

A collection of cosmetic products is arranged on a white surface. In the bottom left, there is a black palette with several shades of red and pink. Above it, a pink-handled brush with a black ferrule and brown bristles lies diagonally. To the right of the palette, a clear plastic brush holder with multiple black brushes is visible. In the bottom right, two lipstick tubes are shown; one has a gold band and the other a black band, both with red lipstick. The background is white with a large, light pink triangular shape in the top right corner.

# 화장품 리뷰데이터 기반 고민 맞춤형 화장품 추천 솔루션

# INDEX

- 
- 01 주제 소개
  - 02 데이터 소개
  - 03 피드백 및 모델
  - 04 실험 결과

## 고민 커뮤니티

초록글 6.10 3:40 | • 지속력 진짜 미친듯이 좋은 립 추천좀 58 ...

1년 전 (2018.6.10 0:47) | 조회 4005 | 👍 0

ㅈ ㄱ ㄴ



## 리뷰 커뮤니티



chohe0214

17세 · 복합성 · ♀ 🍷 평평!

처음엔 다 어둡고 색깔도 별로인줄 알았는데 막장의 정석 발라보니까 저 생각을 왜 했나 싶을 정도로 너무너무 예쁘고 지속력도 좋았어요 ㅈㅅㅅㅅㅅㅅ 꼭 사세요 후회하지 않아요 ㅈㅅㅅㅅㅅㅅ



## 해결책



라스트 벨벳 립 틴트

## 1 고객 맞춤형 솔루션

- 개인의 성향을 파악한 맞춤형 추천
- 마케팅 등 향후 비즈니스 가치 창출



## 2 부수적인 기대효과

- 팀원들 여자친구에게 점수 딸 기회



## 화장품 리뷰데이터 기반 고민 맞춤형 화장품 추천 솔루션

Question



Review

## 화장품 리뷰데이터 기반 고민 맞춤형 화장품 추천 솔루션

### 뷰티 고민글

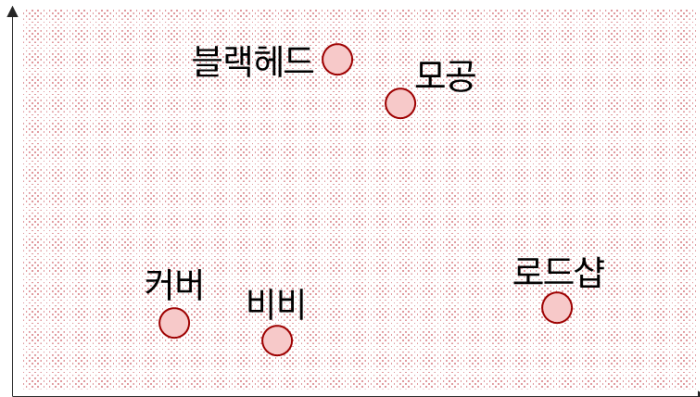
로드샵에서 모공이나 블랙헤드  
잘 커버되는 비비 추천 좀..ㅠㅠㅠ

### 화장품 리뷰

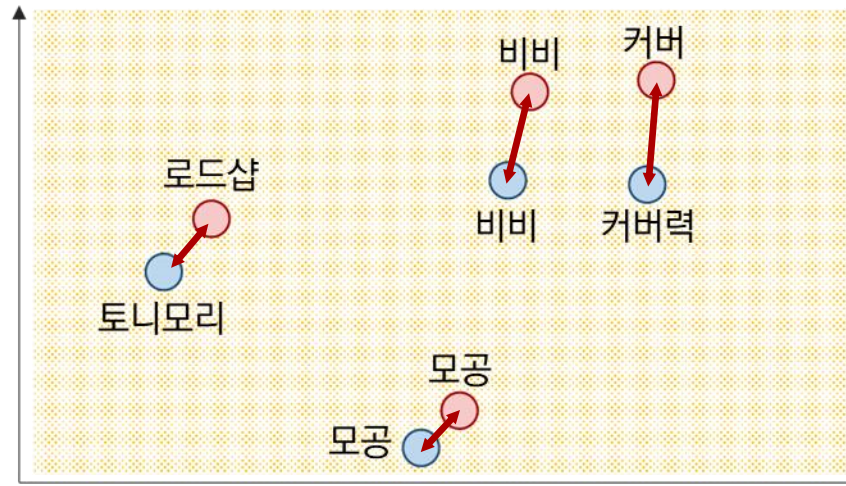
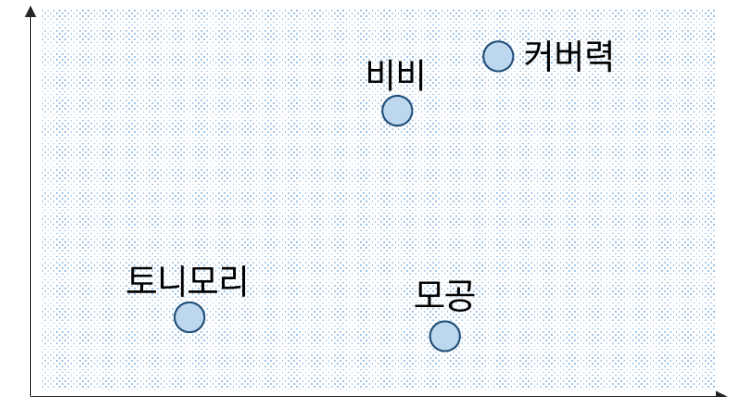
이 비비 덕분에 인생에 안정감을 찾은 느낌이에요 ㅎㅎ  
저처럼 모공 넓은 분들은 모공 부각 안 되구 커버력도 좋은 편!  
저는 토니모리 세일할때 샀는데 그때 사면 가성비도 좋은 것 같구...

