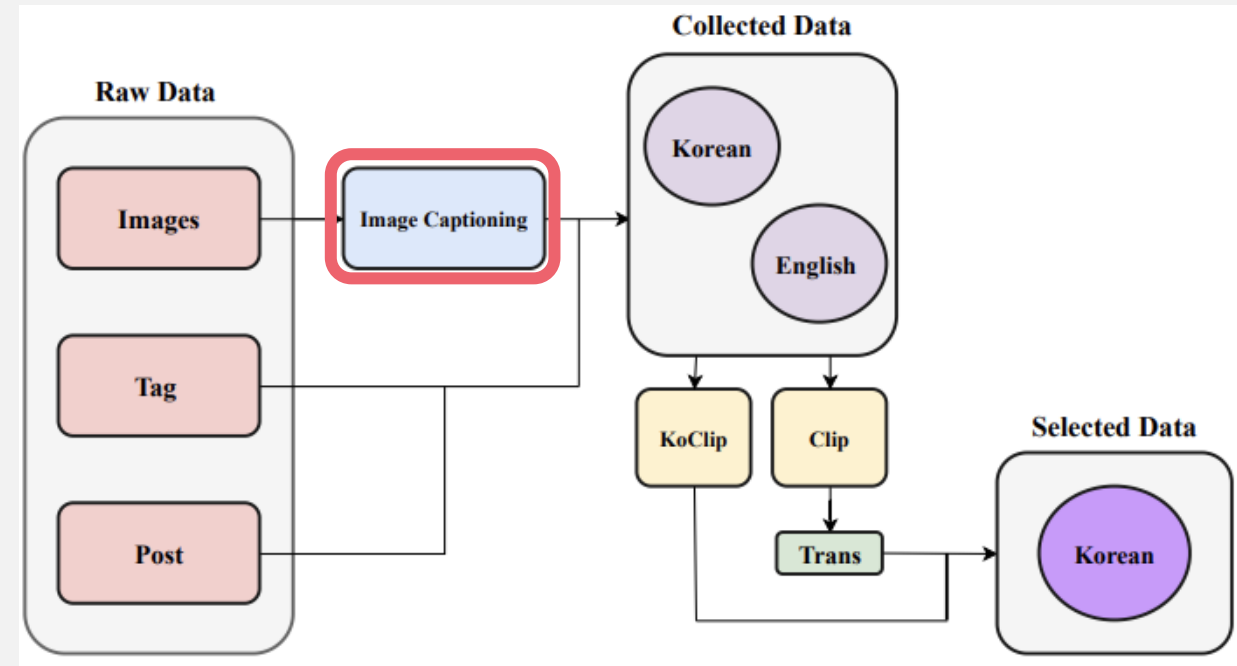


전처리 단계 – Image Captioning

- 이미지를 input으로 넣으면, 이미지에 대한 description을 output으로 추출
- NIC(Neural Image Caption)모델을 사용
- Output으로 영어 description이 도출된다.

