

# 검색어와 상품을 연결하는 스마트한 방법: 네이버 쇼핑 검색 커버리지 개선 기술

송일현 NAVER  
하종우 NAVER

# CONTENTS

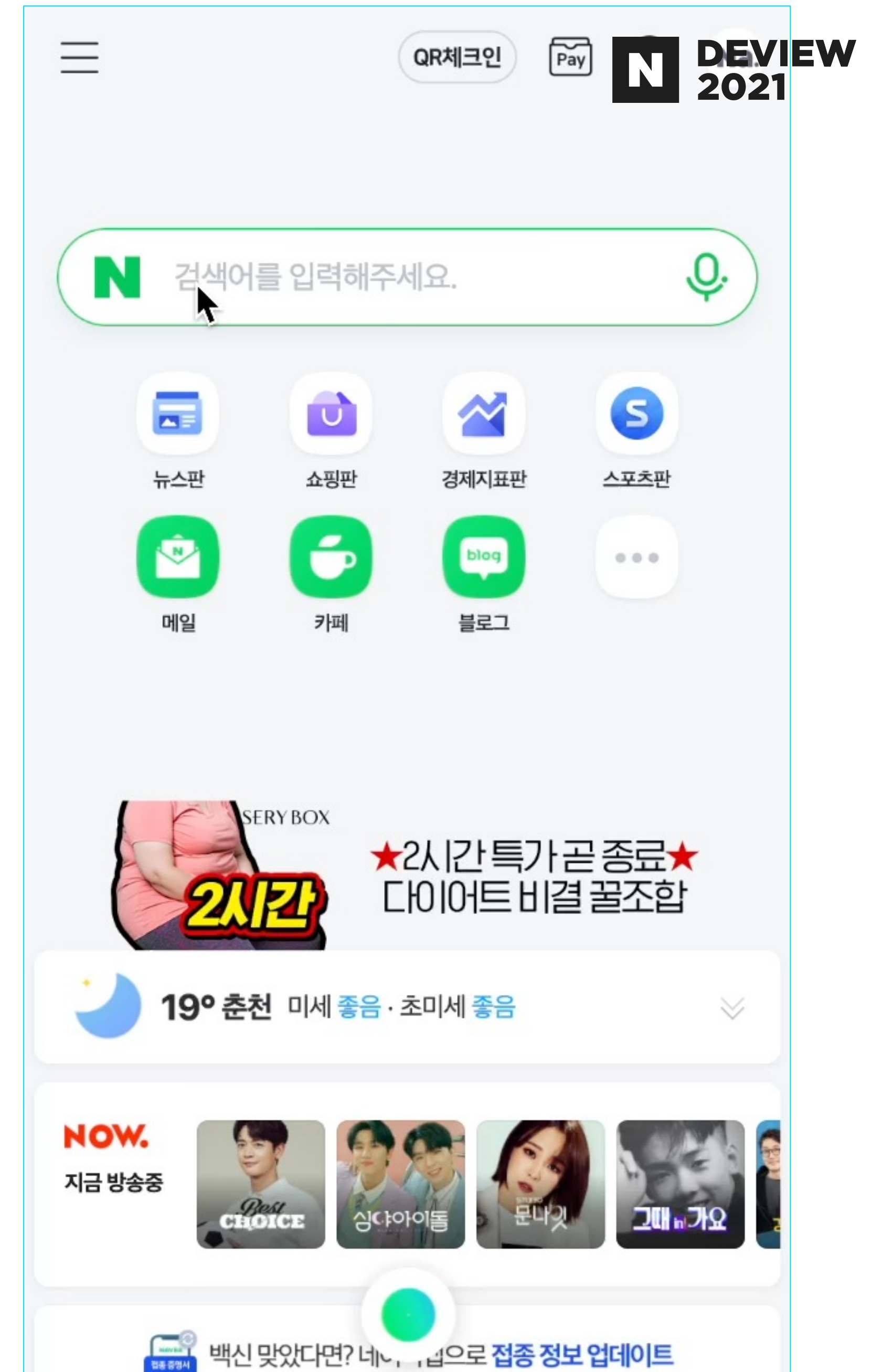
1. 네이버의 쇼핑 검색
2. 검색어 확장: Query Reformulation
3. 검색어 부분 일치 : Query Term Importance
4. BERT 기반 검색: 쇼핑 딥서치

# 1.네이버의 쇼핑 검색

# 1.1 쇼핑 검색 서비스

## 네이버 통합검색의 **쇼핑** 컬렉션 담당

- 쇼핑 의도가 있는 검색어에 '네이버쇼핑' 상품 노출
- 쇼핑 검색어(통합 검색 27%)와 12억 개의 상품 매칭

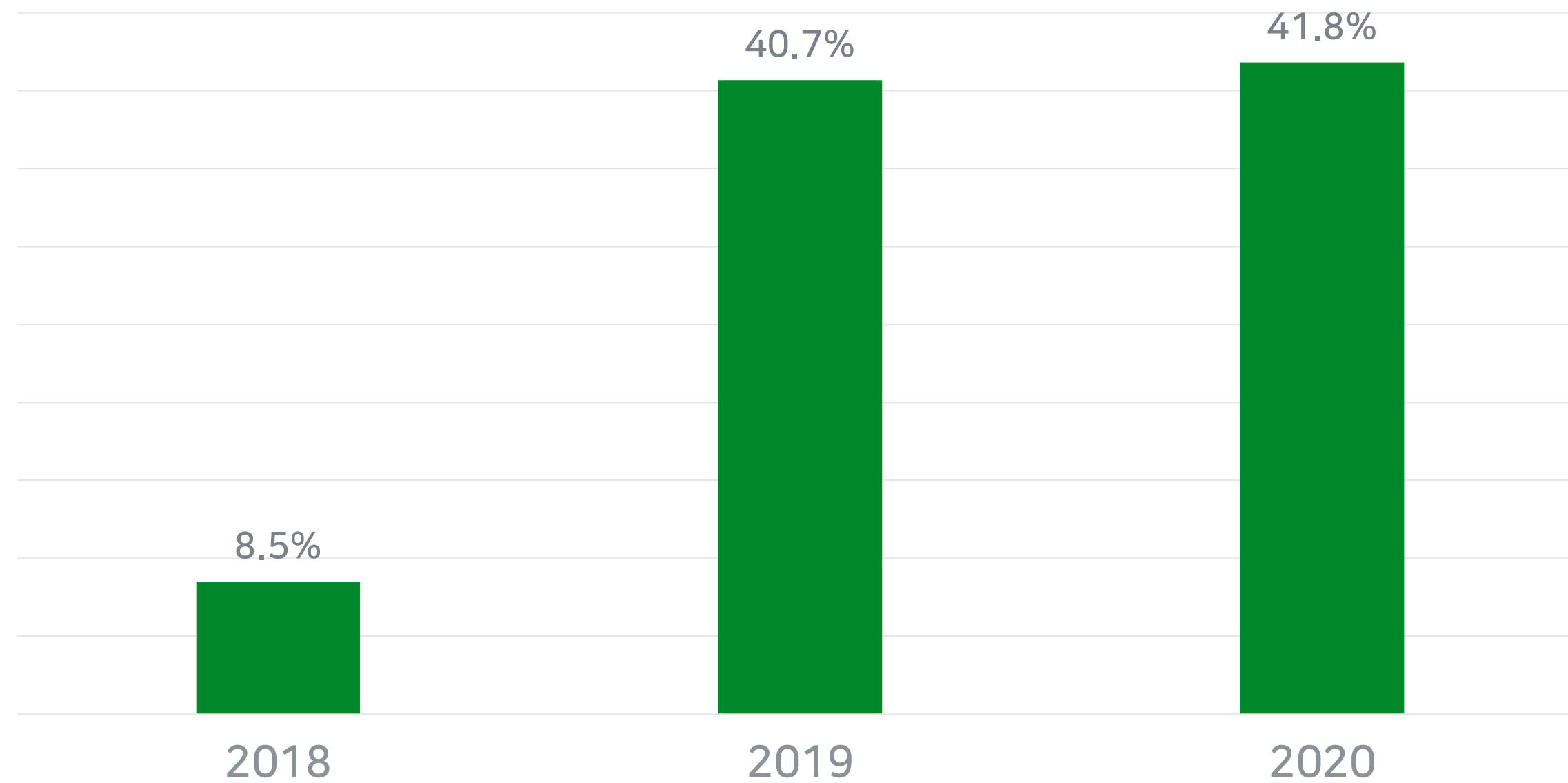


# 1.2 검색 커버리지 개선

점점 다양해지는 검색어 대응 필요

쇼핑 Unique Query Count

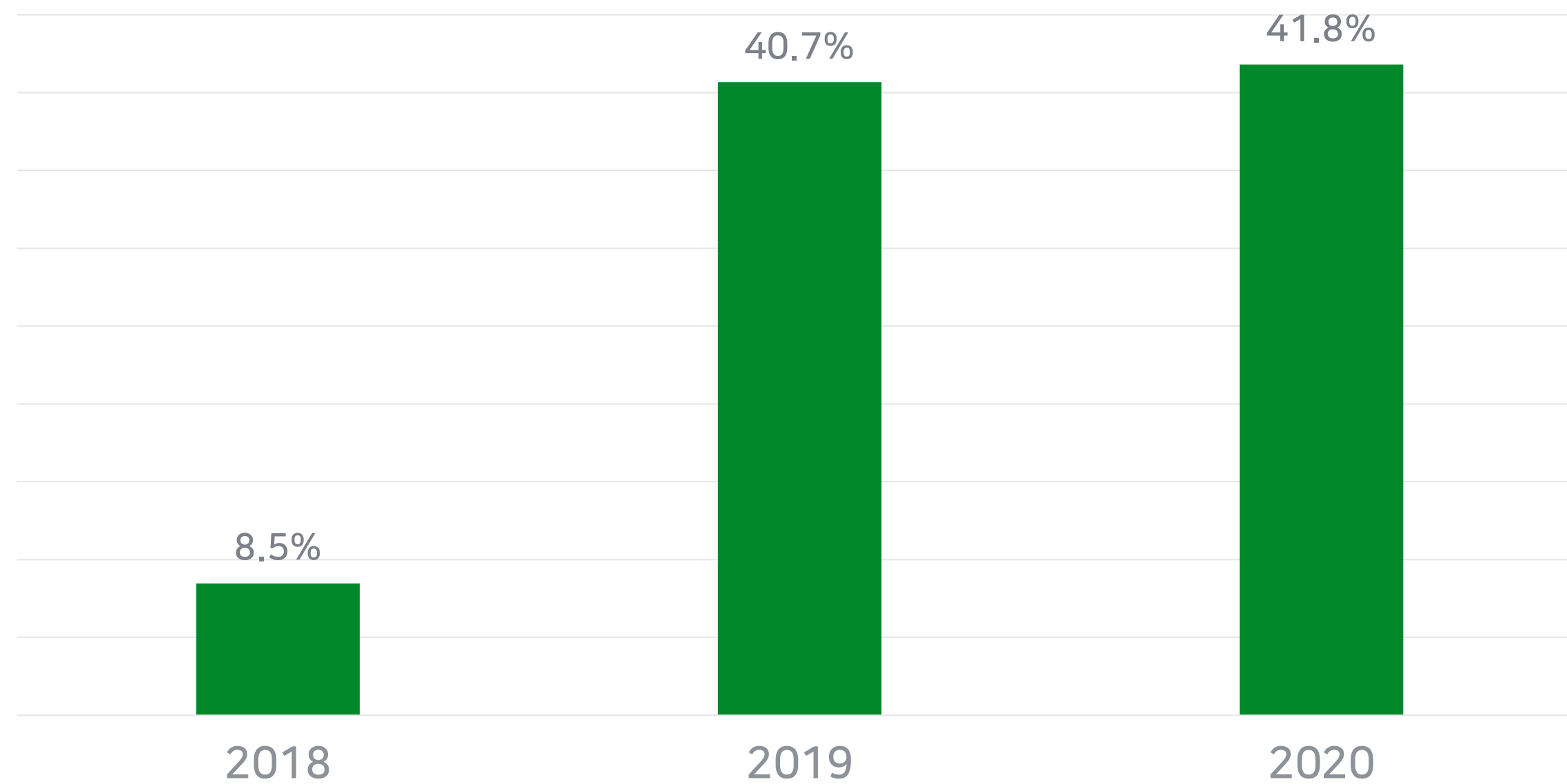
■ 전년대비증가량 (YoY)



# 1.2 검색 커버리지 개선

쇼핑 Unique Query Count

■ 전년대비증가량 (YoY)



점점 다양해지는 검색어 대응 필요

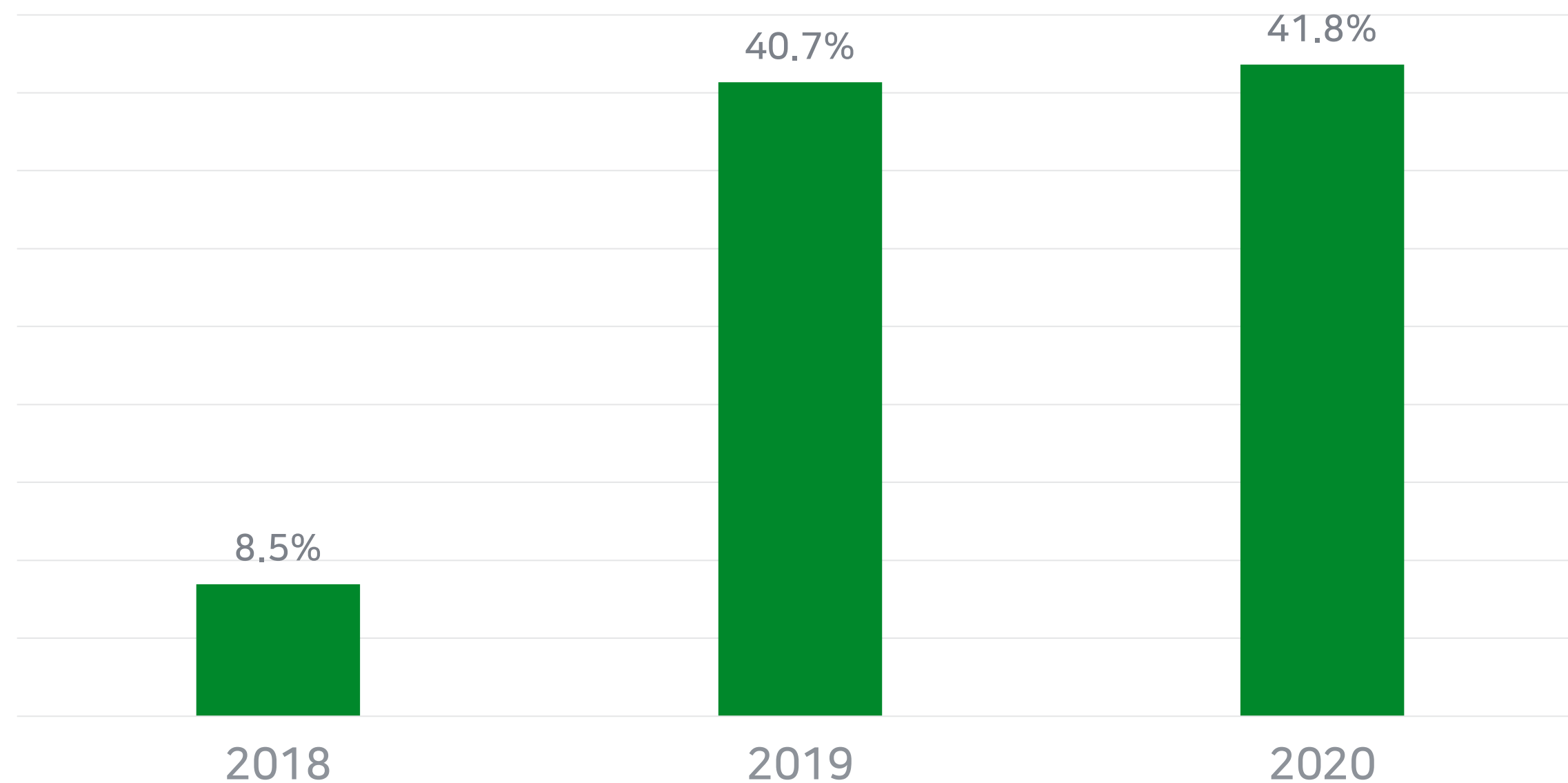
- 검색어 - 상품명과 단어 mismatch

대나무 **이파리** → 대나무 **잎**

# 1.2 검색 커버리지 개선

쇼핑 Unique Query Count

■ 전년대비증가량 (YoY)