### 추천 Solution Space

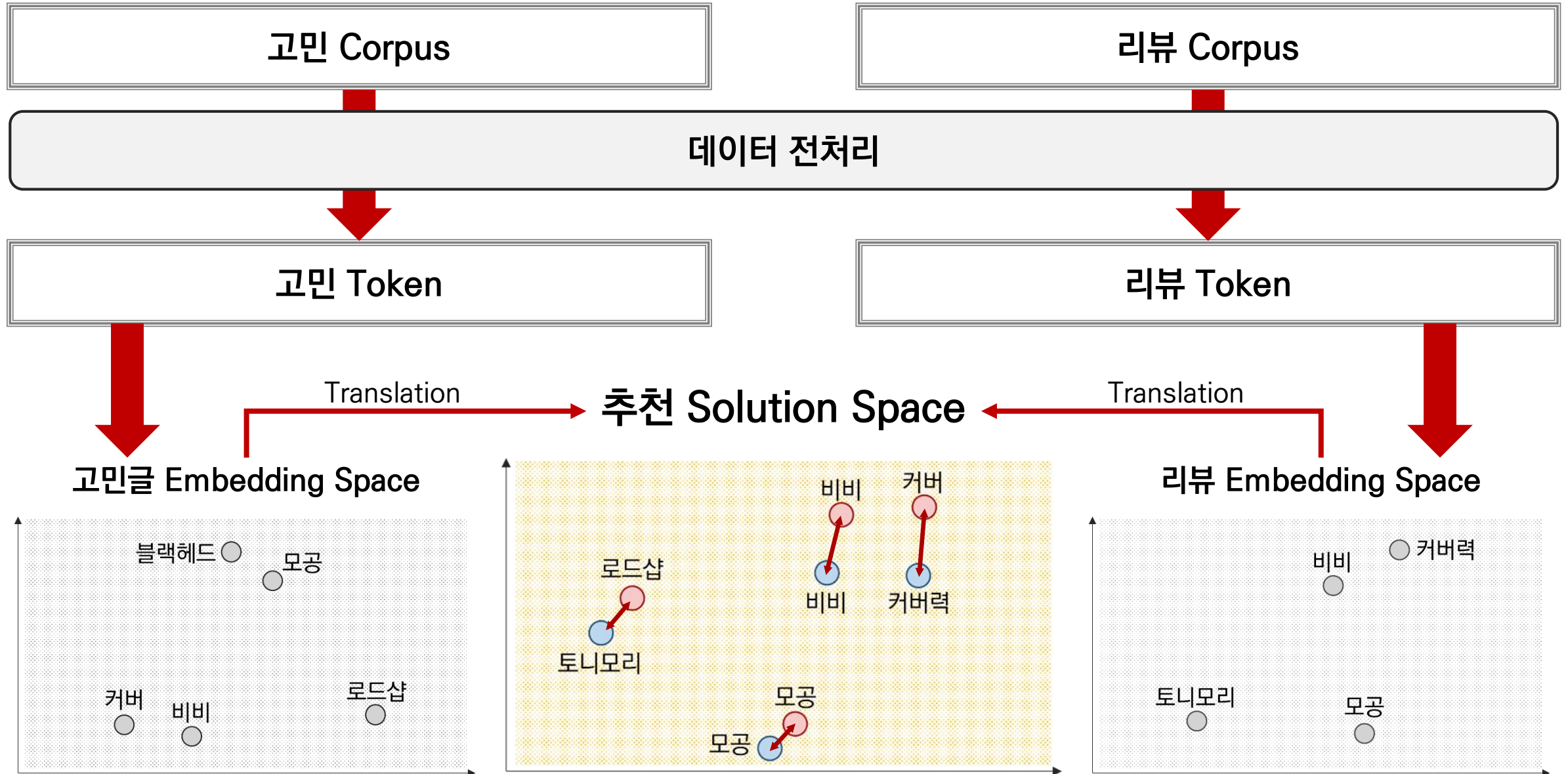
#### 고민글 Embedding Space

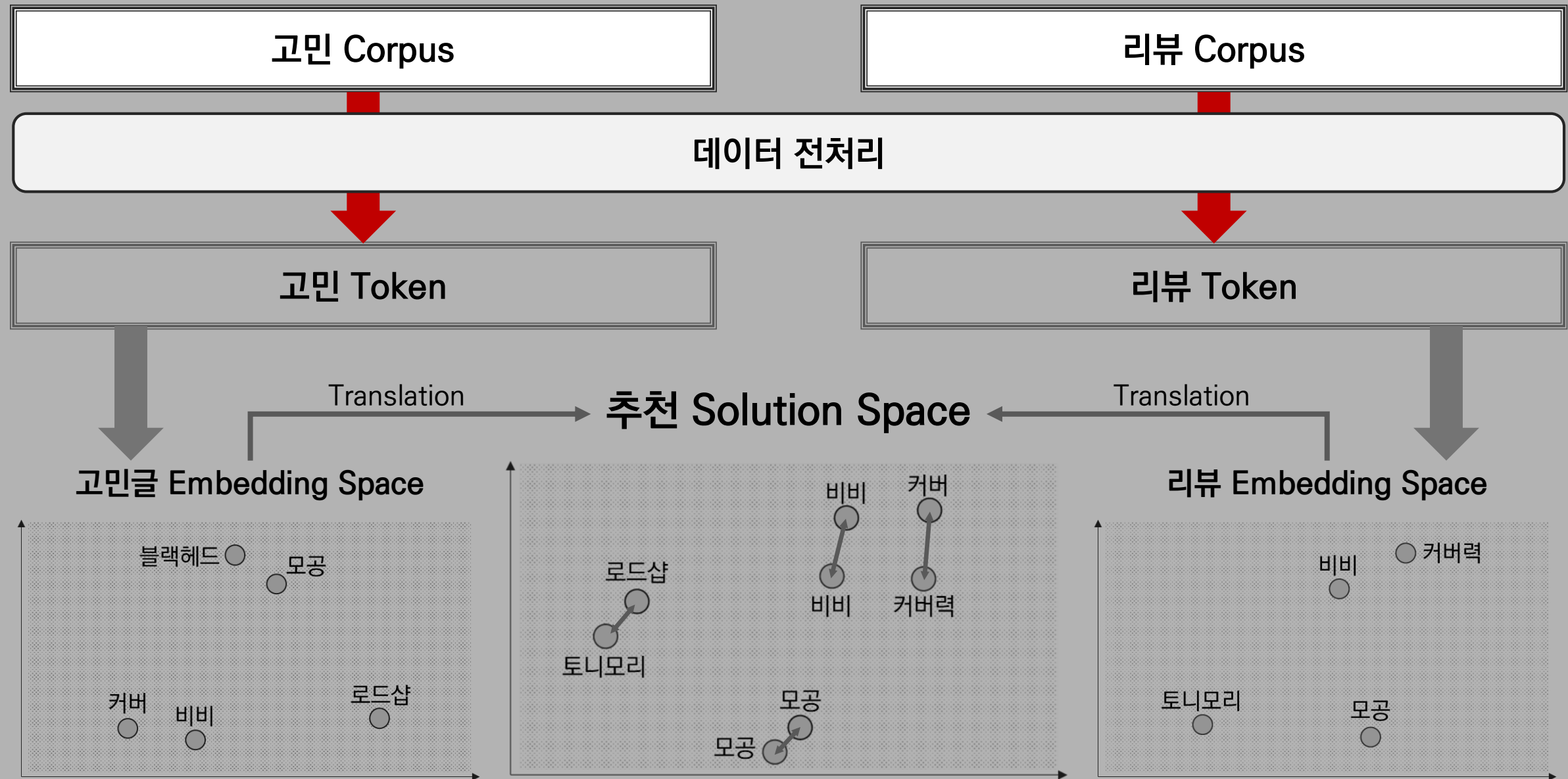


#### 리뷰 Embedding Space











## [ 뷰티 고민글 ]



파우더룸



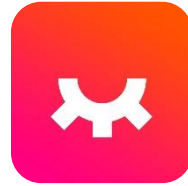
인스티즈



스레딕



더쿠



언니의파우치

- 총 게시글 수 : 244,760개
- 글 제목 및 내용 (댓글 제외)
- 커뮤니티별 게시글 분포
  - 더쿠 : 189,362개
  - 인스티즈 : 32,580개
  - 파우더룸 : 18,801개
  - 언니의파우치 : 3,406개
  - 스레딕 : 971개

## [ 화장품 리뷰 ]



글로우픽

- 총 리뷰 수 : 405,617개
- 총 제품 수 : 11,500개
- 데이터 내용
  - 브랜드 명 / 제품 명
  - 성분 (한글명 / 영어명 / 안전도 지수 / 용도)
  - 리뷰어 정보 (아이디 / 연령 / 피부타입 등)
  - 리뷰 (텍스트 / 평점)

## ① 크롤링 노이즈 제거

- 빈칸 및 오류메시지 제거
- 기타 불필요한 로그 제거

## ② 고민글 제목 및 내용 Concat.

- 고민글의 제목과 내용을 공백을 사이에 두고 Concatenation

## ③ 사용자 사전 추가 및 Tokenizing

- Mecab Tokenizer 활용

### 제품 및 브랜드 명

이니스프리  
아리따움  
입생로랑  
모이스처라이징 립밤 클래식  
아쿠아 글로우 쿠션 파운데이션  
...

### 제품 카테고리

비비 / 씨씨  
워시오프  
선크림  
리무버  
파운데이션  
...

### 뷰티 용어 및 신조어

저렴이 / 고렴이  
수부지 / 지성 / 건성 / 복합성  
화알못 / 화잘알  
여쿨 / 갈웁 / 겨쿨  
톤체성 / 톤궁예 / 톤망진창  
...