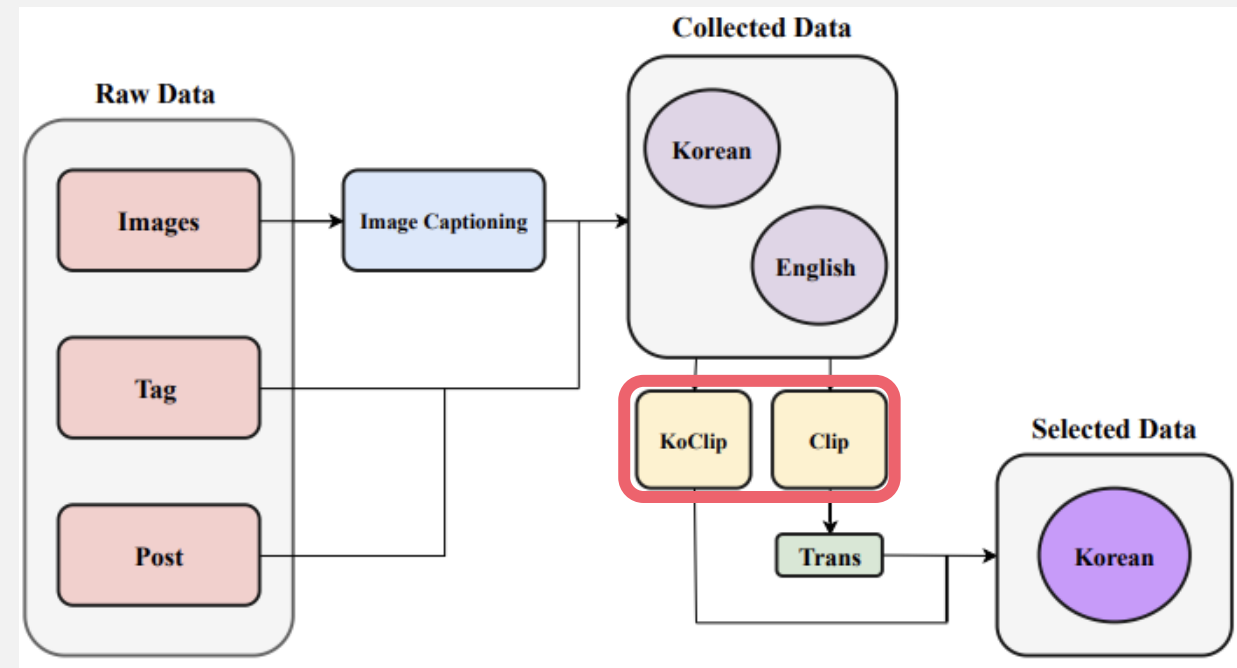


a girl in a red shirt
and a white shirt is
holding a cup .



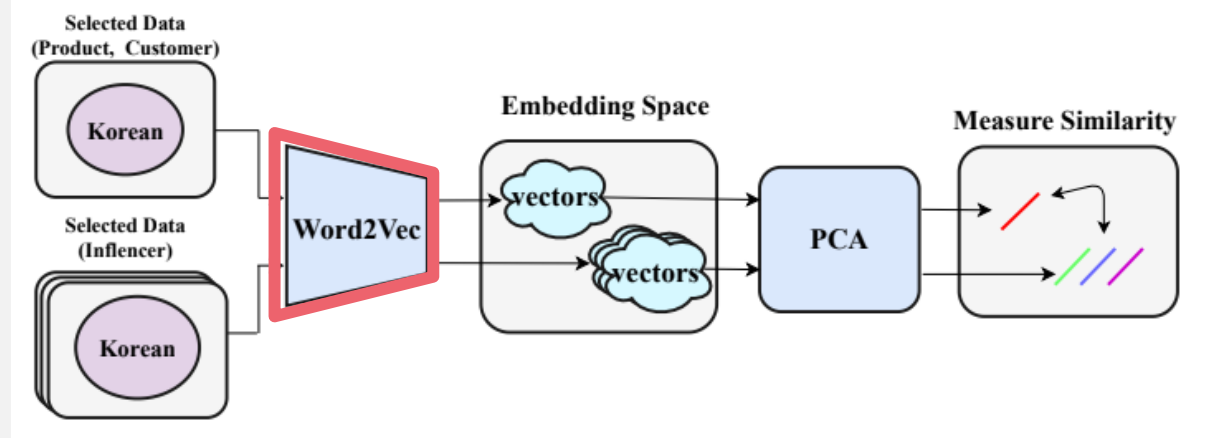
전처리 단계 – CLIP / KoCLIP

- 이미지와 텍스트 사이의 연관성을 측정
- CLIP은 영문 텍스트를, KoCLIP은 한글 텍스트를 입력 받음
- 사진에 대한 텍스트(주어진 것 + Image Captioning으로 생성한 것) 중 사진과 연관 있는 것만 남기는 역할
- Clip의 영문 출력 텍스트는 구글 번역기 통해 한국어로 번역하여 최종 전처리된 selected data는 한글 텍스트로 수집