## 점점 다양해지는 검색어 대응 필요

- 검색어 - 상품명과 단어 mismatch

대나무 **이파리** → 대나무 **잎**

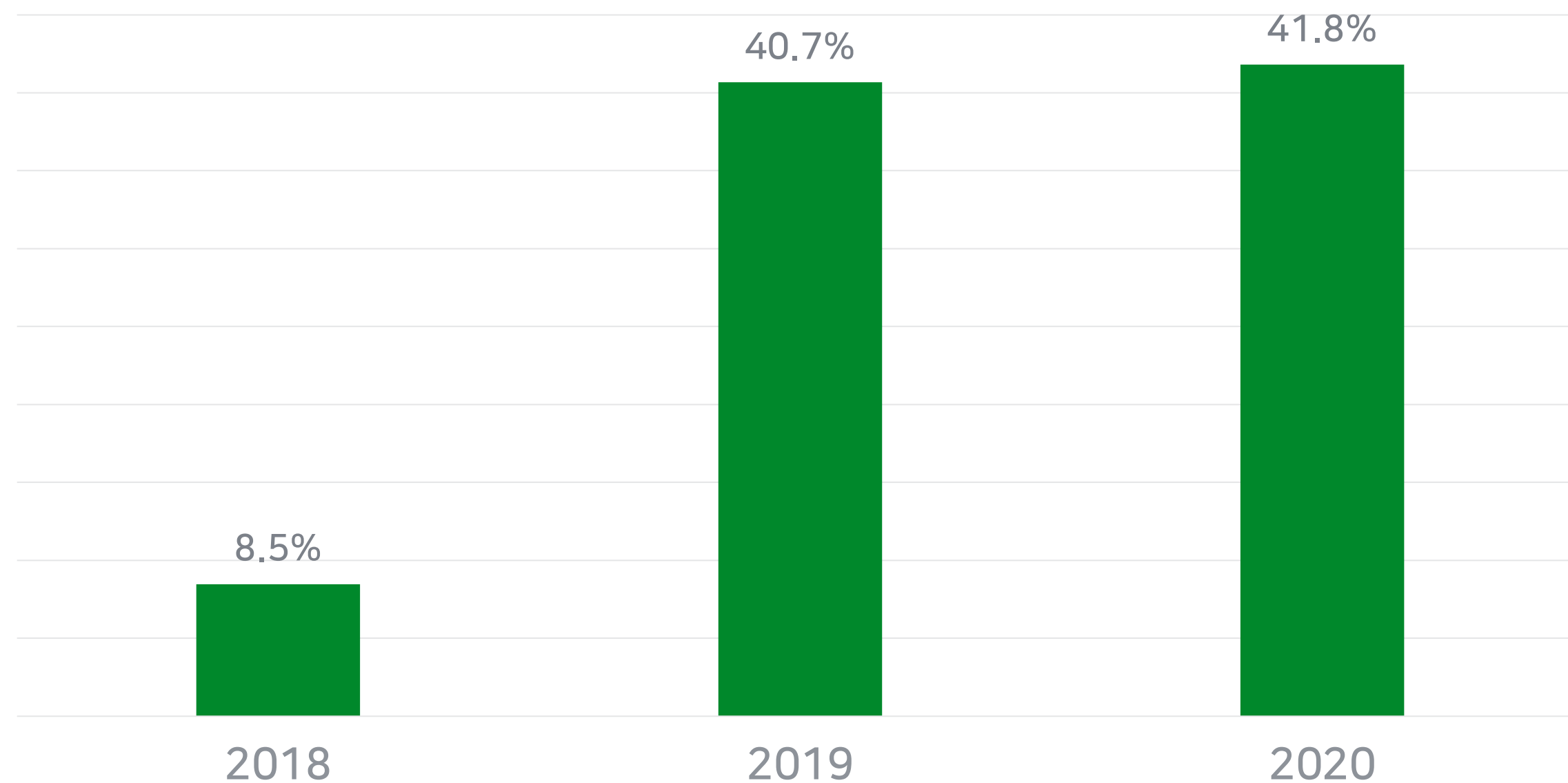
- 사용자의 착각

**studio** tolv → **sudio** tolv

# 1.2 검색 커버리지 개선

쇼핑 Unique Query Count

■ 전년대비증가량 (YoY)



## 점점 다양해지는 검색어 대응 필요

- 검색어 - 상품명과 단어 mismatch

대나무 **이파리** → 대나무 **잎**

- 사용자의 착각

**studio** tolv → **sudio** tolv

- 인물명 - 상품

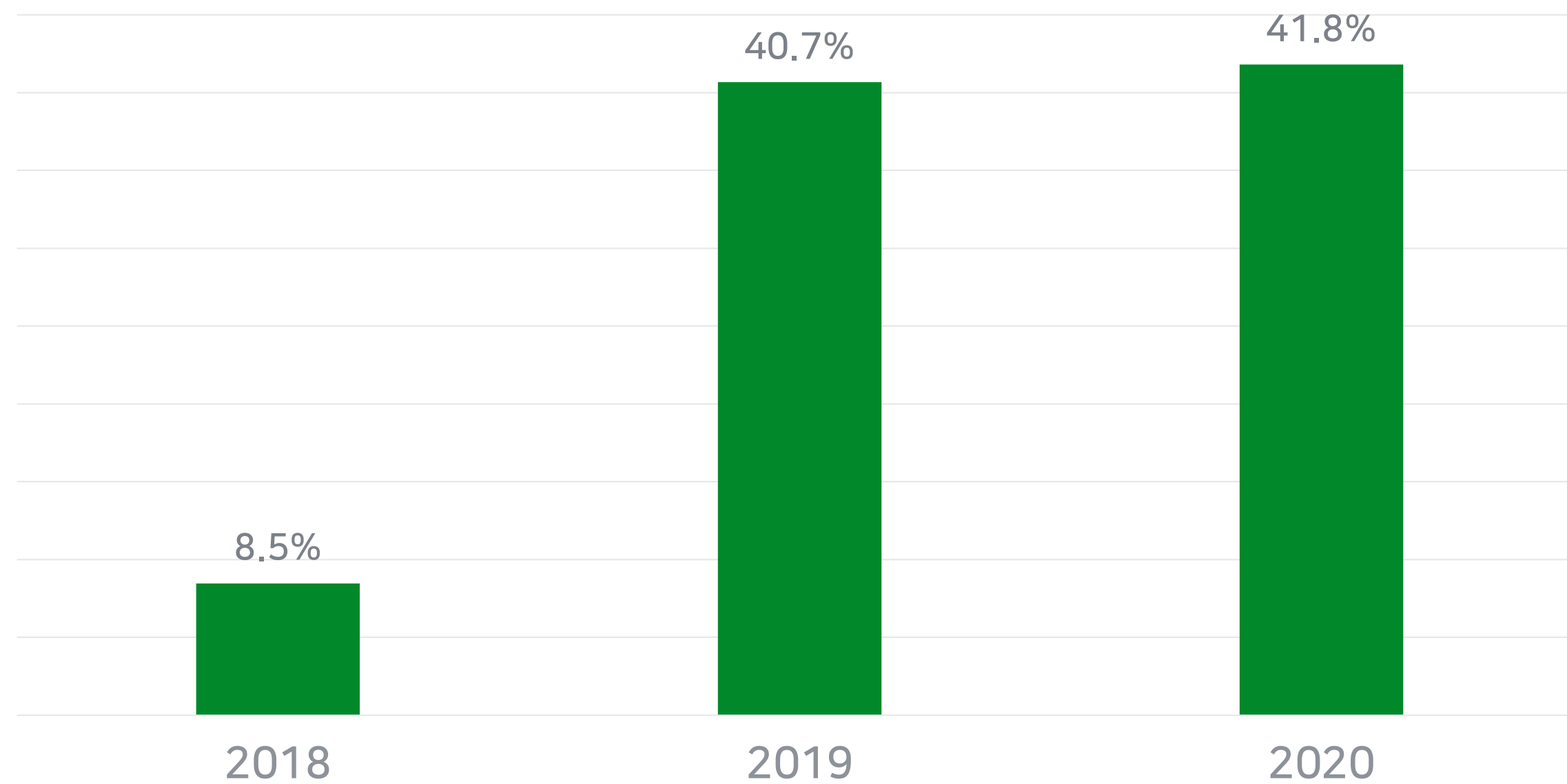
**신세경** 요가복 → **안다르** 요가복



# 1.2 검색 커버리지 개선

쇼핑 Unique Query Count

■ 전년대비증가량 (YoY)



## 점점 다양해지는 검색어 대응 필요

- 검색어 - 상품명과 단어 mismatch

대나무 **이파리** → 대나무 **잎**

- 사용자의 착각

**studio** tolv → **sudio** tolv

- 인물명 - 상품

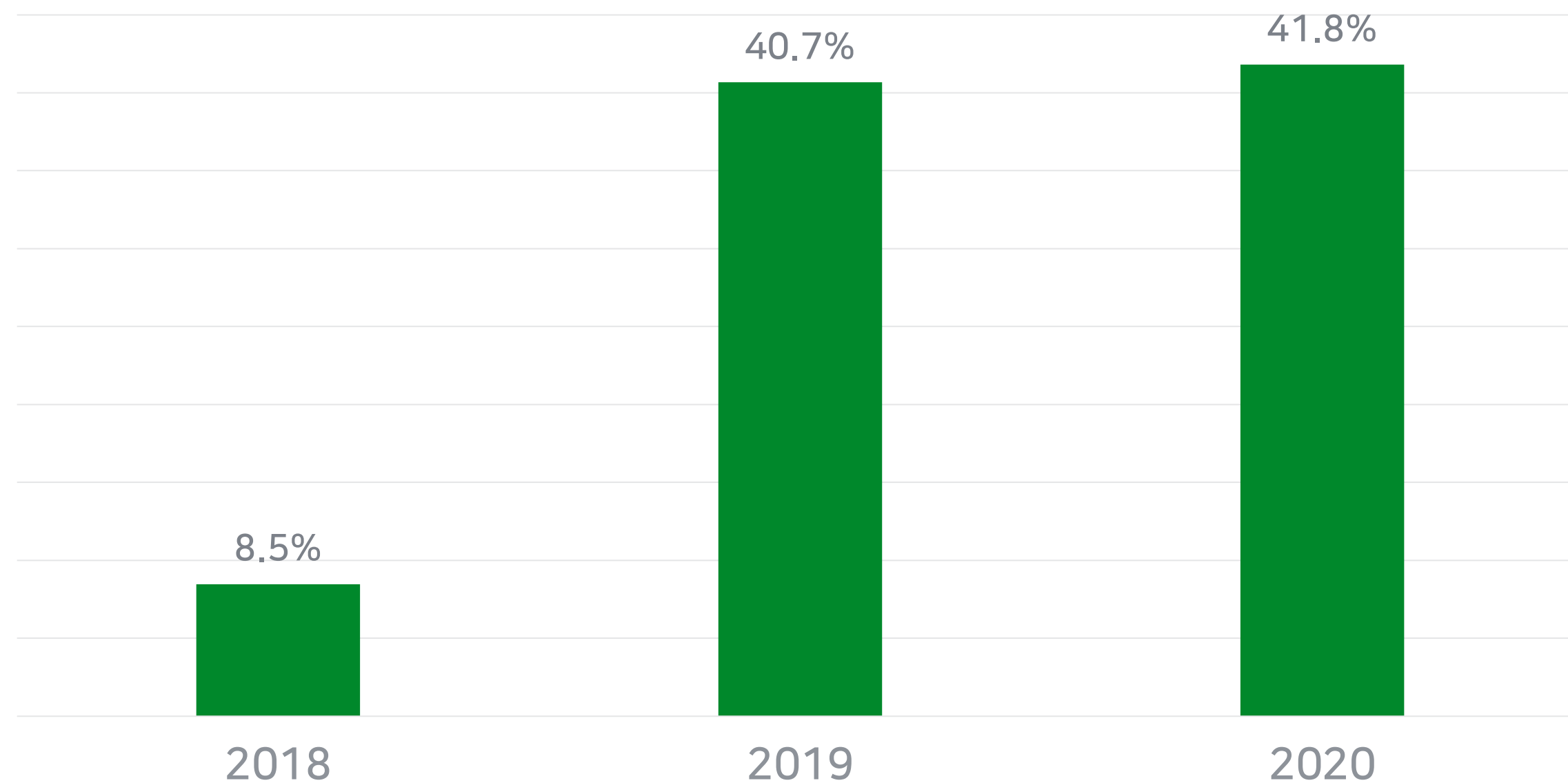
**신세경** 요가복 → **안다르** 요가복

- 자연어 형태의 질의

**뚜껑** 있는 **스테인레스 원통 용기**

# 1.2 검색 커버리지 개선

쇼핑 Unique Query Count  
■ 전년대비증가량 (YoY)