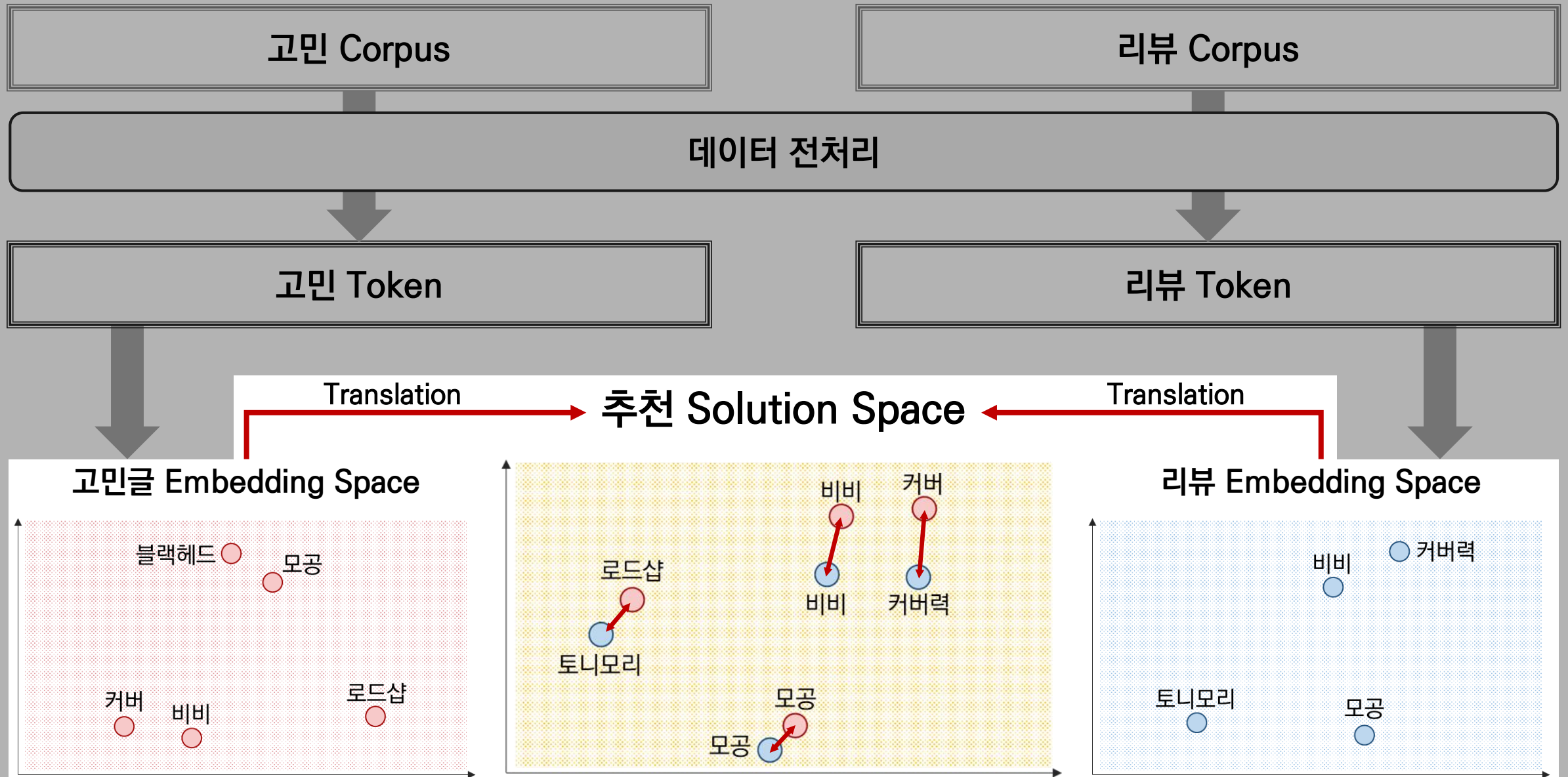
## 4 불용어 처리

- 괄호 / 특수기호 / 링크 / 줄바꿈 / 물음표 / 마침표 / 쉼표 제거
- 관사나 조사 등은 따로 제거하지 않음 (Embedding 시행 후 제품 추천 과정에서 처리)

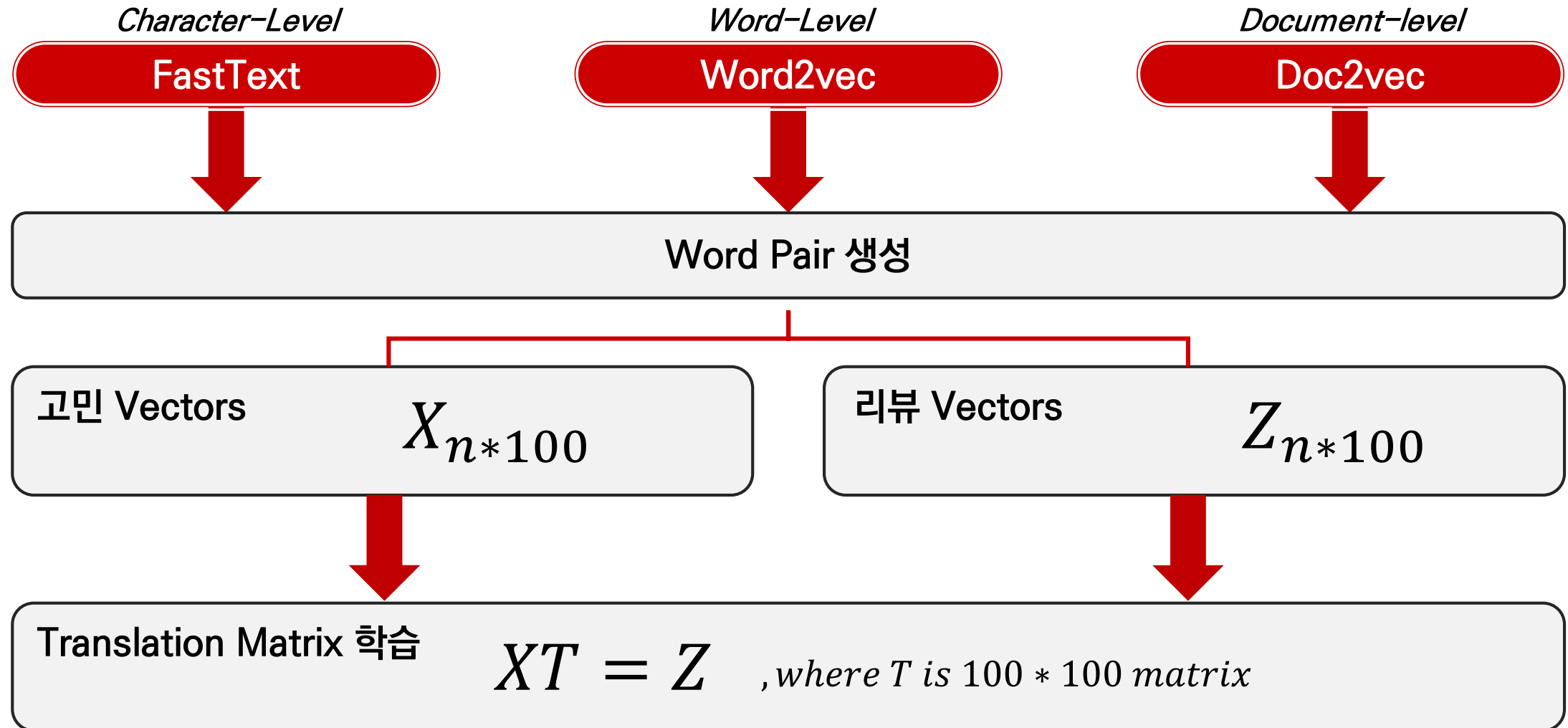
로드샵에서 모공이나 블랙헤드  
잘 커버되는 비비 추천 좀..ㅠㅠㅠ



['로드샵\_NNP', '에서\_JKB',  
'모공\_NNG', '이나\_JC',  
'블랙헤드\_NNP', '잘\_MAG',  
'커버\_NNP', '되\_VV', '는\_ETM',  
'비비\_NNP', '추천\_NNP', '좀\_MAG']