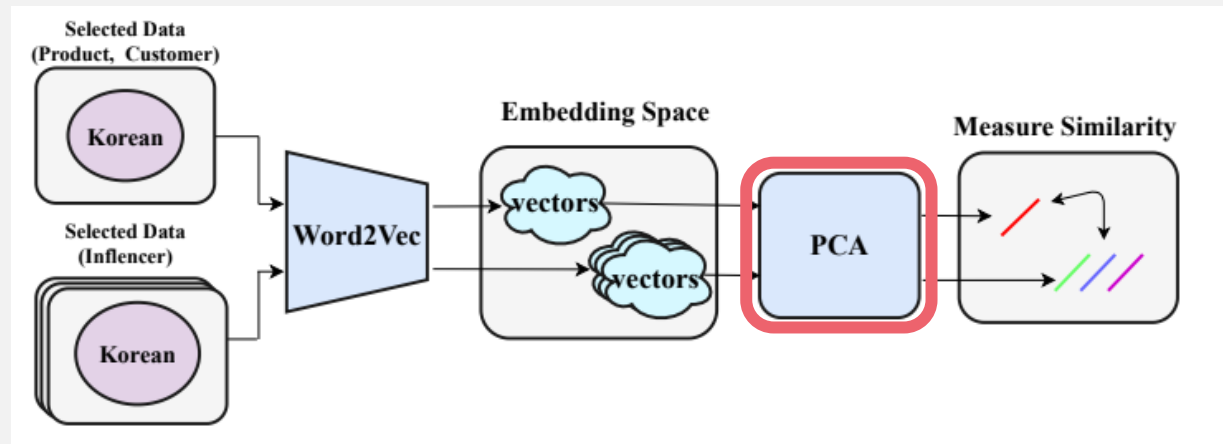
추천 단계 – Word2Vec

- 자연어 처리 방법
- 단어(=토큰)들 사이의 연관성을 고려하여 단어를 적절한 임베딩 공간으로 사영한다.
- 충분히 큰 크기의 한글 Coupus에서 Pretrain 된 Word2Vec 모델을 load한 뒤, 가지고 있는 데이터(상품, 소비자, 인플루언서)를 모두 이용해 이를 FineTuning한다.



추천 단계 – PCA

- 차원축소방법
- 임베딩 공간에 사영된 여러 벡터를 대표하는 featur를 도출
- 주어진 데이터의 분산을 최대한 보존할 수 있는 Component를 계산할 수 있다.
- 집합(상품/소비자/인플루언서) 내 한 객체를 대표하는 feature를 계산
- 서로 다른 집합에 속하는 두 객체 feature 간 cosine similarity를 구한다.



알고리즘 결과 (상품 – 인플루언서)

유사도

	product	@98.c_project	@ys971003__	@malko_bee	@h
0	29960	0.9343547173	0.9226963167	0.917094051	0.945
1	29961	0.9540356246	0.9236141863	0.9170644381	0.921
2	29202	0.874121743	0.8953900435	0.8843493607	0.840
3	29975	0.9156892151	0.8912789218	0.878399951	0.912
4	29977	0.9417293993	0.9460888647	0.9232456713	0.91
5	29978	0.9470140402	0.9618454901	0.9389653085	0.893
6	29980	0.9750101596	0.9618187626	0.9707128413	0.933
7	29981	0.9624632632	0.9254553092	0.9251745776	0.927
8	29469	0.9509209004	0.9433270332	0.9451713992	0.959
9	28964	0.970763627	0.9493069625	0.9380048051	0.930
10	29349	0.9368777151	0.9226099998	0.9260947129	0.943
11	29222	0.9386905747	0.928290987	0.9599004165	0.968
12	29991	0.9334580259	0.9139775272	0.9270638914	0.953
13	29992	0.9162368262	0.887548164	0.9034474768	0.935
14	29996	-0.8969506954	-0.8748139594	-0.8692925163	-0.924
15	28973	0.9419349575	0.9280224157	0.9430215976	0.962
16	30000	0.9503121302	0.9326863358	0.9126643147	0.916

...

⋮

최종 추천 결과

	product	1st	2nd	3rd
0	29960	@by_he.nique	@ap.s_fi1st	@__v.yuum_look__
1	29961	@lamode.seoul	@c_eunnnnnnn	@y_mood_h
2	29202	@yj_mark	@kimyannnnngh	@bloggerbok
3	29975	@ap.s_fi1st	@lamode.seoul	@by_he.nique
4	29977	@loolinmx	@yj_mark	@skuukzky
5	29978	@bloggerbok	@chaileeson	@skuukzky
6	29980	@so_j2	@sympa_young	@one_r_k
7	29981	@lamode.seoul	@y_mood_h	@337janggoon
8	29469	@__v.yuum_look__	@jung_staas	@0nefence
9	28964	@lamode.seoul	@c_eunnnnnnn	@y_mood_h
10	29349	@__v.yuum_look__	@so_love_so_	@ap.s_fi1st
11	29222	@so_h_appy	@jung_staas	@hotneul
12	29991	@jung_staas	@so_h_appy	@so_j2
13	29992	@so_h_appy	@jung_staas	@hotneul
14	29996	@jindalorian	@zxcvr0626	@jin_pyo_is
15	28973	@so_h_appy	@jung_staas	@hotneul
16	30000	@jin.wonder	@lamode.seoul	@__v.yuum_look__