## 점점 다양해지는 검색어 대응 필요

- 검색어 - 상품명과 단어 mismatch  
대나무 **이파리** → 대나무 **잎**

- 사용자의 착각

**studio** tolv → **sudio** tolv

- 인물명 - 상품

**신세경** 요가복 → **안다르** 요가복

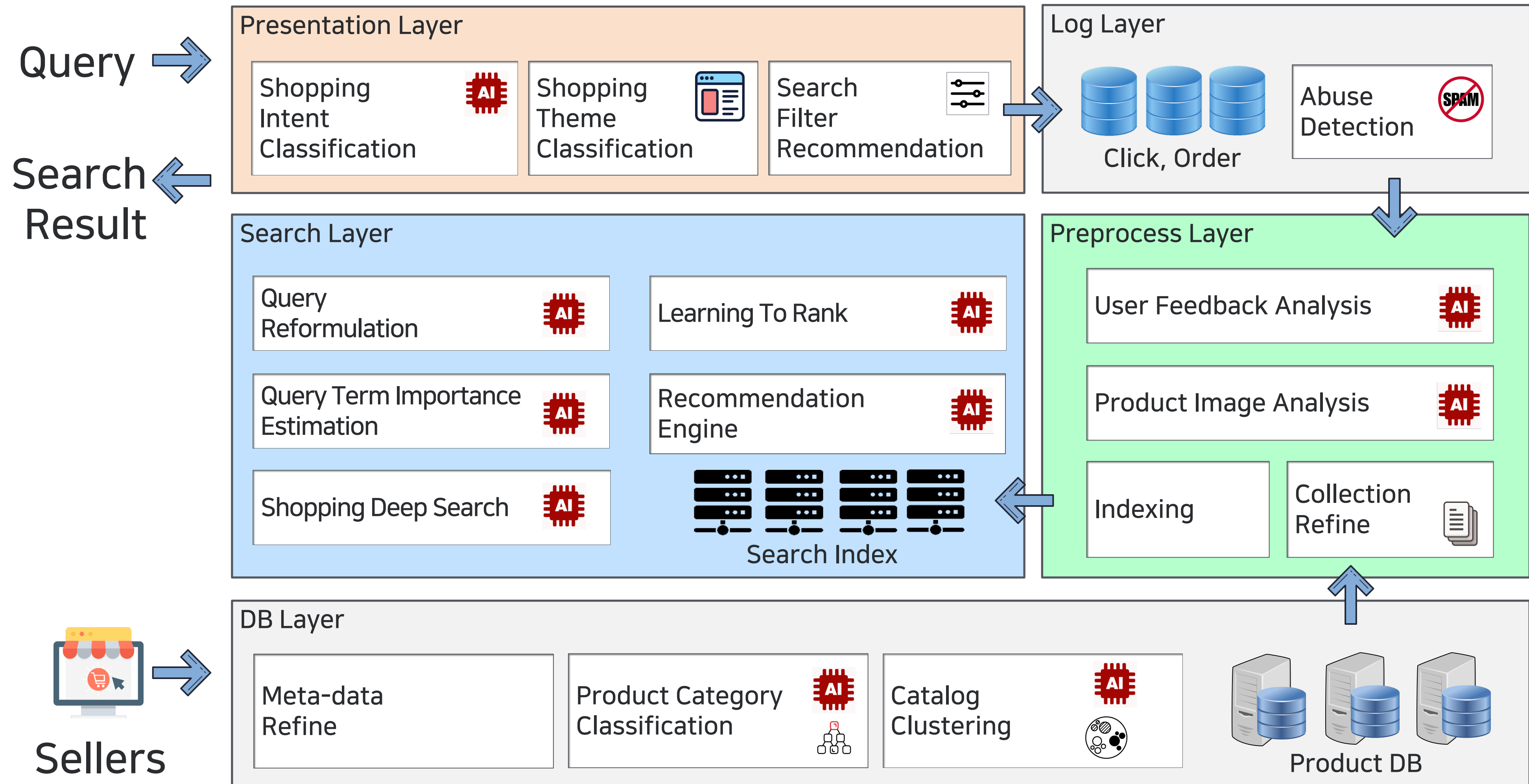
- 자연어 형태의 질의

**뚜껑** 있는 **스테인레스 원통 용기**

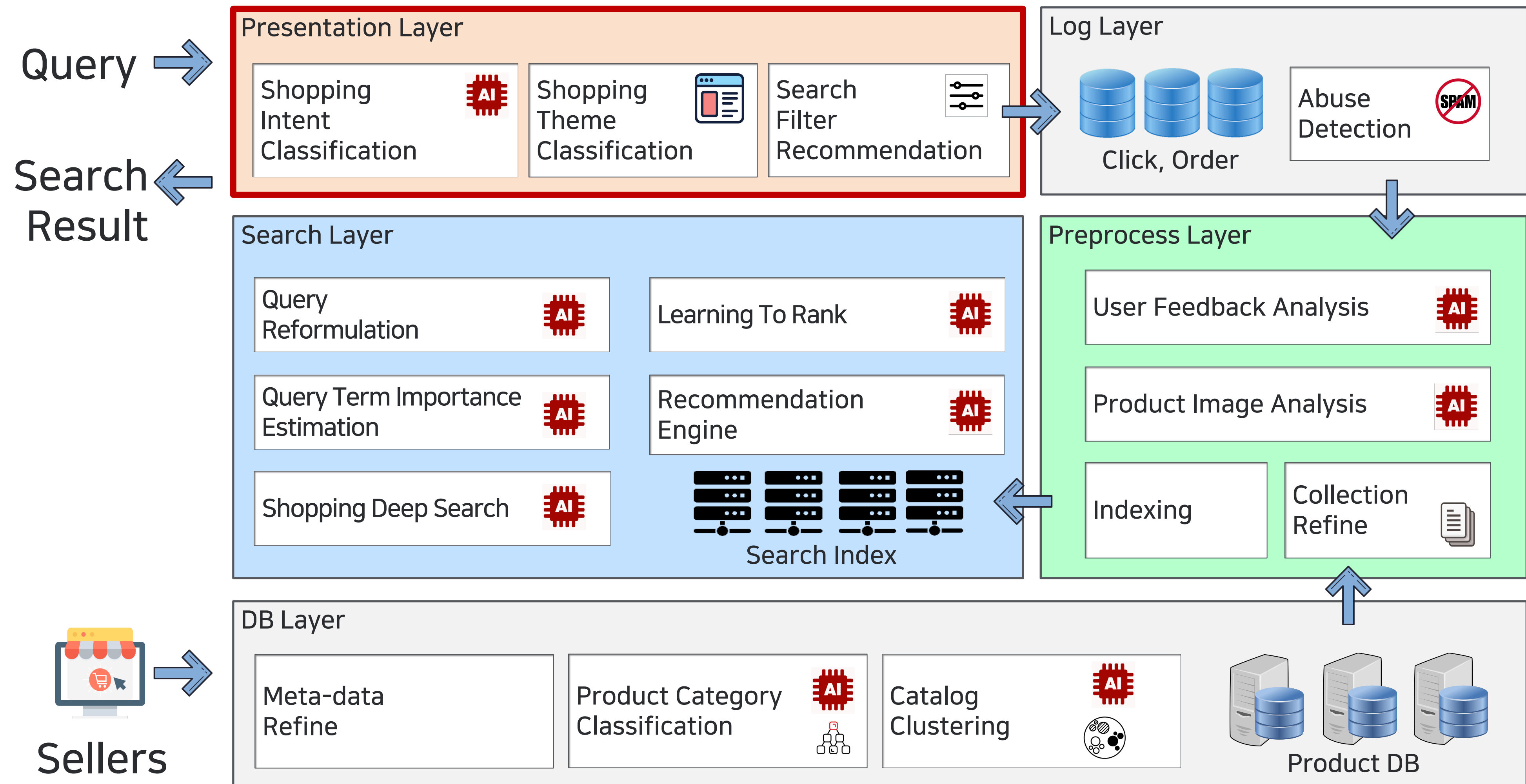
- 복합 의도

**온풍기** 내돈내산

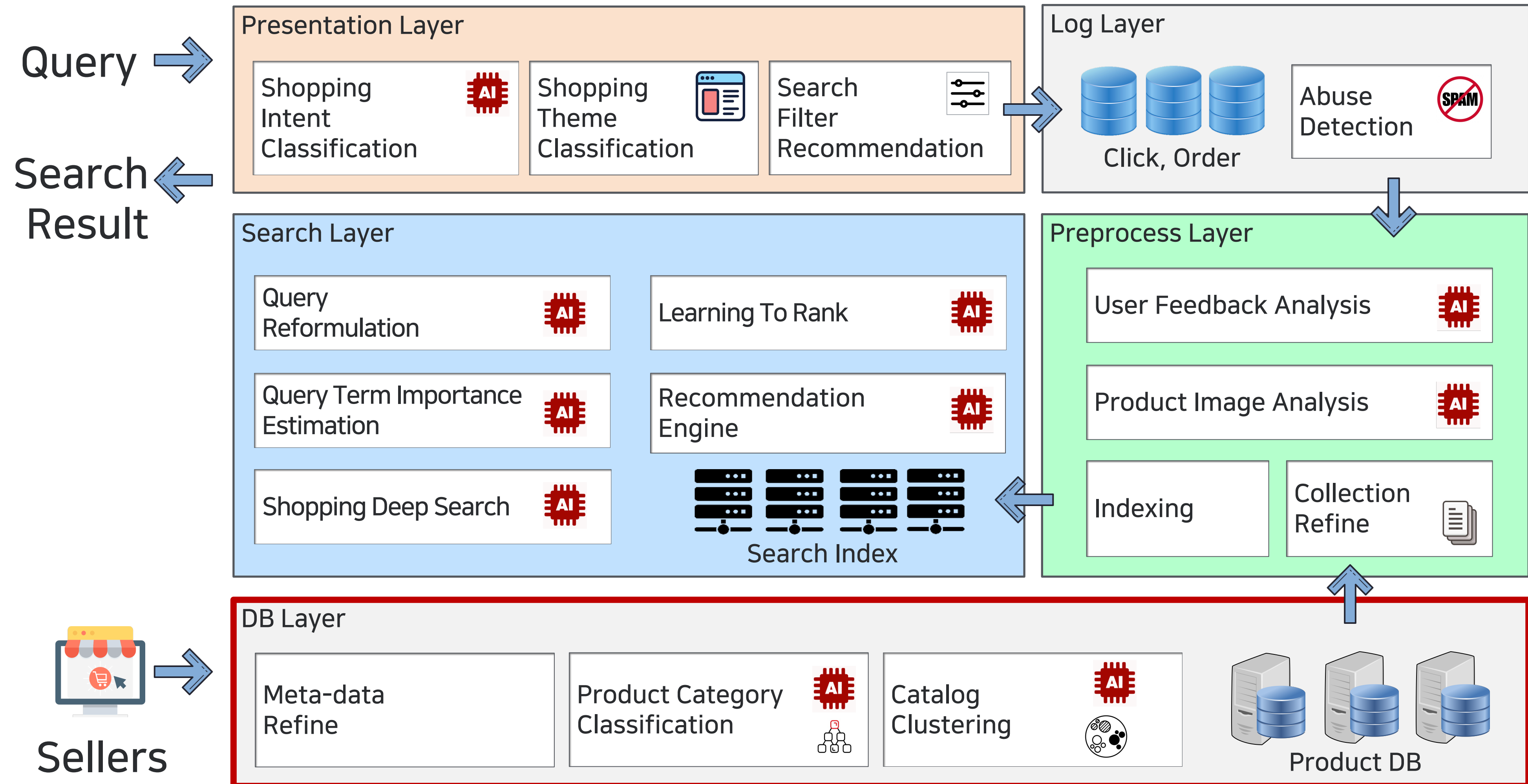
# 1.3 쇼핑 검색 기술 Overview



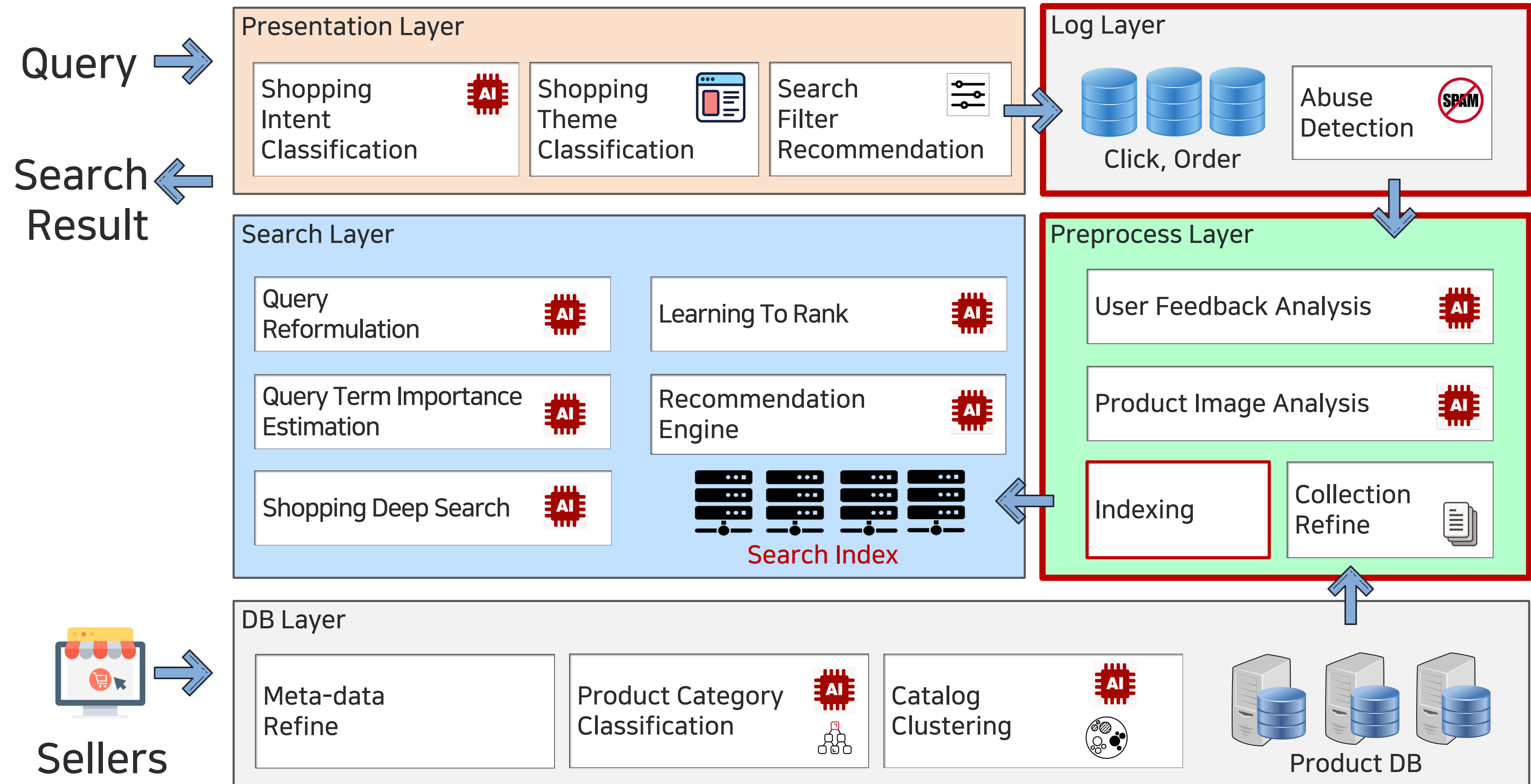
# 1.3 쇼핑 검색 기술 Overview



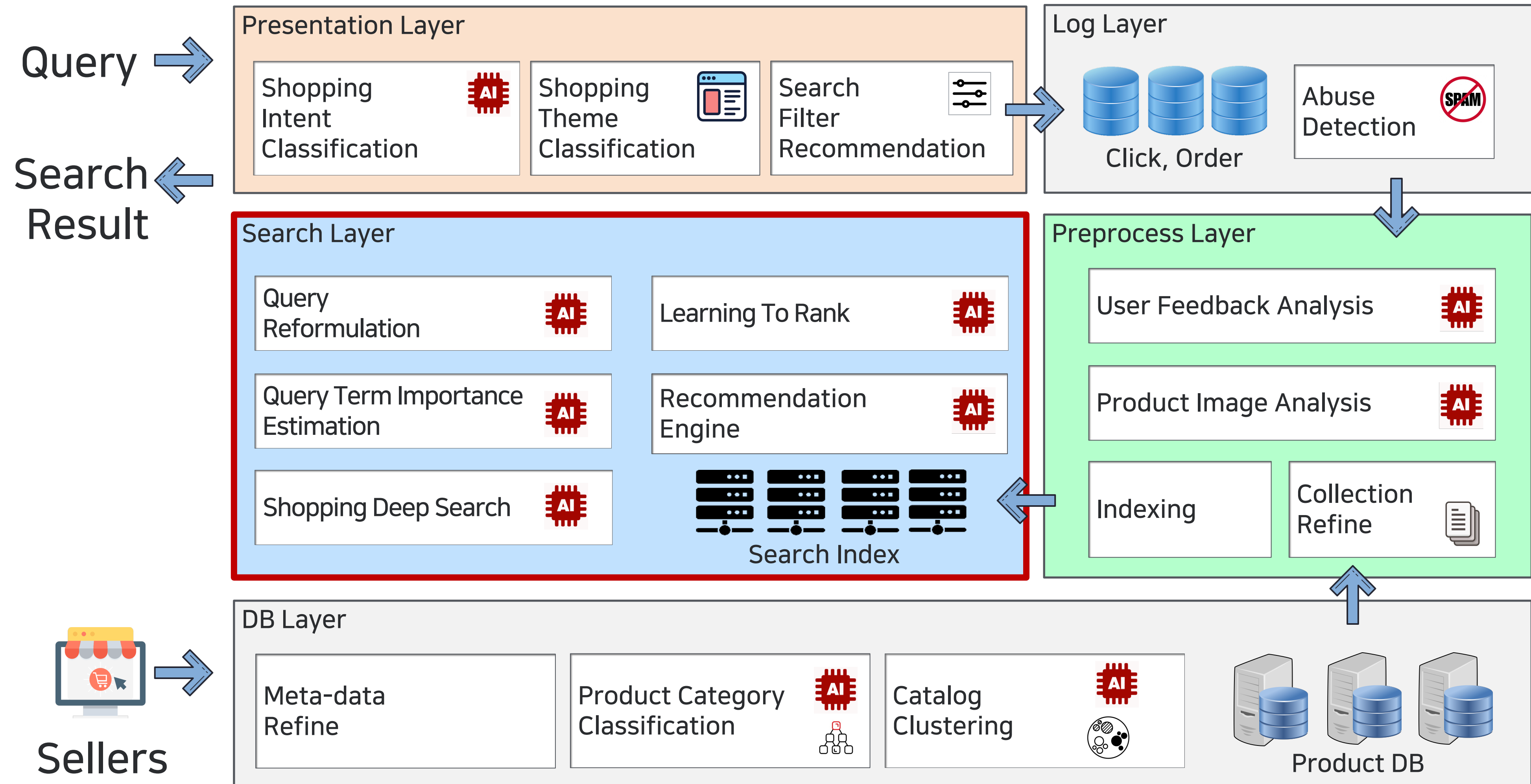
# 1.3 쇼핑 검색 기술 Overview



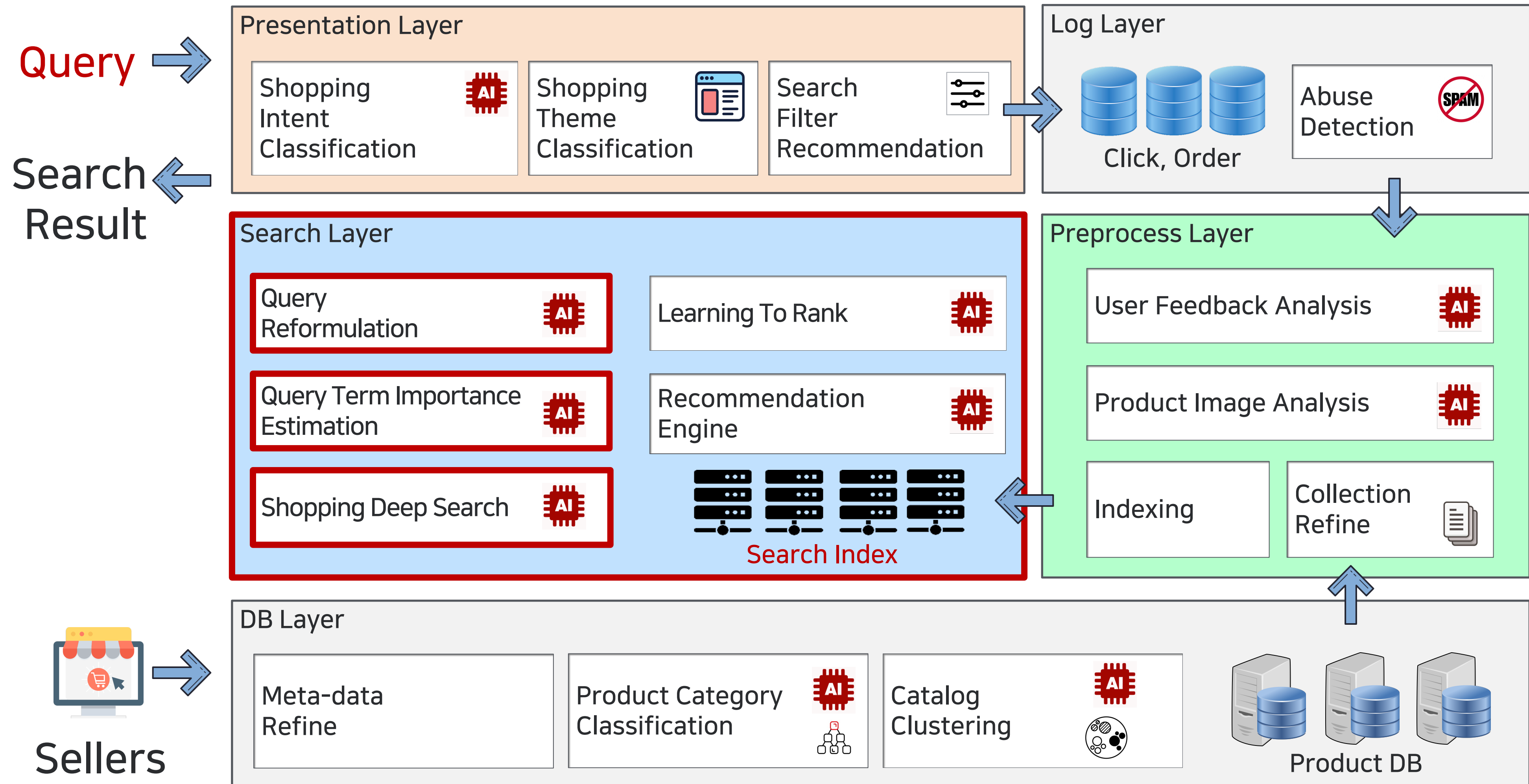
# 1.3 쇼핑 검색 기술 Overview



# 1.3 쇼핑 검색 기술 Overview

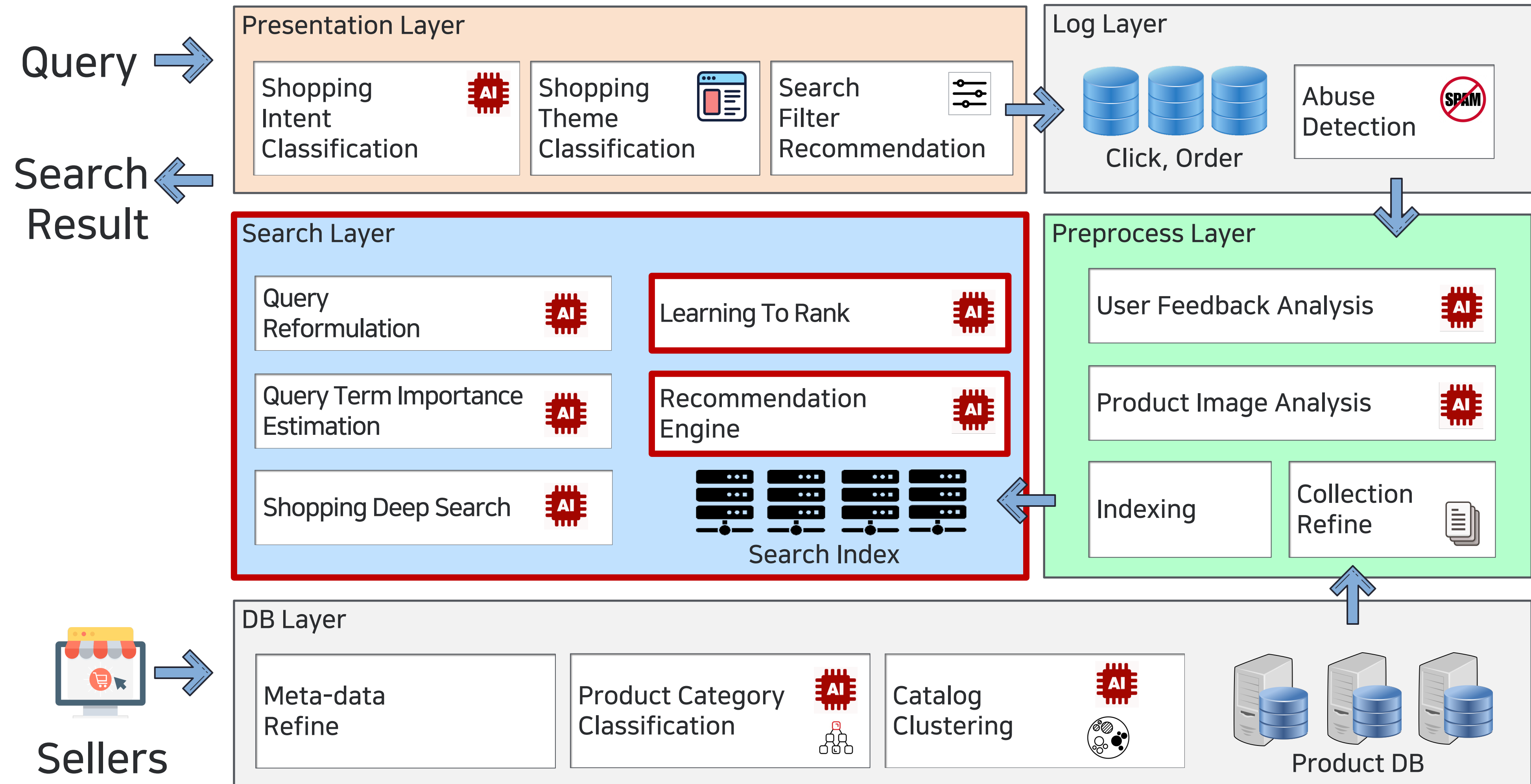


# 1.3 쇼핑 검색 기술 Overview

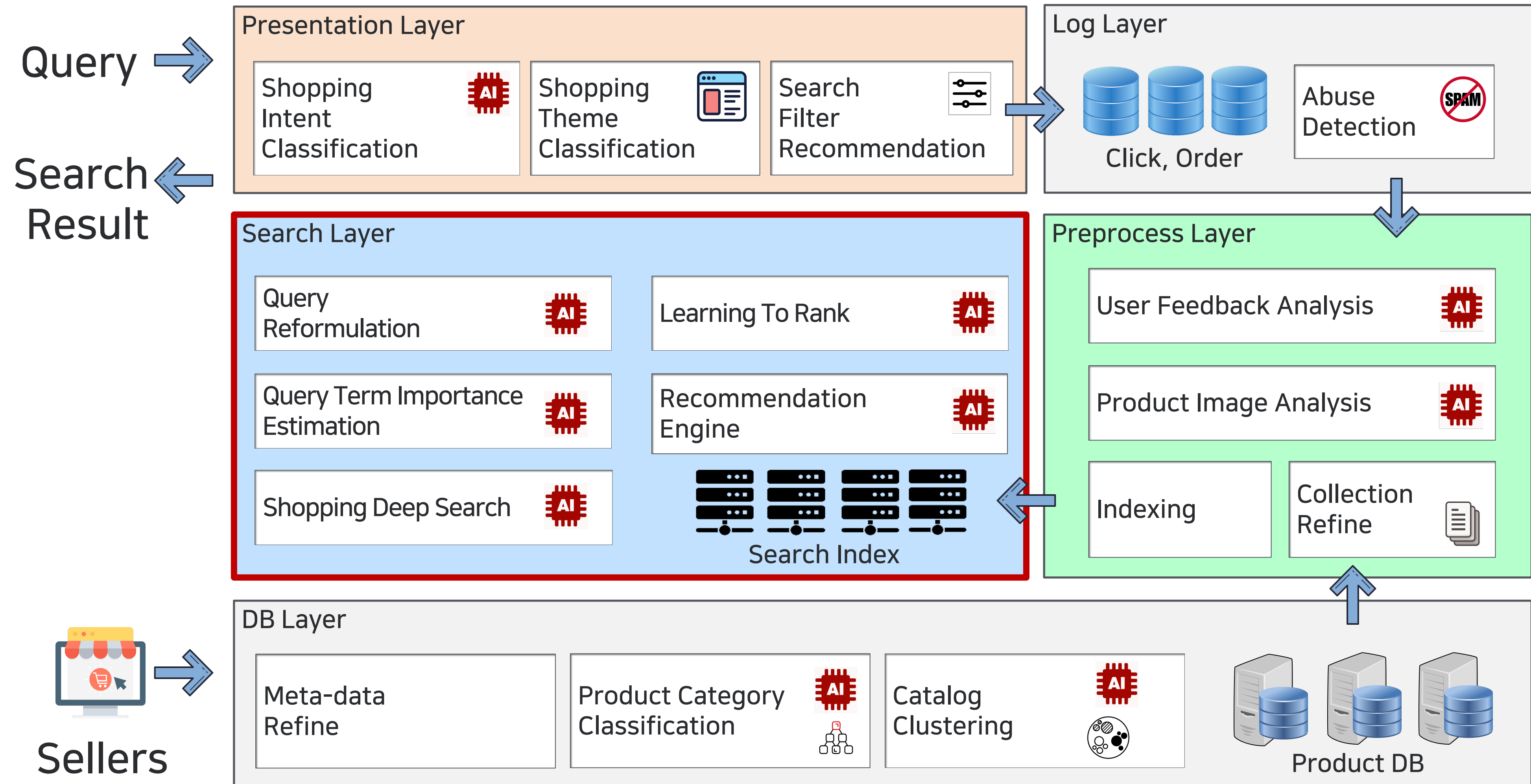




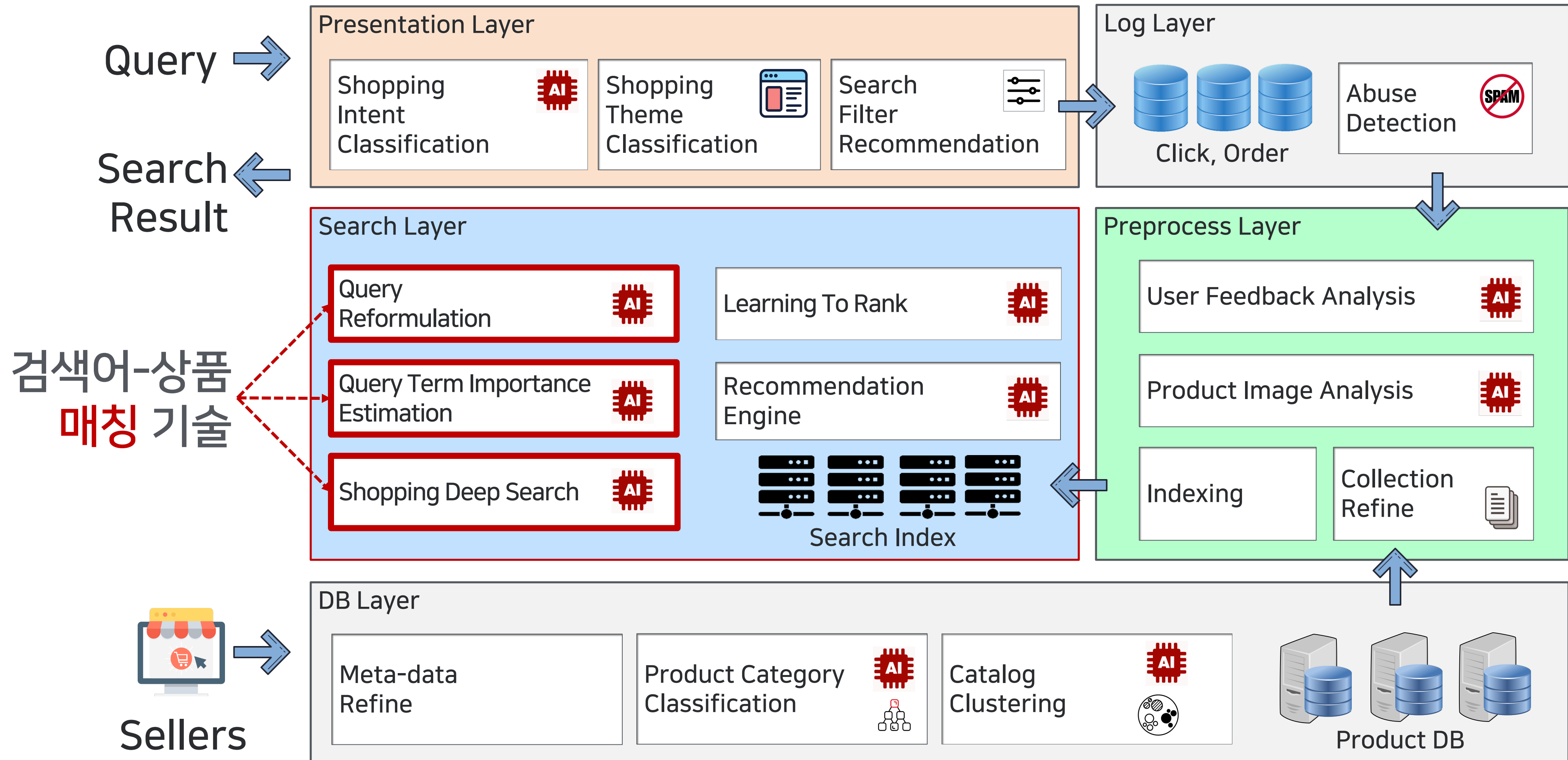
# 1.3 쇼핑 검색 기술 Overview



# 1.3 쇼핑 검색 기술 Overview



# 1.3 쇼핑 검색 기술 Overview



## 2. 검색어 확장 : Query Reformulation

## 2.1 Query Reformulation (QR)

검색어를 더 좋은 쇼핑 검색 결과를 얻을 수 있는 형태로 변환



옥동자 돈까스 ?

## 2.1 Query Reformulation (QR)

검색어를 더 좋은 쇼핑 검색 결과를 얻을 수 있는 형태로 변환

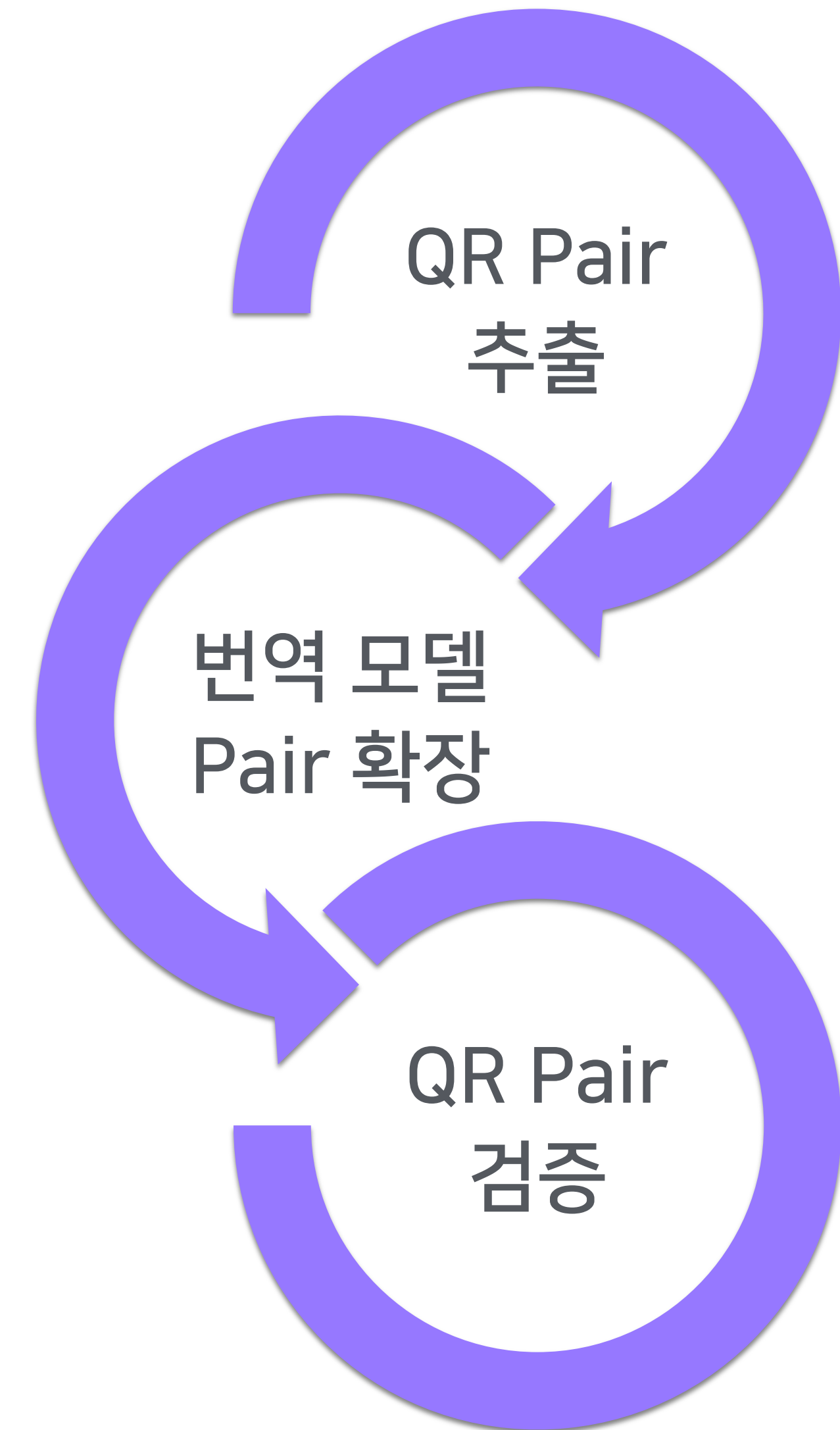


옥동자 돈까스 ? → 옥주부 돈까스 !

# 2.1 Query Reformulation (QR)

## QR Pair (Q1, Q2) 추출 과정

- 사용자 검색 로그를 통한 검색어 QR Pair 추출