# Translation Matrix



- $T$  : Non-linear relationship 으로 확장 가능



## 1 Cosine Similarity

Input 고민글  $x$ 에 대해 Translation 수행 :  $x^T = x'$

Translation된  $x'$ 과 리뷰 vector 간 Cosine 유사도 측정 :  $sim = \frac{x' \cdot z_i}{|x'| * |z_i|}$

$x'$ 과 Cosine 유사도가 가장 높은 리뷰 vector 추천 :  $z_{select} = \operatorname{argmax}_{z_i}(sim)$

- 고민글의 **핵심어 및 제품 카테고리**를 선별하지 못함
- word기반 Translation을 Document에 적용한 것이 효과적이지 못했을 것으로 추정
- **길이가 긴 리뷰에서 Cosine 유사도가 크게 산출** (문장길이에 대한 Robustness 확보 필요)

## 2 TF-IDF Based Method

고민 Token

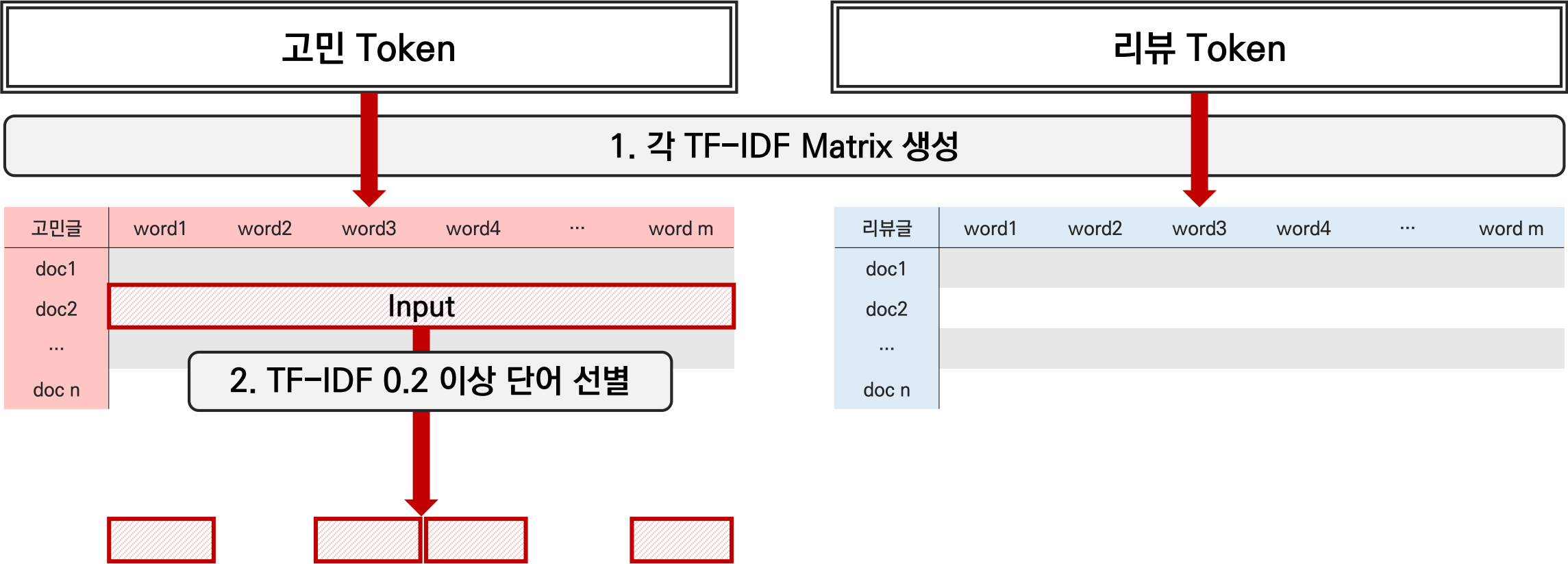
리뷰 Token

1. 각 TF-IDF Matrix 생성

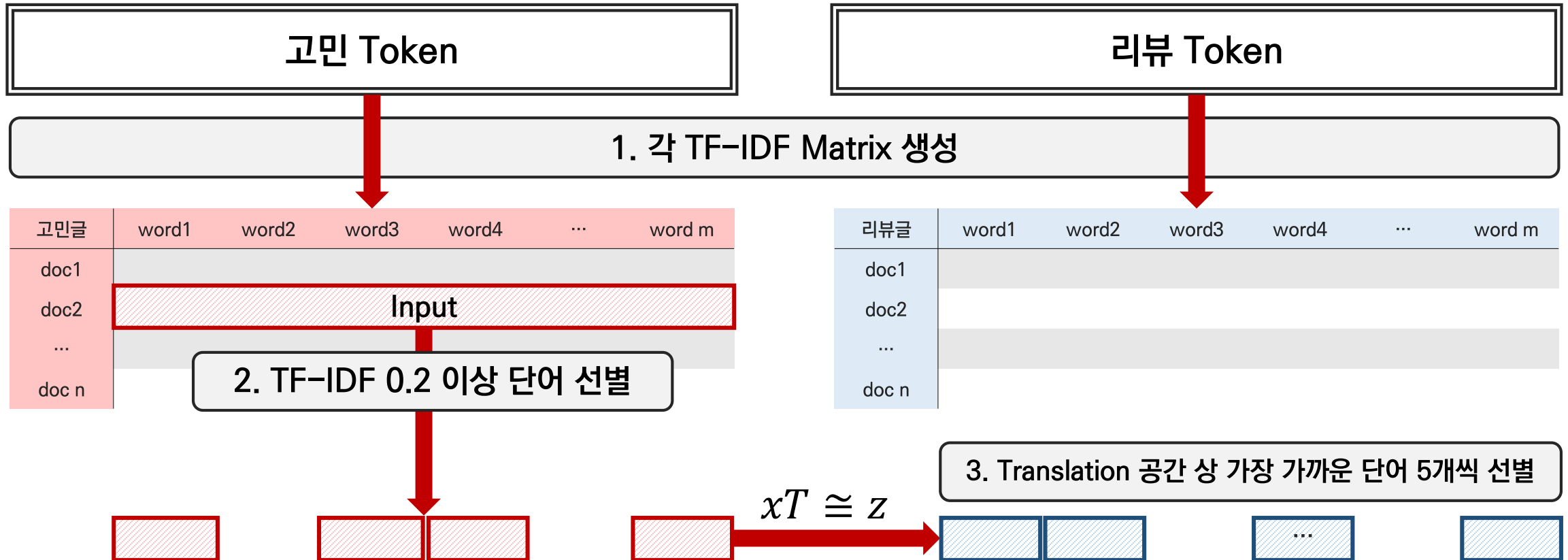
고민글	word1	word2	word3	word4	...	word m
doc1						
doc2						
...						
doc n						