⋮

알고리즘 결과 (소비자 – 인플루언서)

유사도

	customer_name	@98.c_project	@ys971003__	@malko_bee	@hotneul	@jung_staas	@oneh6_	@dosirak_hansol	@j0ng_v	
0	@kxungho00	0.9781163457	0.9826476826	0.9675105267	0.9268311325	0.9814590881	0.9563510578	0.9774128289	0.95259	
1	@agree_seop	-0.6151088777	-0.6099531288	-0.5924101704	-0.4785684632	-0.5460134198	-0.6542346514	-0.581567022	-0.656036	...
2	@dlwlrma	0.9308180811	0.9207614896	0.8938477379	0.9005692303	0.9469854134	0.8806722904	0.9177863736	0.864003	
3	@j_g_ok	-0.8246188025	-0.872770212	-0.8766349758	-0.7286292524	-0.8174754024	-0.8531032648	-0.8794519783	-0.8959	

최종 추천 결과

	customer_name	1st	2nd	3rd
0	@kxungho00	@0nefence	@y_.dulcet	@so_love_so_
1	@agree_seop	@jin_pyo_is	@jindalorian	@dismas_
2	@dlwlrma	@jin.wonder	@_v.yuum_look__	@ap.s_f1st
3	@j_g_ok	@jin_pyo_is	@jindalorian	@zxcvr0626

추천 결과 예시 (상품 – 인플루언서)

<상품 기준 추천 인플루언서 결과 예시>

product(29202)

MUSINSA



1st(@yj_mark)



2nd(@kimyannnnngh)



3rd(@bloggerbok)



product(29992)

MUSINSA



1st(@so_happy)



2nd(@jung_staas)



3rd(@hotneul)



...

추천 결과 예시 (소비자 – 인플루언서)

< 소비자 기준 추천 인플루언서 결과 예시 >

customer(@agree_seop)

1st(@jin_pyo_is)

2nd(@jjindalorian)

3rd(@dismas_)



customer(@kxungho00)

1st(@0nefence)

2nd(@y._dulcet)

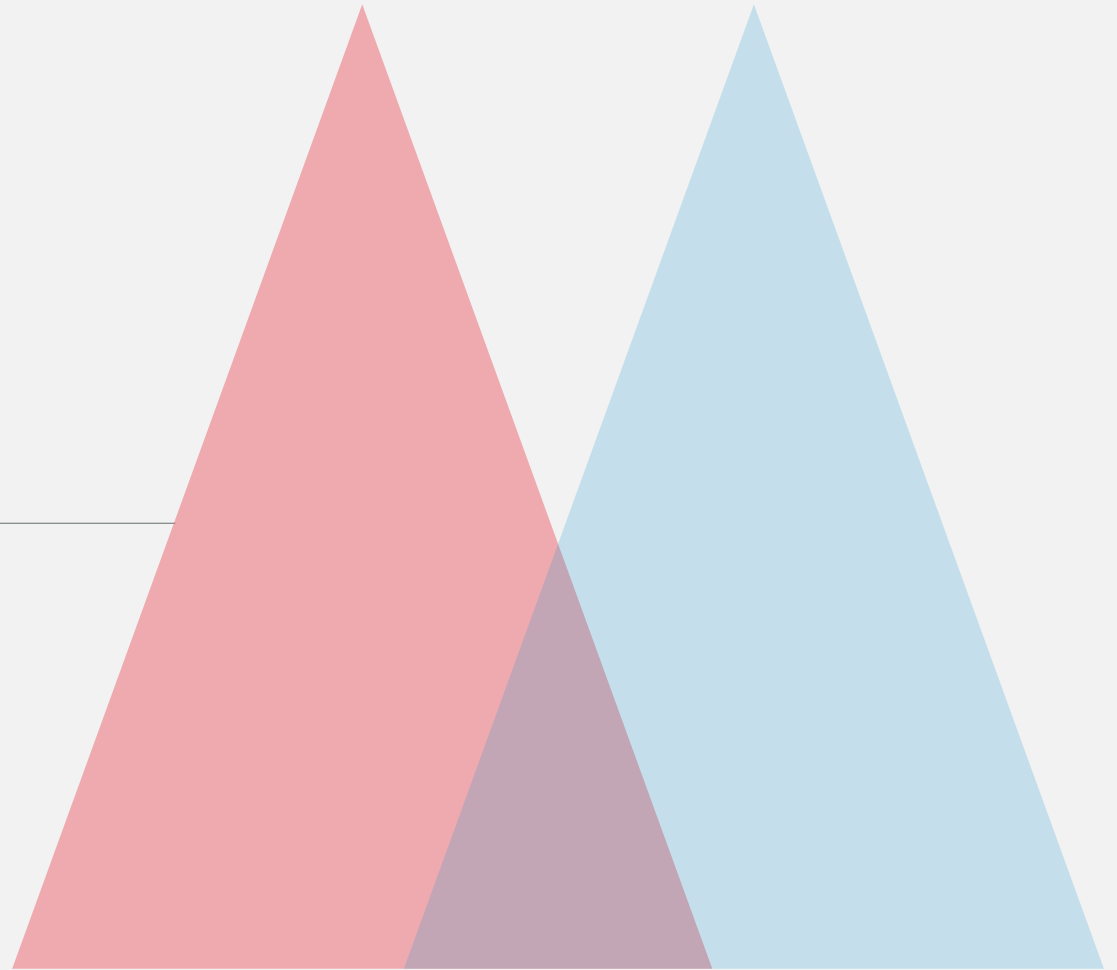
3rd(@so_love_so_)