- 번역 모델을 이용한 QR Pair 확장
- 사용자 반응을 통한 QR Pair 검증



# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

QR Pair  
추출



# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생

QR Pair  
추출

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생

2. 검색어 유사도 :  $Sim(Q1, Q2)$

- 자소 / 음절 / 쇼핑 entity (브랜드, 카테고리, 속성)

QR Pair  
추출

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생

2. 검색어 유사도 :  $Sim(Q1, Q2)$

- 자소 / 음절 / 쇼핑 entity (브랜드, 카테고리, 속성)

- 검색어 embedding vector

QR Pair  
추출

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

QR Pair  
추출

- 추출 기준
  1. 일정 시간 간격 안에 검색어 입력 발생
  2. 검색어 유사도
  3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$ 
    - 상품명, 쇼핑 검색어의 생성 확률 (language modeling)

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

QR Pair  
추출

- 추출 기준
  1. 일정 시간 간격 안에 검색어 입력 발생
  2. 검색어 유사도
  3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$ 
    - 상품명, 쇼핑 검색어의 생성 확률 (language modeling)
    - $P(\text{"옥동자 돈까스"}) < P(\text{"옥주부 돈까스"})$
    - $P(\text{"날씨"}) < P(\text{"원피스"})$
    - $P(\text{"노투북"}) < P(\text{"노트북"})$

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