리뷰글	word1	word2	word3	word4	...	word m
doc1						
doc2						
...						
doc n						

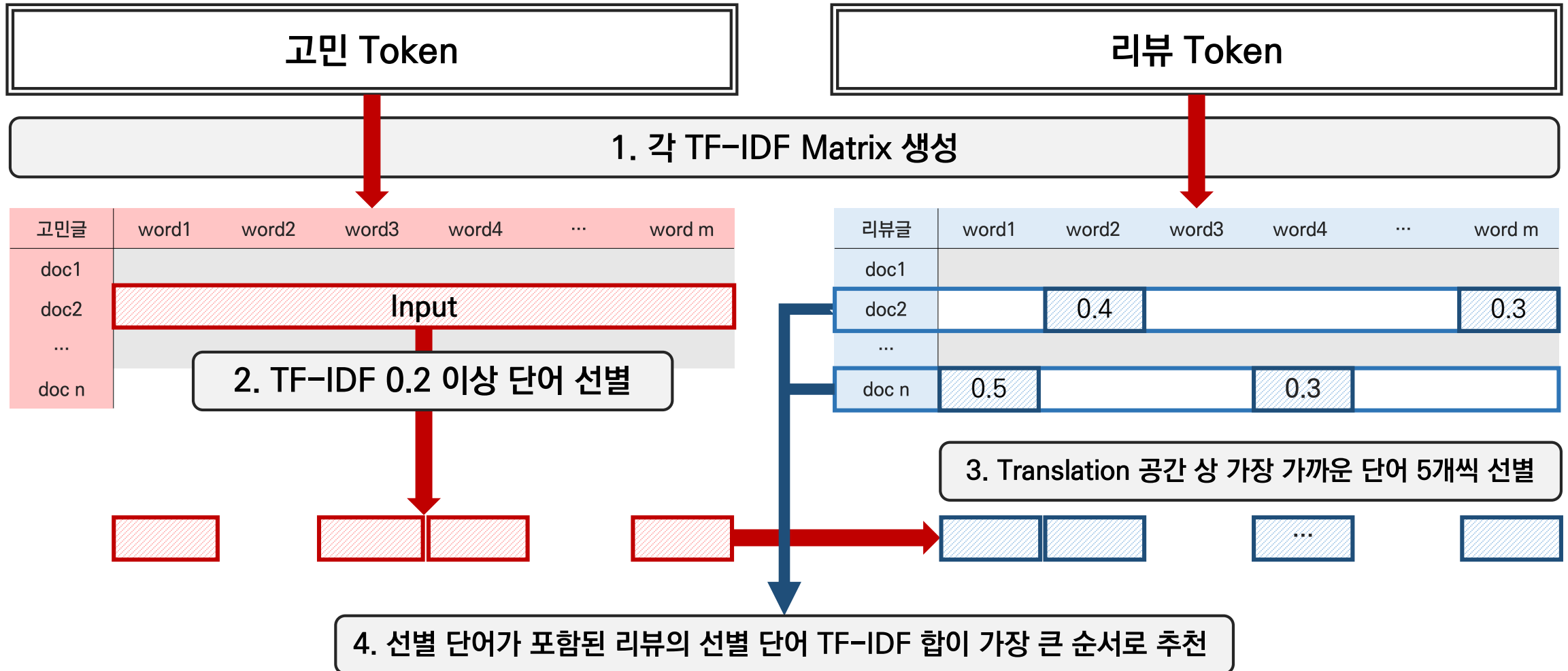
## 2 TF-IDF Based Method



## 2 TF-IDF Based Method



## ② TF-IDF Based Method





## 2 TF-IDF Based Method

- Cosine similarity 대비 **핵심어 반영비율이 높음** (카테고리 추천 방향은 개선 필요)
- 리뷰 길이에 **상대적으로 Robust함**

### [ 뷰티 고민글 ]

“**애꽃살** 1도 없는데 **애꽃살**에 뭐 바르면 좋을까  $\pi\pi\pi$ ?”

### [ 유사 리뷰글 ] (아이새도우 추천 리뷰)

“**애교살**에 바르면 짱짱 이뻐요  $\neg\neg$ ”

“언더 **애교살**용으로 쓰는데 부드러워서 좋아요”

“**애교살** 새도우로 제격이에요. 펄도 작고 **애교살**에 바르면 원래 내 **애교살**인듯 자연스럽게 표현 가능해요”

## 대표 피드백

“**화해 어플** 데이터가 도움을 줄 것이다”

“메타 정보 데이터가 도움을 줄 것이다”

“고민글에서 제품을 명확하게 추정할 수 있다면  
리뷰글 제품 매칭에 도움을 줄 것이다”

“같은 공간에서 임베딩이 좋을 것 같다”



## 답변

웹 페이지가 아닌 **DB 직접** 접근 요청식



해당 코드를 모방하지 않으면,  
**데이터 수집 어려움**

## 대표 피드백

“화해 어플 데이터가 도움을 줄 것이다”

“**메타 정보** 데이터가 도움을 줄 것이다”

“**고민글에서 제품을 명확하게 추정**할 수 있다면  
리뷰글 제품 매칭에 도움을 줄 것이다”

“같은 공간에서 임베딩이 좋을 것 같다”



## 답변

- **카테고리를 분류하는 모델**을 이용하여 Word Vectors Fine-Tuning
- Word Embedding 시 **메타정보를 Token으로 함께 반영**하여 학습

## 대표 피드백

“화해 어플 데이터가 도움을 줄 것이다”

“메타 정보 데이터가 도움을 줄 것이다”

“고민글에서 제품을 명확하게 추정할 수 있다면  
리뷰글 제품 매칭에 도움을 줄 것이다”

“같은 공간에서 임베딩이 좋을 것 같다”

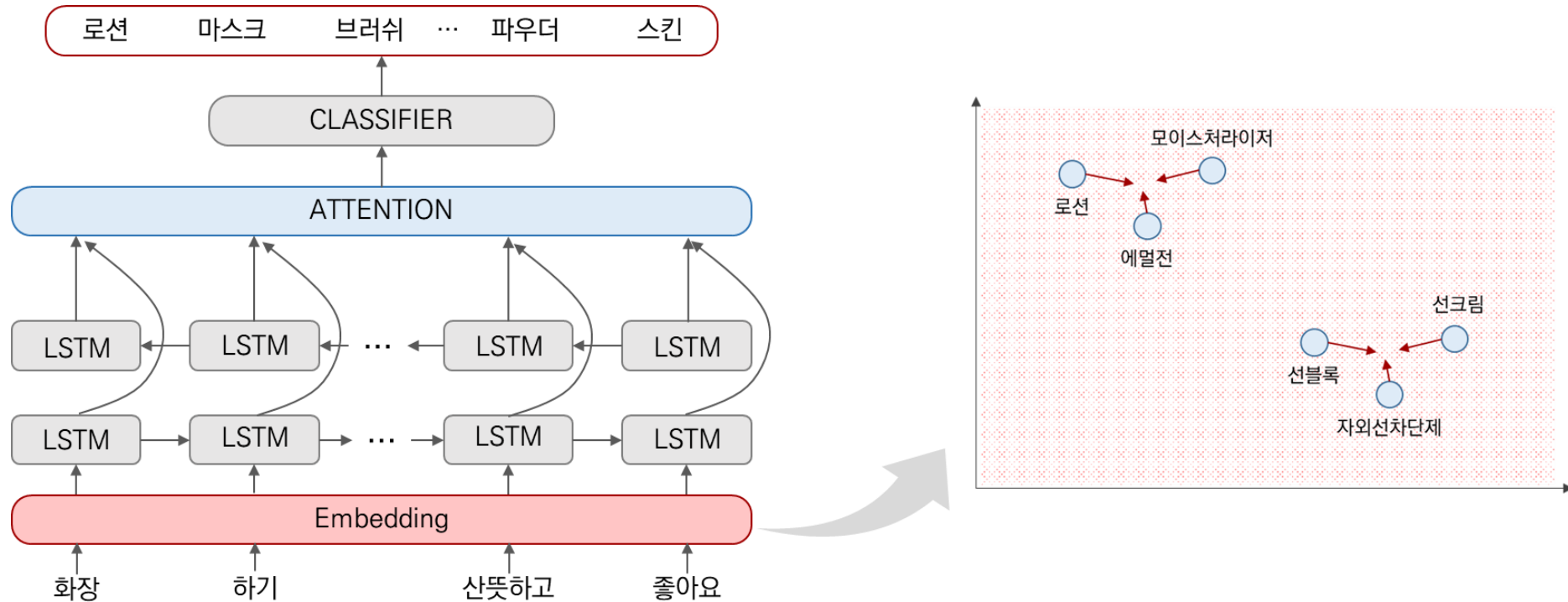


## 답변

- **카테고리를 분류하는 모델**을 이용하여 Word Vectors Fine-Tuning
- Word Embedding 시 **메타정보를 Token으로 함께 반영**하여 학습
- 고민 데이터와 리뷰 데이터를 **합친 후** 임베딩