...

004

플랫폼 서비스 기획



알고리즘 분석

- 우리의 알고리즘은 사진과 글(생략 가능)을 가진 객체의 집합이 여럿 존재할 때, 한 집합의 한 객체와 유사한 다른 집합의 객체 간 순위를 매긴다고 요약 가능하다.
- 우리 알고리즘의 사진과 글을 지닌 어떤 객체라도 적용된다는 강점을 가진다.
- 한 객체 내에서 사진과 글이 그룹으로 존재하더라도 작동한다.

<예시>

- 영화 작품 포스터 + 시놉시스
- 상품 구매 후기 사진 + 내용



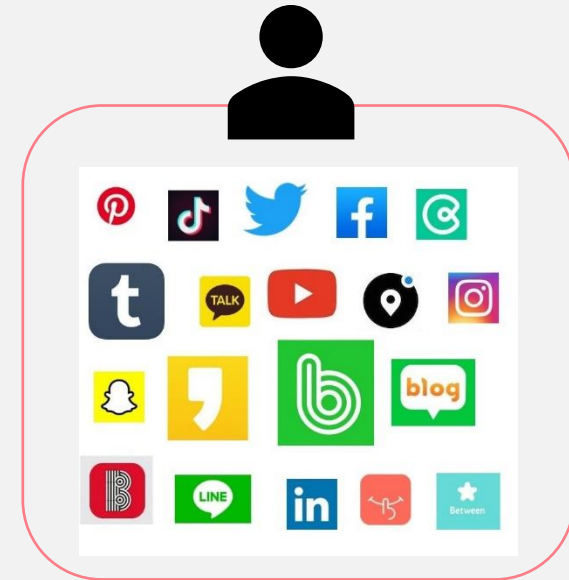
기획안

모도리 (빈틈없이 아주 여무진 사람)