- 추출 기준
  1. 일정 시간 간격 안에 검색어 입력 발생
  2. 검색어 유사도
  3. 검색어 생성 확률
  4. 검색 결과 상품 수 :  $C(Q1) < C(Q2)$

QR Pair  
추출

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생
2. 검색어 유사도
3. 검색어 생성 확률
4. 검색 결과 상품 수

**5. 검색 결과 만족도 :  $CTR(Q1) < CTR(Q2)$**

- 전체 세션 로그로 집계한 CC/QC

QR Pair  
추출

# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생
2. 검색어 유사도 :  $Sim(Q1, Q2)$
3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$
4. 검색 결과 상품 수 :  $C(Q1) < C(Q2)$
5. 검색 결과 만족도 :  $CTR(Q1) < CTR(Q2)$

$$\lambda_1 Sim(Q1, Q2) + \lambda_2 (P(Q2) - P(Q1)) + \lambda_3 (C(Q2) - C(Q1)) + \lambda_4 (CTR(Q2) - CTR(Q1)) \geq threshold$$

QR Pair  
추출



# 2.2 QR Pair(Q1,Q2) 추출

예) Q1: 옥동자 돈까스 → Q2: 옥주부 돈까스

검색 세션  
(노출/클릭) 로그

Session Id	Query	Expose items	Click items	Date Time	Duration
78x7WScQ	옥동자 돈까스	0	0	2021-11-11 13:00:02	
78x7WScQ	옥주부 돈까스	20	2	2021-11-11 13:00:05	3s

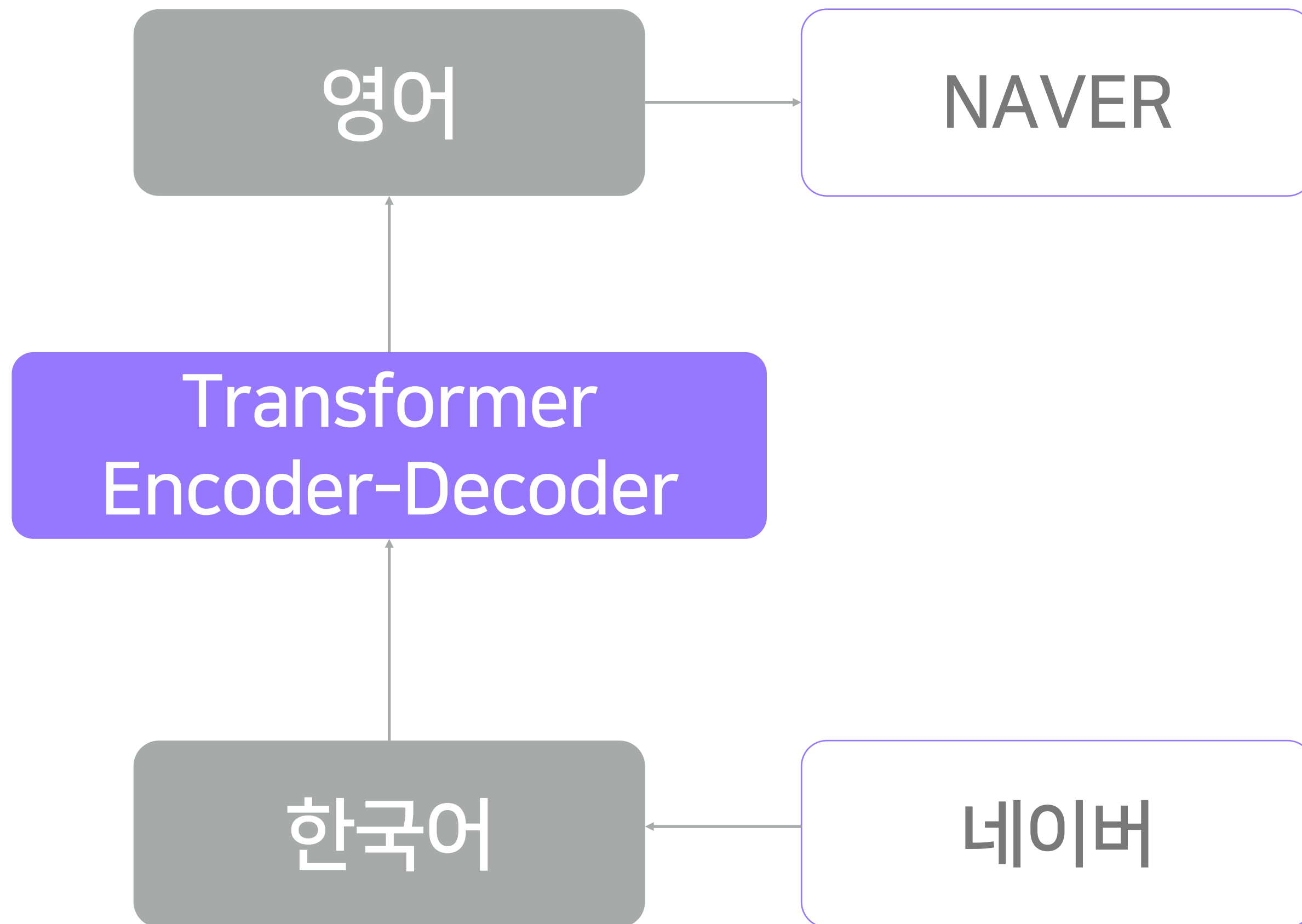
- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생
2. 검색어 유사도 :  $Sim(Q1, Q2)$
3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$
4. 검색 결과 상품 수 :  $C(Q1) < C(Q2)$
5. 검색 결과 만족도 :  $CTR(Q1) < CTR(Q2)$