## 1 Word Vectors Tuning

- Word2Vec으로 학습시킨 Pre-trained Word Vectors를 사용
- 리뷰글을 이용해 카테고리(63개)를 분류하는 **biLSTM-Attention 모델**을 만들
- 해당 **고민글**에 대한 카테고리를 제대로 **추론**(Inference)하는지 확인





## 1 Word Vectors Tuning

- Word2Vec으로 학습시킨 Pre-trained Word Vectors를 사용
- 리뷰글을 이용해 카테고리(63개)를 분류하는 **biLSTM-Attention 모델**을 만들
- 해당 **고민글**에 대한 카테고리를 제대로 **추론**(Inference)하는지 확인

[ 뷰티 고민글에 대한 카테고리 후보 ] (제대로 된 추천 리뷰를 못 찾음)

“웜톤 핑크레드 **립** 추천해줄덕? 원래 씨쉬어 카바나 같은걸로 사려다가 더 붉은걸로 찾는중이야!”

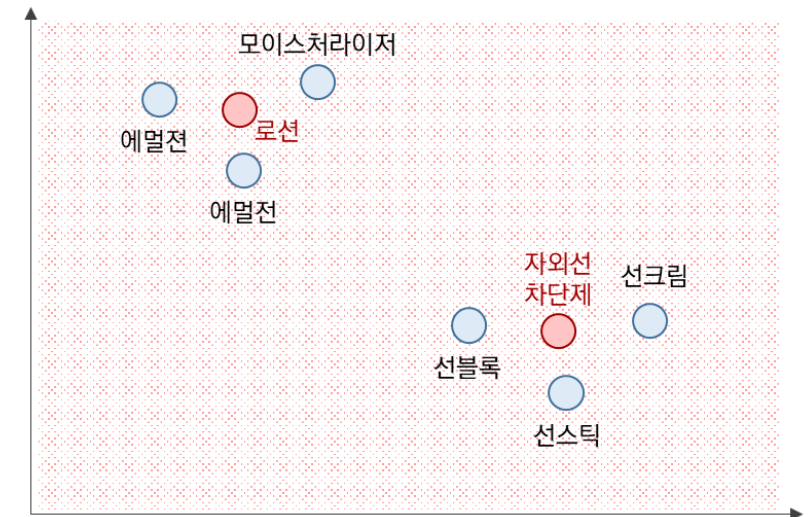
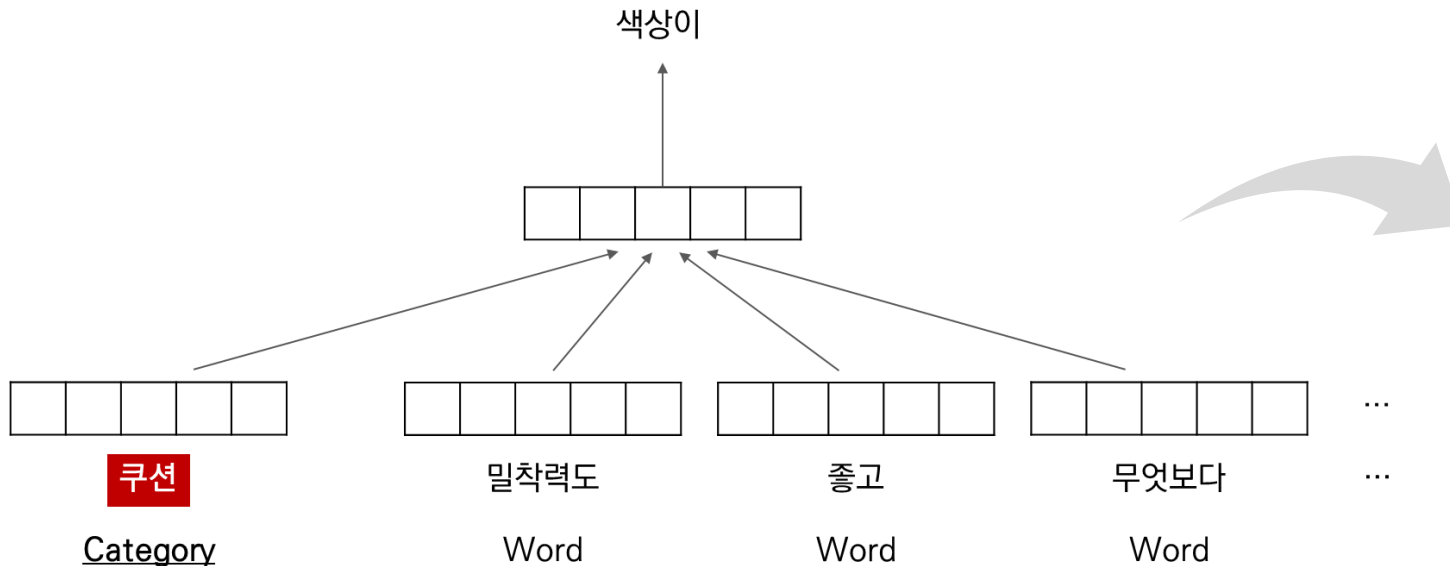
↳ 클렌징, 바디, 브러쉬, 기타, 기능성

“민감성 수부지 덕들 **수분크림** 정착한 거 알려주라. 댕 부탁해 크림 말고 다른 보습템 추천도 고마워”

↳ 클렌징, 마스크, 기능성, 바디, 스킨

## 2 Embed Category

- 카테고리를 추가하여 임베딩 진행
- 해당 **고민글**에 대한 카테고리를 제대로 **추론**(Inference)하는지 확인



## 2 Embed Category

- 카테고리를 추가하여 임베딩 진행
- 해당 **고민글**에 대한 카테고리를 제대로 **추론**(Inference)하는지 확인

[ 뷰티 고민글에 대한 카테고리 후보 ] (제대로 된 추천 리뷰를 못 찾음)

“블러셔 브러시 추천좀 빨리 부탁해 ㅠㅠ”

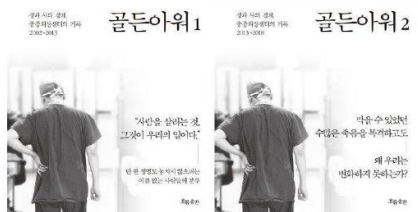
↳ 헤어스타일링, 헤어바디, 트리트먼트팩, 샴푸, 팬티라이너

“로드샵 눈썹그리는거는 어디꺼가 좋음?”