- 한 사람의 모든 SNS, 기록 플랫폼을 연동하는 플랫폼. 플랫폼의 플랫폼.
- 플랫폼 초월적인 사진+글 정보를 바탕으로 다른 사용자를 추천

<기능>

1. 나의 여러 플랫폼을 한꺼번에 관리
2. 서로 다른 플랫폼 간의 사용자를 추천. 추천에 사용할 나와 상대의 플랫폼을 선택 가능



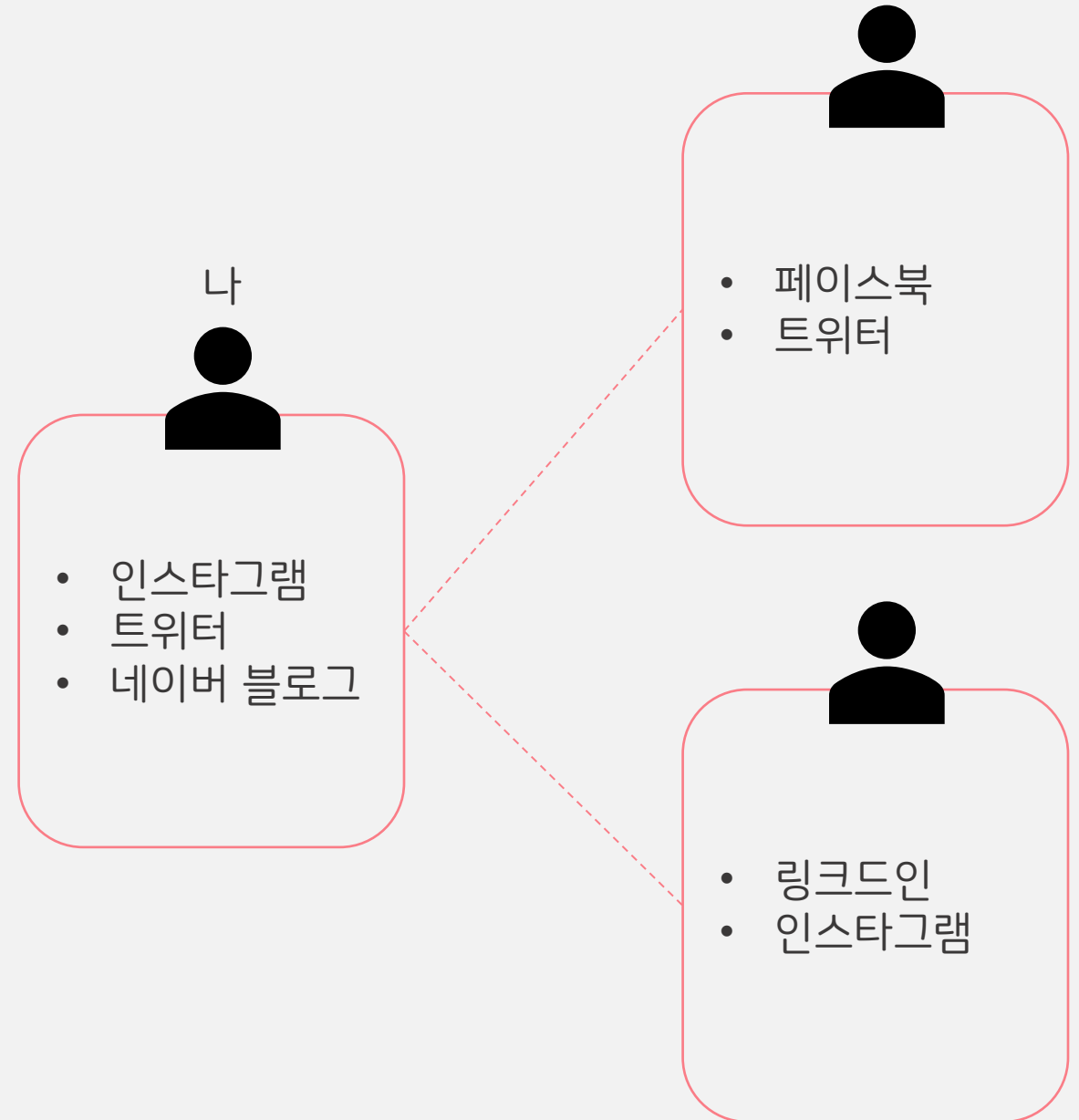
기획안

모도리 (빈틈없이 아주 여무진 사람)

- 한 사람의 모든 SNS, 기록 플랫폼을 연동하는 플랫폼. 플랫폼의 플랫폼.
- 플랫폼 초월적인 사진+글 정보를 바탕으로 다른 사용자를 추천

<기능>

1. 나의 여러 플랫폼을 한꺼번에 관리
2. 서로 다른 플랫폼 간의 사용자를 추천. 추천에 사용할 나와 상대의 플랫폼을 선택 가능



기획안

모도리 (빈틈없이 아주 여무진 사람)

- 한 사람의 모든 SNS를 플랫폼을 연동하는 플랫폼. 플랫폼의 플랫폼.
- 플랫폼 초월적인 사진+글 정보를 바탕으로 다른 사용자를 추천

<기능>

1. 나의 여러 플랫폼을 한꺼번에 관리
2. 서로 다른 플랫폼 간 사용자 추천. 추천에 사용할 나와 상대의 플랫폼을 선택 가능

<예상 문제>

연동할 플랫폼 회사가 있어야 하며, 많아야 한다.

<해결 방안>

연동 플랫폼 회사의 존재를 알릴 수 있고, 거대 플랫폼 회사의 경우 유입을 늘리는 창구로 활용 가능하다.

감사합니다.