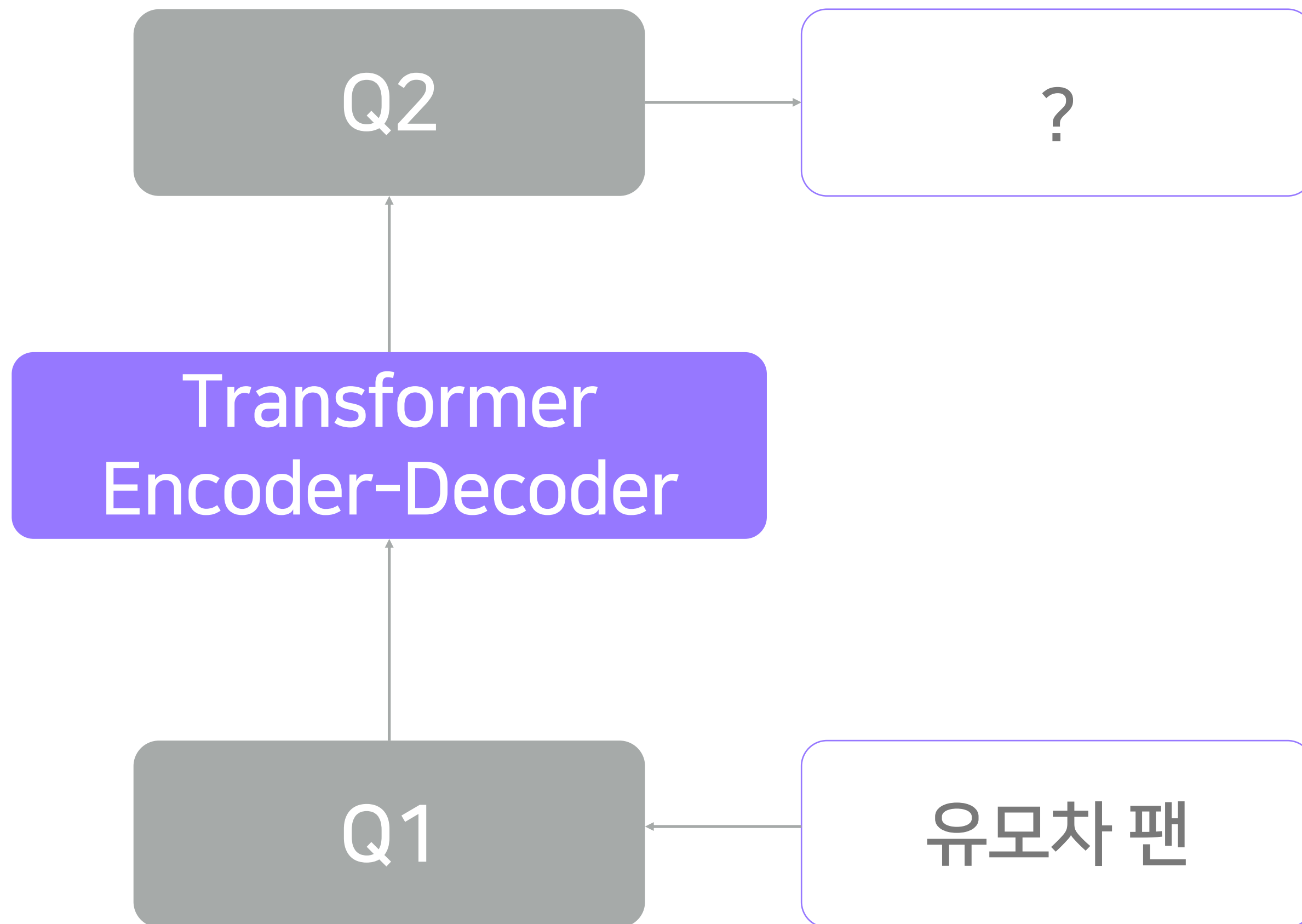
$$\lambda_1 Sim(Q1, Q2) + \lambda_2 (P(Q2) - P(Q1)) + \lambda_3 (C(Q2) - C(Q1)) + \lambda_4 (CTR(Q2) - CTR(Q1)) \geq threshold$$

QR Pair  
추출

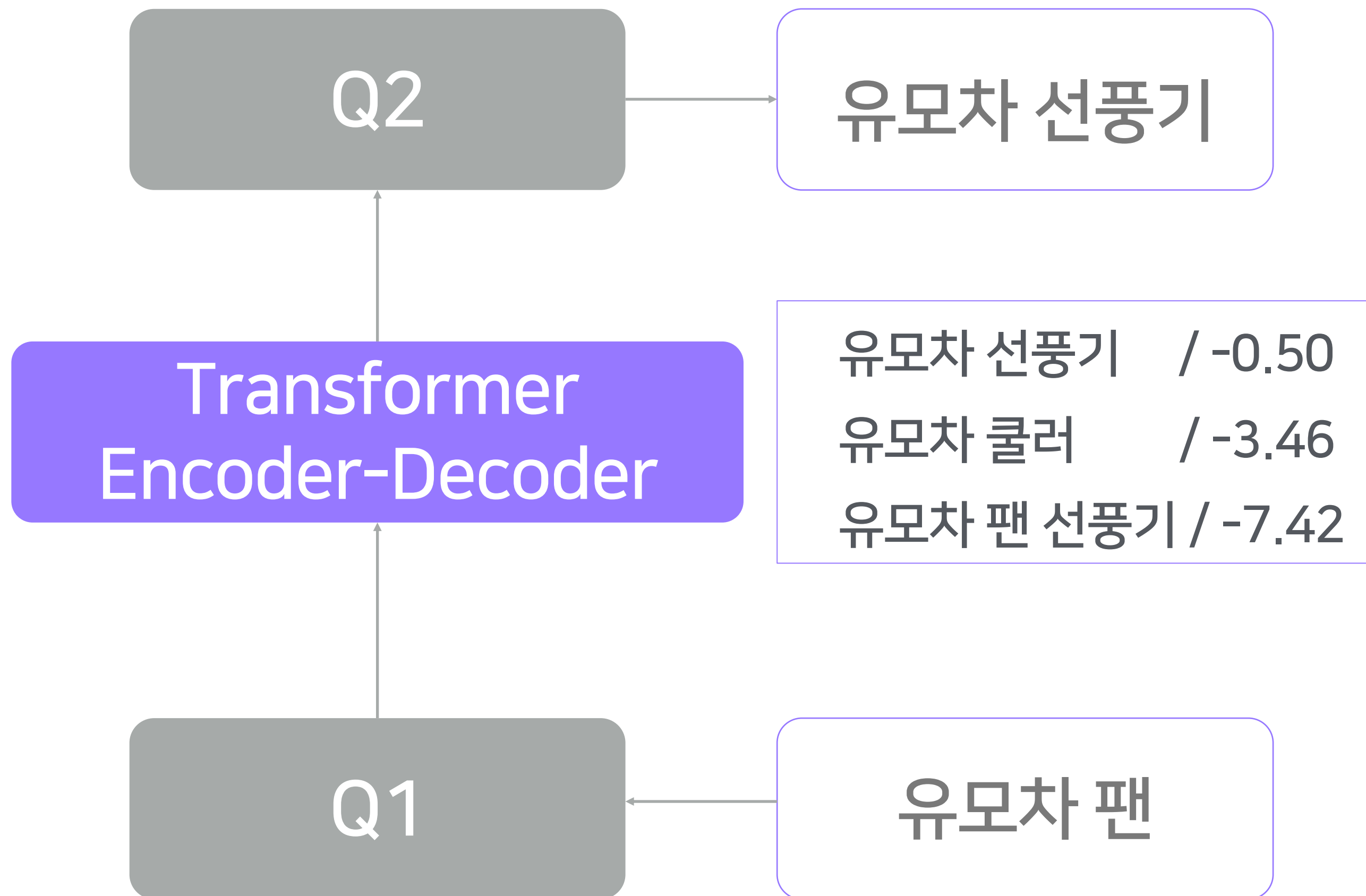
## 2.3 번역 모델을 이용한 QR Pair 확장



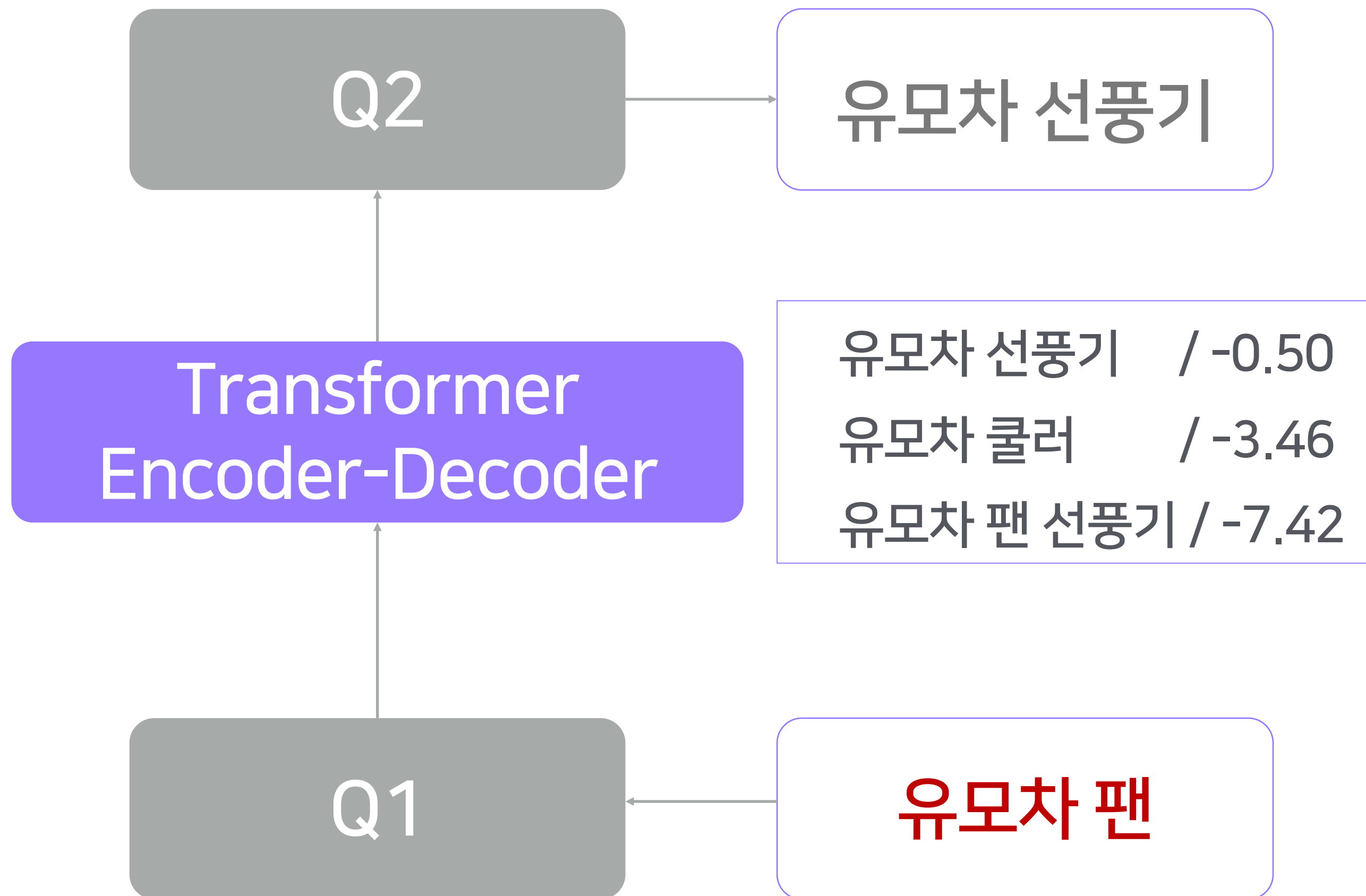
## 2.3 번역 모델을 이용한 QR Pair 확장



# 2.3 번역 모델을 이용한 QR Pair 확장



# 2.3 번역 모델을 이용한 QR Pair 확장



'유모차 팬'에 대한 검색 결과입니다.

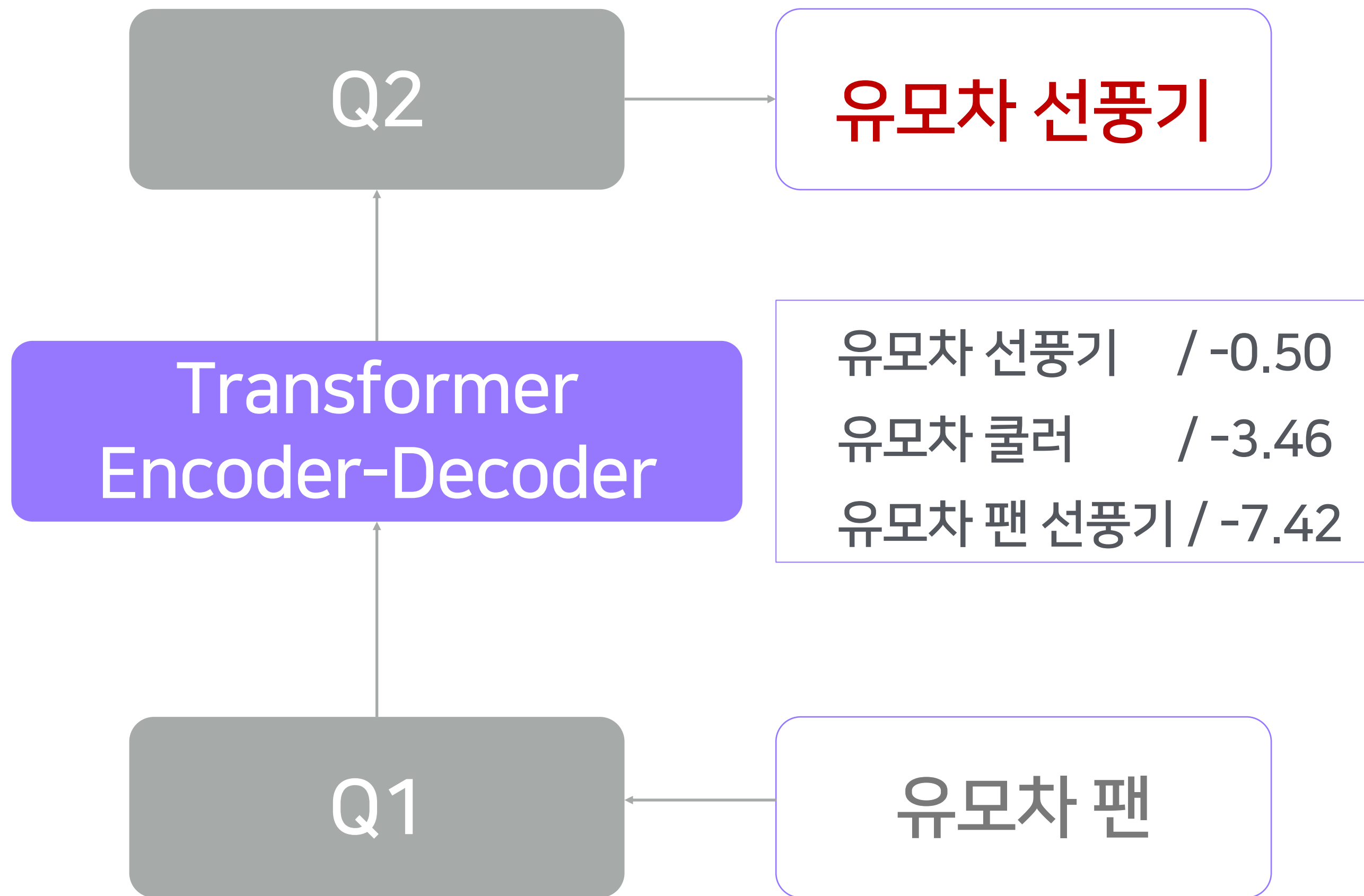
전체	가격비교	네이버페이
8,056	53	3,813

네이버 랭킹순 · 낮은 가격순 · 높은 가격순 · 등록일순 · 리뷰 많은순

휴대용 미니 클립 유모차 팬 유연한 굴절식 충전식  
**최저 32,770원** 판매처 3  
 출산/육아 > 유모차 > 유모차용품 > 기타유모차용품  
 등록일 2021.09. · 찜하기 0 · 정보 수정요청

해의 124568 배터리로 작동되는 유모차 팬 3단 속도 회  
**최저 65,700원** 판매처 4  
 출산/육아 > 유모차 > 유모차용품 > 기타유모차용품  
 등록일 2021.08. · 찜하기 0 · 정보 수정요청

# 2.3 번역 모델을 이용한 QR Pair 확장



'유모차 선풍기'에 대한 검색 결과입니다.