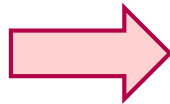
↳ 크림, 페이스오일, 바디로션, 트리트먼트팩, 립밤

## 3 Dictionary-based Approach

- 수작업을 통해 115개의 카테고리에 대한 딕셔너리를 만듦 (네이버 연관 검색어 / 글로우픽 참고)
- 고민글에서 **카테고리**(key)의 **키워드**(value)를 하나라도 포함한다면 해당 카테고리를 **후보군**으로 추가
- TF-IDF 방법을 적용하여 뷰티 고민글과 유사한 리뷰글을 비교함



"문제는 한국 사회에서 시스템이 필요하다고 지시를 내릴 사람은 많은데 전통적으로 "노가다"를 뭘 사람은 없다는 겁니다. 이런 일은 남이 해야 하는 거라 생각하죠. 아니면 남이 했다가 자기한테 해가 되면 안 되니까 오만가지 이유를 대서 이런 일은 하면 안 된다고 하고, 이런 일이 의료계에서만 있는 줄 알았어요. 사회 전반이 바뀌지 않으면 이 문제는 나아지지 않아요"라고 했다.



```
categories = {
    "스킨": ["스킨", "토너", "로션", "워터", "부스터"],
    "로션": ["로션", "젤", "모이스처라이저", "모이스처라이저", "에멀전", "에멀전", "밀크", "모이스처라이저"],
    "에센스": ["에센스", "세럼", "앰플"],
    "크림": ["크림", "젤"],
    "미스트": ["미스트"],
    "메이크업픽서": ["메이크업픽서", "픽서", "스프레이", "미스트"],
    "페이스오일": ["페이스오일", "오일", "세럼", "앰플"],
    "클렌징폼": ["클렌징폼", "클렌징", "폼클렌징", "폼클렌저", "폼", "클렌저", "클렌저"],
    "클렌징오일": ["클렌징오일", "오일", "클렌징", "클렌저", "클렌저"],
    "클렌징밀크": ["클렌징밀크", "밀크", "클렌징", "클렌저", "클렌저"],
    "클렌징크림": ["클렌징크림", "크림", "클렌징", "클렌저", "클렌저"],
    "클렌징워터": ["클렌징워터", "워터", "클렌징", "클렌저", "클렌저"],
    "클렌징젤": ["클렌징젤", "젤", "클렌징", "클렌저", "클렌저"],
    "스크럽/필링": ["스크럽", "필링", "바디스크럽", "젤", "필링젤", "필링패드"],
    "포인트리무버": ["포인트리무버", "리무버", "클리너"],
    "클렌징티슈": ["클렌징티슈", "티슈", "클렌징", "클렌저", "클렌저"],
    "클렌징비누": ["클렌징비누", "비누", "클렌징", "클렌저", "클렌저"],
    "메이크업베이스": ["메이크업베이스", "베이스", "쿠션"],
    "톤업크림": ["톤업크림", "크림", "톤업", "로션"],
    "베이스프라이머": ["베이스프라이머", "베이스", "프라이머"],
    "포인트프라이머": ["포인트프라이머", "프라이머"],
    "파운데이션": ["파운데이션", "파데", "쿠션", ""],
    "비비크림": ["비비크림", "비비", "BB", "bb", "BB크림", "bb크림"],
    "씨씨크림": ["씨씨크림", "씨씨", "CC", "cc", "CC크림", "cc크림"],
    "쿠션타입": ["쿠션타입", "쿠션", "파운데이션", "파데"],
    "컨실러": ["컨실러"],
    "팩트": ["팩트", "파우더"],
    "파우더": ["파우더"],
    "트윈케익": ["트윈케익", "트윈", "팩트", "케익", "케이크", "케일", "트윈케일", "트윈케이크", "트윈팩트"],
    "립스틱": ["립", "립스틱", "립팔라"],
    "립글로스/락커": ["립글로스", "립글로즈", "글로스", "글로즈", "락커", "입술", "립컬러", "립라커", "립라커"],
    "립틴트": ["립틴트", "틴트", "입술"],
    "립밤": ["립밤", "입술"],
    "립라이너": ["립라이너", "라이너", "입술", "초크", "펜슬"],
    "아이라이너-펜슬&젤": ["아이라이너", "펜슬", "젤", "아라", "아이펜슬", "라이너", "젤라이너"],
    "아이라이너-리퀴드": ["아이라이너", "리퀴드", "아라", "아이라이너"]
}
```

## 3 Dictionary-based Approach

- **수작업**을 통해 115개의 카테고리에 대한 딕셔너리를 만듦 (네이버 연관 검색어 / 글로우픽 참고)
- 고민글에서 **카테고리**(key)의 **키워드**(value)를 하나라도 포함한다면 해당 카테고리를 **후보군**으로 추가
- TF-IDF 방법을 적용하여 뷰티 고민글과 유사한 리뷰글을 비교함

### [ 뷰티 고민글 ]

“**워터프루프** **아이라이너** 뭐가 좋아?? 고민중...”

### [ 유사 리뷰글 ]

“진짜 강추합니다!! 지금까지 써 본 펜슬라이너 중에 제일 안 번지는것 같아요. 지금까지 써본 펜슬라이너 중에 가장 좋습니다!!” (더샘, 에코소울 **워터프루프** 젤 **아이라이너**)

“개 얇은데 개 잘그려짐 ... 인생 **아이라이너** 유목생활은 끝남” (클리오, 샤프 쏘 심플 워터프루프 펜슬 라이너)



## 3 Dictionary-based Approach

- **수작업**을 통해 115개의 카테고리에 대한 딕셔너리를 만듦 (네이버 연관 검색어 / 글로우픽 참고)
- 고민글에서 **카테고리**(key)의 **키워드**(value)를 하나라도 포함한다면 해당 카테고리를 **후보군**으로 추가
- TF-IDF 방법을 적용하여 뷰티 고민글과 유사한 리뷰글을 비교함

### [ 뷰티 고민글 ]

“**프라이머나 메이크업베이스** 로드샵꺼 중에 좋은거 추천 해주라!! **모공** 조금이라두 커버되는 걸루ㅠㅠ ... 아니면 모공커버 아니더라도 좋거나 유명한 거 알려주 ♡”

### [ 유사 리뷰글 ]

“로드샵 중 정말 최고를 자랑하는 **프라이머**예요♡ 지복합에 **모공**, 요철 잔 주름 다 있는데 바르는 순간 **모공 커버**, 주름 커버, 요철까지 정리해주는 효자템♡” (이니스프리, 노세범 블러 프라이머)

“인생 **프라이머**!!! 모공은 커버는 안되는데 엄청 실키한 마무리감이 특징 주름 사이에 잘 들어가서 파데 주름 끼임이 없어져요!!!” (네이처리퍼블릭, 진생 로얄 실크 프라이머)

## 3 Dictionary-based Approach

### [ 뷰티 고민글 ]

“내가 이마랑 볼구석 쪽이랑 턱 쪽이 **여드름** 진짜 심한데  $\pi\pi$  **팩**이나 **폼클렌징** 좋은거 추천해줘”

#### 기존 TF-IDF

스킨프로 트리트먼트 에센스

- “민감성에도 순하고 초크초크 해요 답답하 지도 않고 가벼운데 촉촉해서 좋은 것 같아요”

블랙헤드 워시 클리너

- “미간이나 턱쪽에 피지 쫌 빼주는거 같아요! 좋습니다당”