전체	가격비교	네이버페이	백화점/홈쇼핑
38,703	385	13,769	5,144

네이버 랭킹순 · 낮은 가격순 · 높은 가격순 · 등록일순 · 리뷰 많은순 · 쇼핑물선택

**베리어 휴대용 유모차 선풍기**

최저 15,700원 판매처 127

출산/육아 > 유모차 > 유모차용품 > 기타유모차용품

리뷰 ★★★★★ 817 · 등록일 2021.03. · 찜하기 88 · 정보 수정요청

**엘츠 휴대용 미니 유모차 선풍기 4종컬러**

최저 25,000원 판매처 11

출산/육아 > 유모차 > 유모차용품 > 기타유모차용품

리뷰 ★★★★★ 673 · 등록일 2019.03. · 찜하기 26 · 정보 수정요청

**휴대용 미니 클립 유모차 팬 유연한 굴절식 충전식**

최저 32,770원 판매처 3

출산/육아 > 유모차 > 유모차용품 > 기타유모차용품

등록일 2021.09. · 찜하기 0 · 정보 수정요청

**해외 124568 배터리로 작동되는 유모차 팬 3단 속도 회**

최저 65,700원 판매처 4

출산/육아 > 유모차 > 유모차용품 > 기타유모차용품

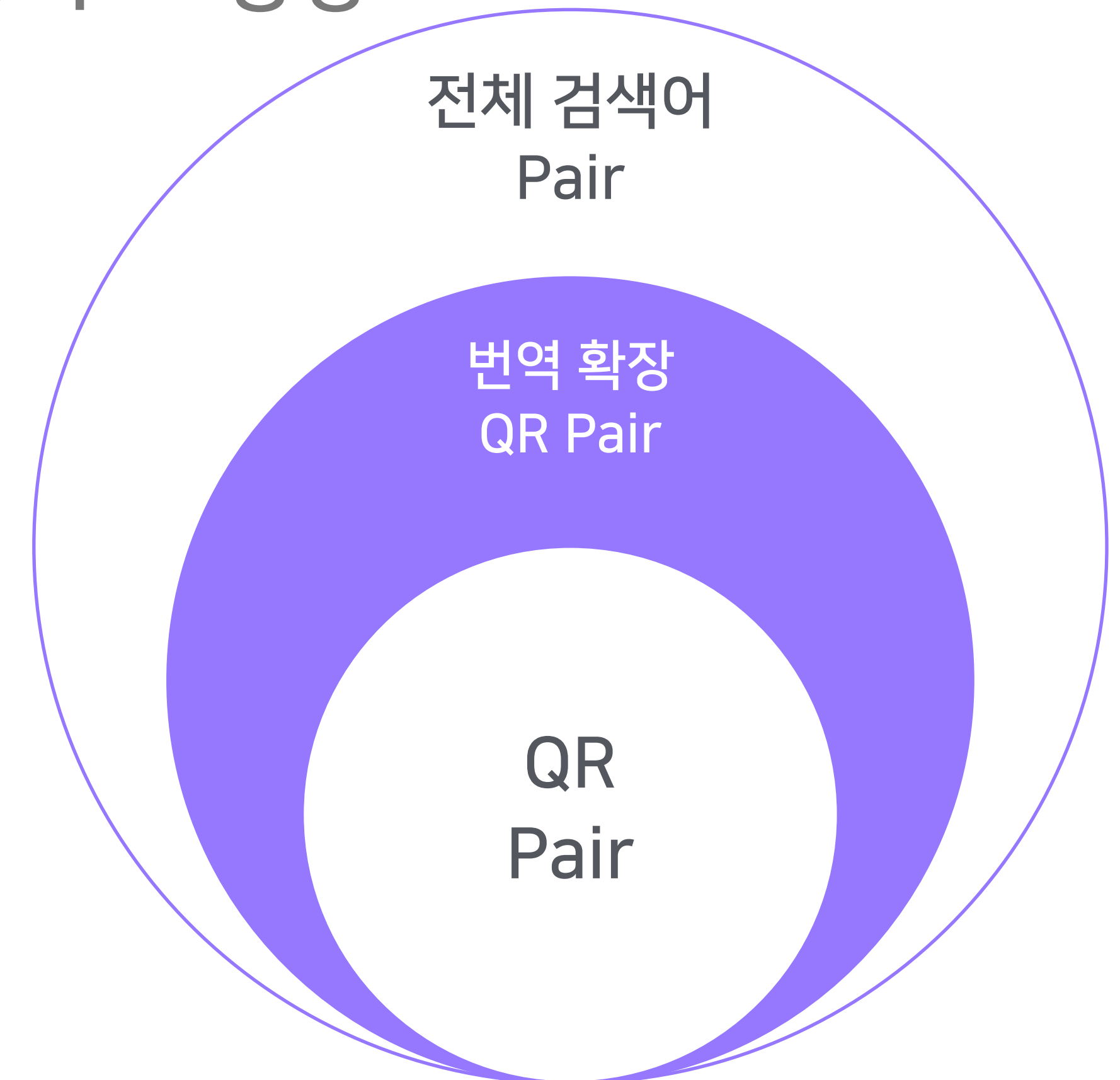
등록일 2021.08. · 찜하기 0 · 정보 수정요청

## 2.3 번역 모델을 이용한 QR Pair 확장

- 학습 데이터 : 2.2 에서 추출한 QR Pair (Q1, Q2)
- 2.2 에서 추출되지 않은 Low CTR 검색어의 QR Pair 후보 생성
- 연속해서 입력된 이력이 있는 검색어 Pair 로 필터링

### - 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생
2. 검색어 유사도 :  $Sim(Q1, Q2)$
3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$
4. 검색 결과 상품 수 :  $C(Q1) < C(Q2)$
5. 검색 결과 만족도 :  $CTR(Q1) < CTR(Q2)$



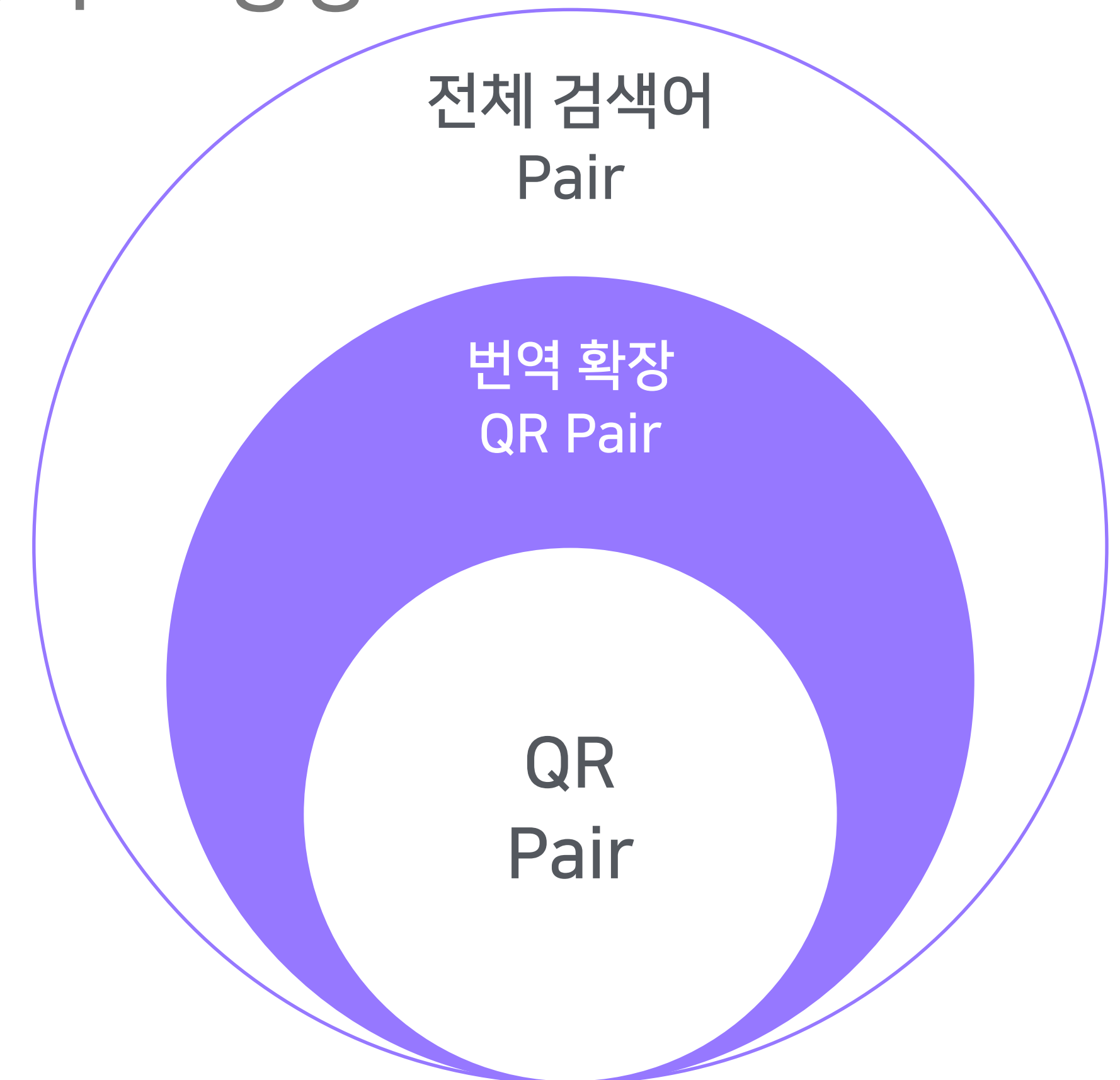
## 2.3 번역 모델을 이용한 QR Pair 확장

- 학습 데이터 : 2.2 에서 추출한 QR Pair (Q1, Q2)
- 2.2 에서 추출되지 않은 Low CTR 검색어의 QR Pair 후보 생성
- 연속해서 입력된 이력이 있는 검색어 Pair 로 필터링

- 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생
2. 검색어 유사도 :  $Sim(Q1, Q2)$
3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$
4. 검색 결과 상품 수 :  $C(Q1) < C(Q2)$
5. 검색 결과 만족도 :  $CTR(Q1) < CTR(Q2)$

→ 15% QR Pair 커버리지 증가





## 2.4 사용자 피드백을 통한 검증

사용자의 피드백을 통해 QR Pair 사후 검증

- QR 적용 전/후의 CTR 비교

## 2.4 사용자 피드백을 통한 검증

### 사용자의 피드백을 통해 QR Pair 사후 검증

- QR 적용 전/후의 CTR 비교
  - '<기존 검색어>로만 검색하기' 클릭 비율
- ⇒ [ positive pair, negative pair ]

N 옥동자 돈까스 ✕ 🔊

N 네이버쇼핑 ⓘ 다른 사이트 더보기

연관된 키워드로 검색한 결과를 포함합니다.  
'옥동자 돈까스'로만 검색하기 >

전체
3만원이하
3만~6만원
6만원이상

옥주부돈까스 제주돼지통까스 등심6개  
고구마3개 치즈3개 소스

최저 35,750원

판매처 9

★ 4.6 (1,063) | 찜 87

만족도 포장 , 배송도 좋았습니다

ⓘ

맛 저번에 홈쇼핑으로 한 번 사 먹고 나서 너무 맛있어서

# 2.4 사용자 피드백을 통한 검증

## 사용자의 피드백을 통해 QR Pair 사후 검증

- QR 적용 전/후의 CTR 비교
- '<기존 검색어>로만 검색하기' 클릭 비율
- ⇒ [ positive pair, negative pair ]
- QR Pair 추출 자동화 및 고도화 활용

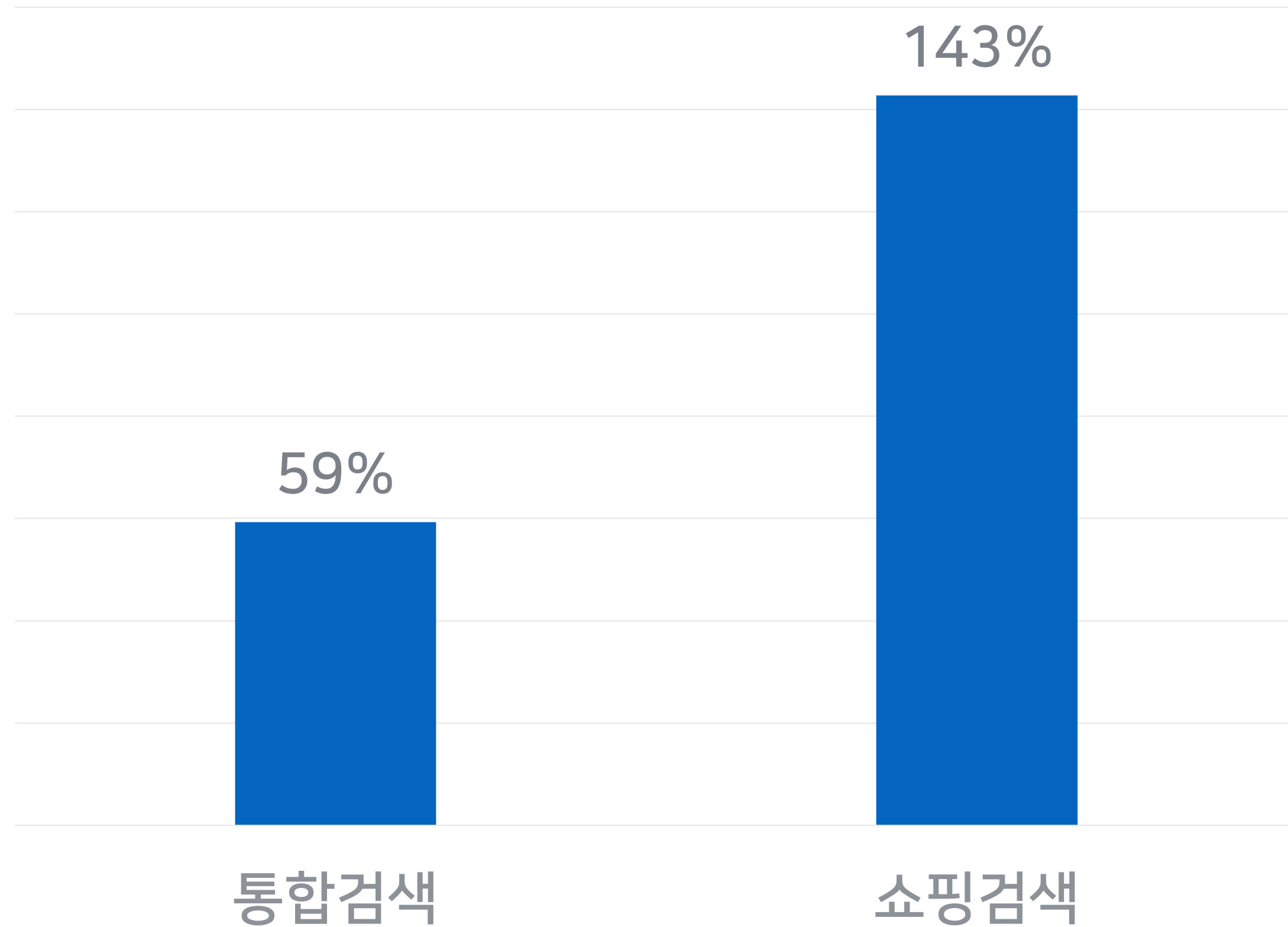
### - 추출 기준

1. 일정 시간 간격 안에 검색어 입력 발생
2. 검색어 유사도 :  $Sim(Q1, Q2)$
3. 검색어 생성 확률 :  $P(Q1) < P(Q2)$
4. 검색 결과 상품 수 :  $C(Q1) < C(Q2)$
5. 검색 결과 만족도 :  $CTR(Q1) < CTR(Q2)$

$$\lambda_1 Sim(Q1, Q2) + \lambda_2 (P(Q2) - P(Q1)) + \lambda_3 (C(Q2) - C(Q1)) + \lambda_4 (CTR(Q2) - CTR(Q1)) \geq threshold$$