#### Dic-based

포어 브라이트닝 웨이셜 폼

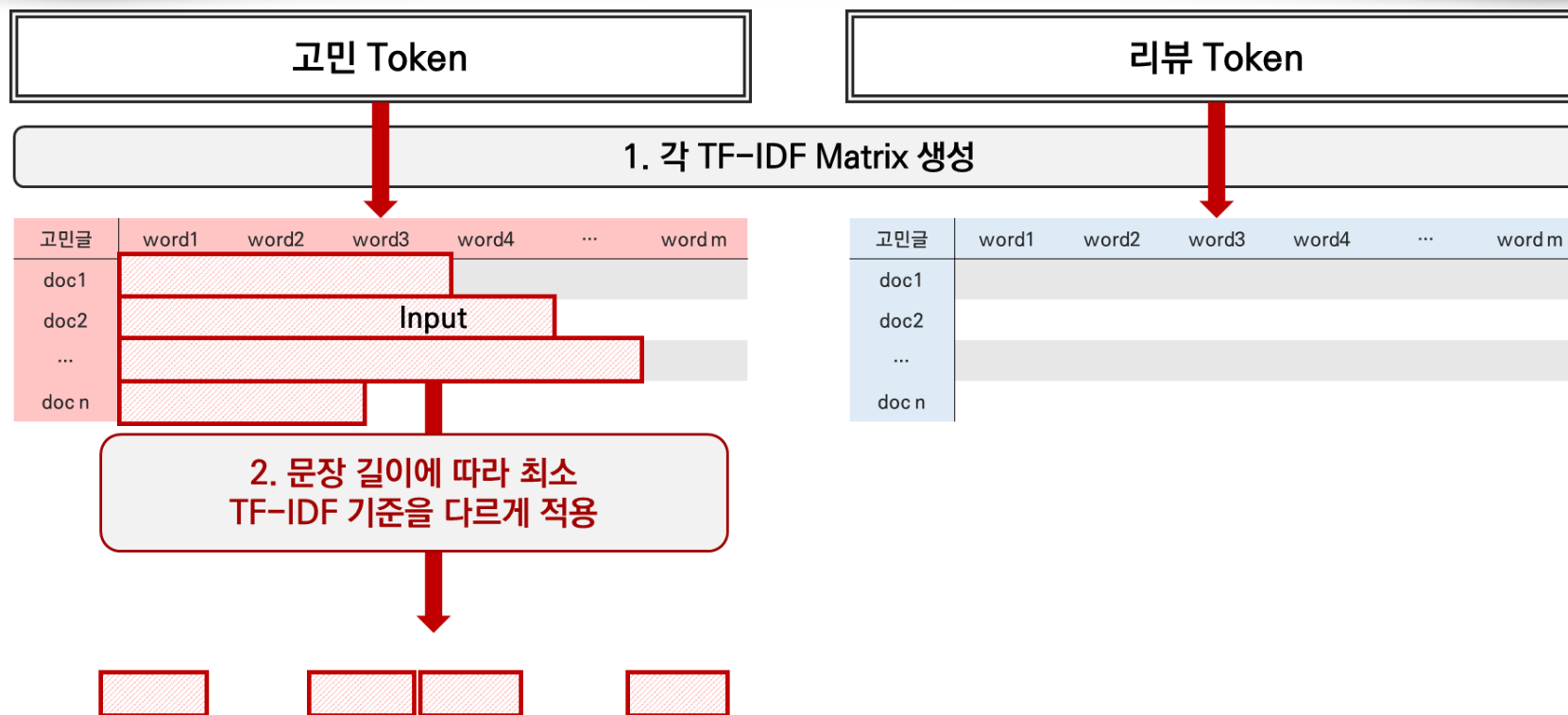
- “학생 때 한창썸뻤던 **클렌징 폼** ! 당시 유분 폭발로 **여드름** 종종 나던 시절이었는데 , 이 클렌징 폼 이 굉장히 산뜻 해서 좋았다”

진액 캡슐**팩** 비자&알로에 [트러블]

- “마스크팩 보다는 수분크림 느낌인데 그래도 좋아요!! ”

## 문장 길이 반영

- 고민글에서 문장 길이에 상관없이 **일정한** 단어의 개수를 추출하기 위함
- 고민글에 대해 **문장 길이**를 반영하여 **TF-IDF 기준을 다르게** 적용



## 1 제품 카테고리 반영

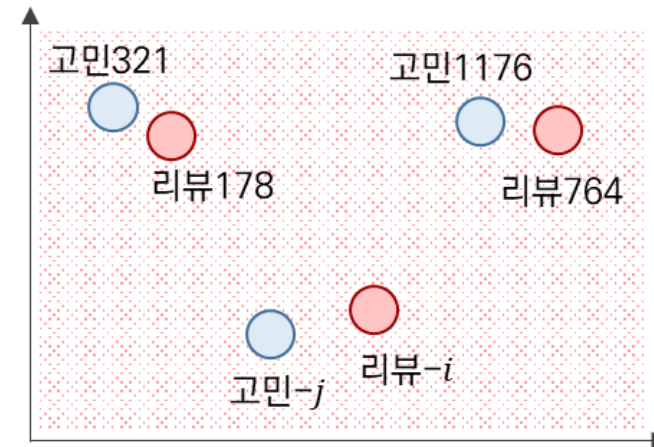
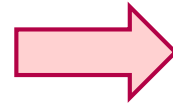
- Word Vectors Fine-Tuning : biLSTM-Attention 분류 모델을 이용
- Embed Category : 제품 카테고리를 함께 임베딩
- Dictionary-based : 제품 카테고리 딕셔너리 구축

## 2 문장 길이 반영

- 고민글에 대해 **문장 길이**를 반영하여 **TF-IDF 기준을 다르게** 적용

## 고민 & 리뷰 Together

- 중간 발표 피드백을 반영해 고민 데이터와 리뷰 데이터를 합친 후 임베딩 진행
- Cosine similarity 를 이용해 뷰티 고민글과 유사한 리뷰글을 비교함 (상위 5개)
- TF-IDF 방법을 적용하여 뷰티 고민글과 유사한 리뷰글을 비교함



## 고민 & 리뷰 Together

- 중간 발표 피드백을 반영해 고민 데이터와 리뷰 데이터를 합친 후 임베딩 진행
- **Cosine similarity** 를 이용해 뷰티 고민글과 유사한 리뷰글을 비교함 (상위 5개)
- TF-IDF 방법을 적용하여 뷰티 고민글과 유사한 리뷰글을 비교함

### [ 뷰티 고민글 ]

“수분 부족 지성 로라 파우더 추천 줘요. 오늘 저녁 출국이라 품절 풀리길 기다렸지만...(생략)”

### [ 유사 리뷰글 ] (제대로 된 추천 리뷰를 못 찾음)

“가격 만크이나 제값 하는 핸드크림 ... 끈적함 없이 촉촉함 유지 ...” (록시땅, 시어버터 핸드크림)

“아침용 클렌저로 쓰면 ... 촉촉한 마무리감 좋아하시면 사세요~” (시드물, 티트리 마일드 클렌저)

## 고민 & 리뷰 Together

- 중간 발표 피드백을 반영해 고민 데이터와 리뷰 데이터를 합친 후 임베딩 진행
- Cosine similarity 를 이용해 뷰티 고민글과 유사한 리뷰글을 비교함 (상위 5개)
- **TF-IDF 방법**을 적용하여 뷰티 고민글과 유사한 리뷰글을 비교함

### [ 뷰티 고민글 ]

“이쁜 무몽이들아 로드샵 **클렌징오일** 추천해줘!! 내 피부는 건성 그치만 튼튼하다ㅋㅋ”

### [ 유사 리뷰글 ] (제대로 된 추천 리뷰를 못 찾음)

“색도 선명하고 잘 그려져서 좋아요~~” (클리오, 워터프루프 **펜라이너**)

“발색도 좋고 지속력도 좋고 색도 너무 예쁜데 블렌딩이 잘 안된다는 점ㅜㅜ” (이니스프리, 리얼핏 매트 리퀴드)

Q & A