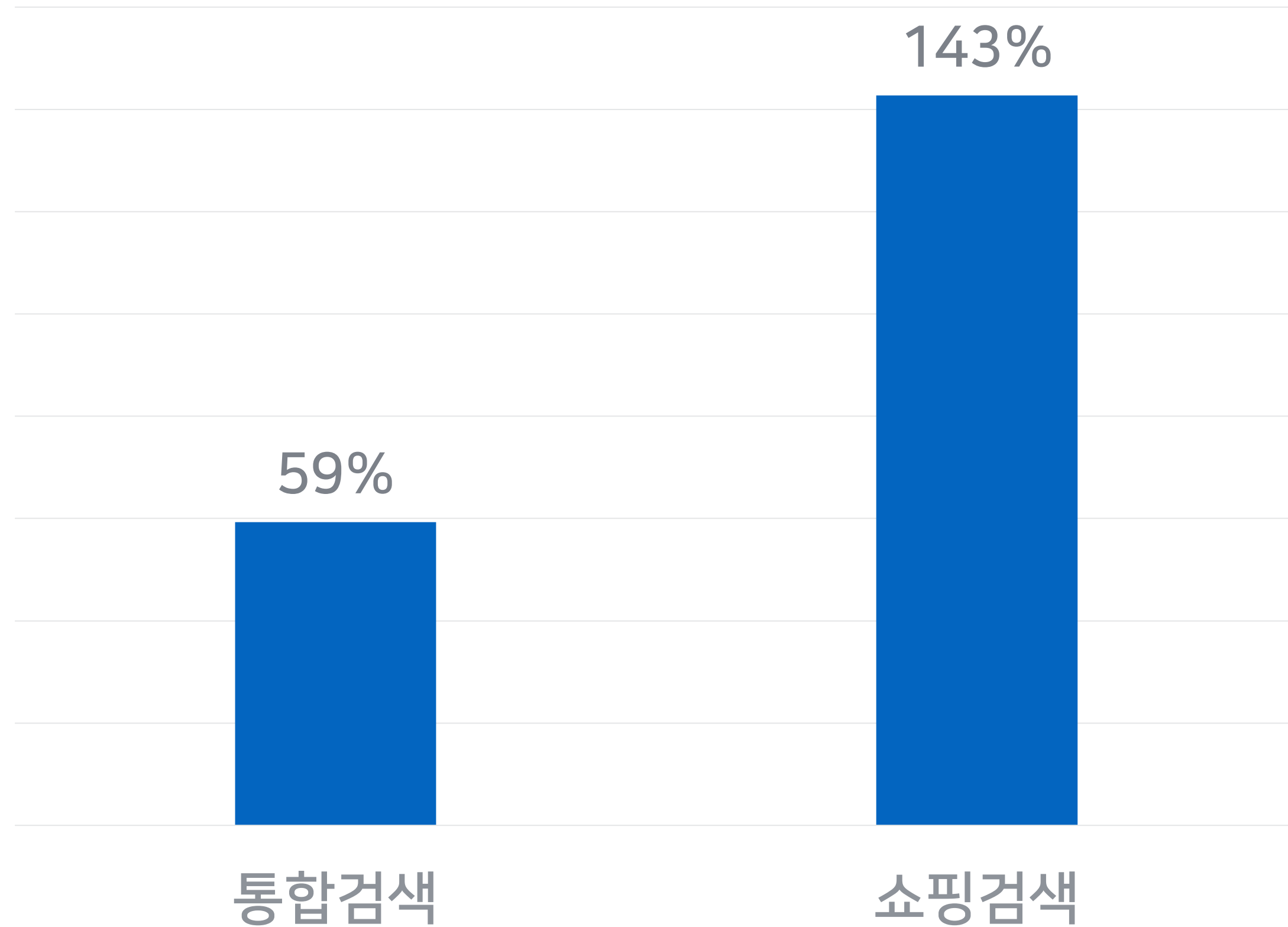
## 2.5 QR 적용 효과

QR 적용 전/후 CTR 개선 (%)

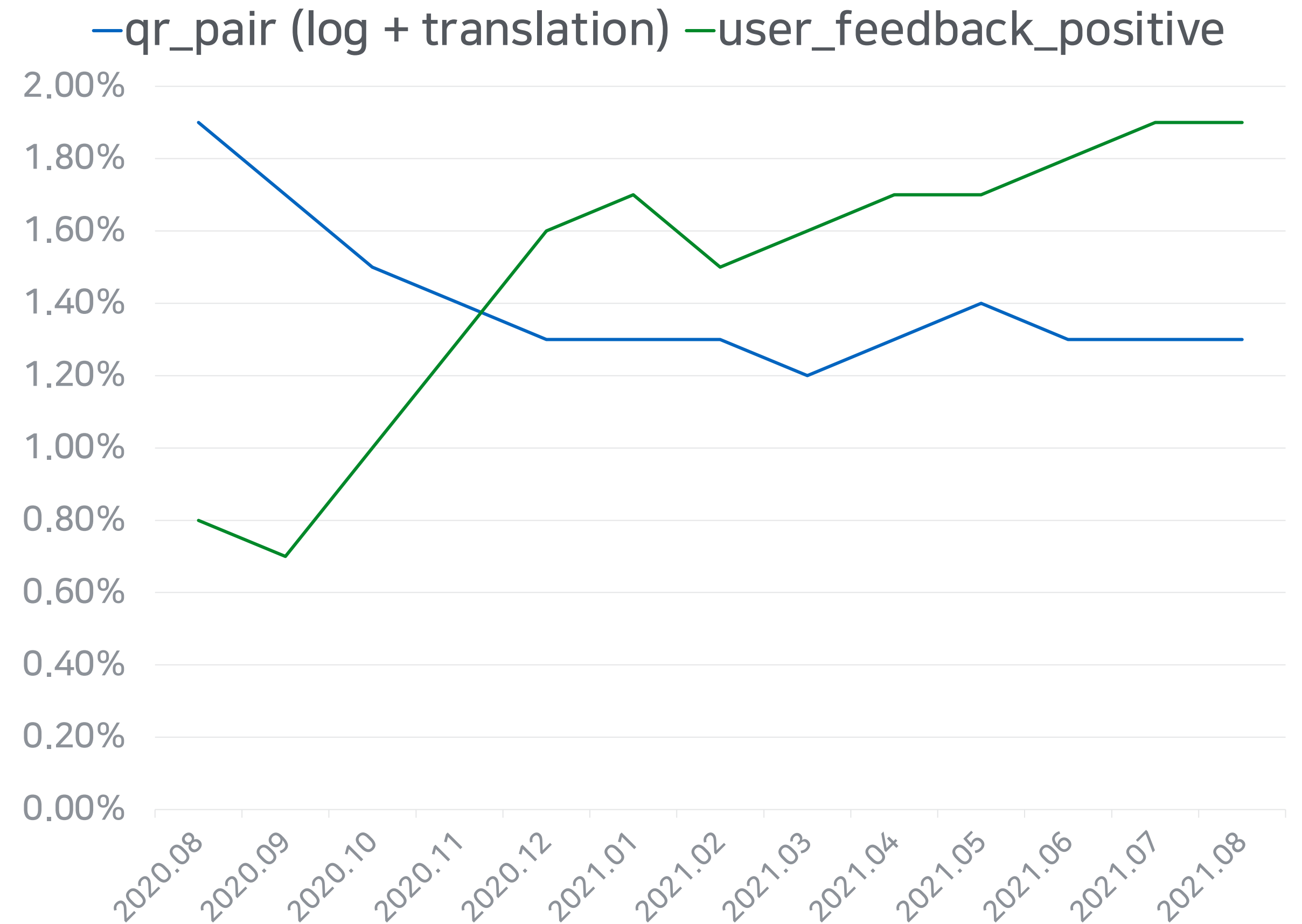


# 2.5 QR 적용 효과

## QR 적용 전/후 CTR 개선 (%)



## Type 별 UQC Ratio



# 3. 검색어 부분 일치 : Query Term Importance

# 3.1 검색어 부분 일치

## 검색 결과에 반드시 포함되어야 하는 중요 단어 선별

- 문장형 검색어

**톰포드 패블러스** 판매하는 곳

Query : **OR** ("판매", "곳")

**AND** ("톰포드", "패블러스")

**N** 톰포드 패블러스 판매하는 곳
✕

**N** 네이버쇼핑 ⓘ
다른 사이트 더보기

	향수종류	가격
<b>전체</b>	남녀공용향수	여성향수
남성향수		

**톰포드 패블러스 오 드 퍼퓸**

최저 **231,840원**

판매처 17

★ 5.0 (49) | 찜 64

# 3.1 검색어 부분 일치

## 검색 결과에 반드시 포함되어야 하는 중요 단어 선별

- 문장형 검색어

**톰포드 패블러스** 판매하는 곳

- 생략된 상품 코드

**as9371ikt.akor**



**[LG전자]코드제로 A9S 올인원타워 무선청소기 AS9371IKT**

**1,012,000원**

디지털/가전 > 생활가전 > 청소기 > 무선청소기

[롯데카드 5% 청구할인 : 961,400원] 신규 15%쿠폰 / 첫구매 10%쿠폰 / OK캐시백 1%적립

등록일 2021.08. · 찜하기 1 · 신고하기



# 3.1 검색어 부분 일치

## 검색 결과에 반드시 포함되어야 하는 중요 단어 선별

- 문장형 검색어

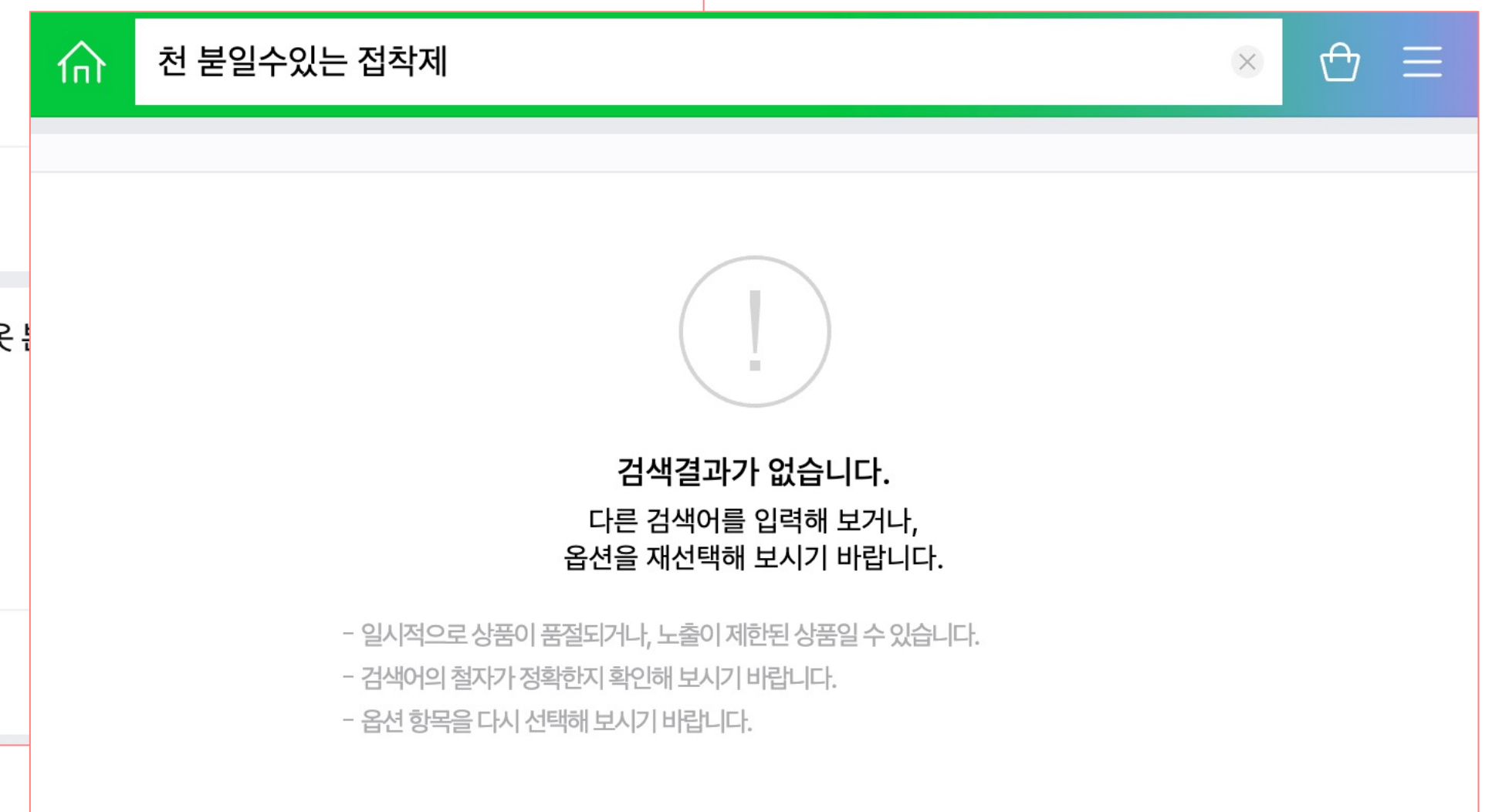
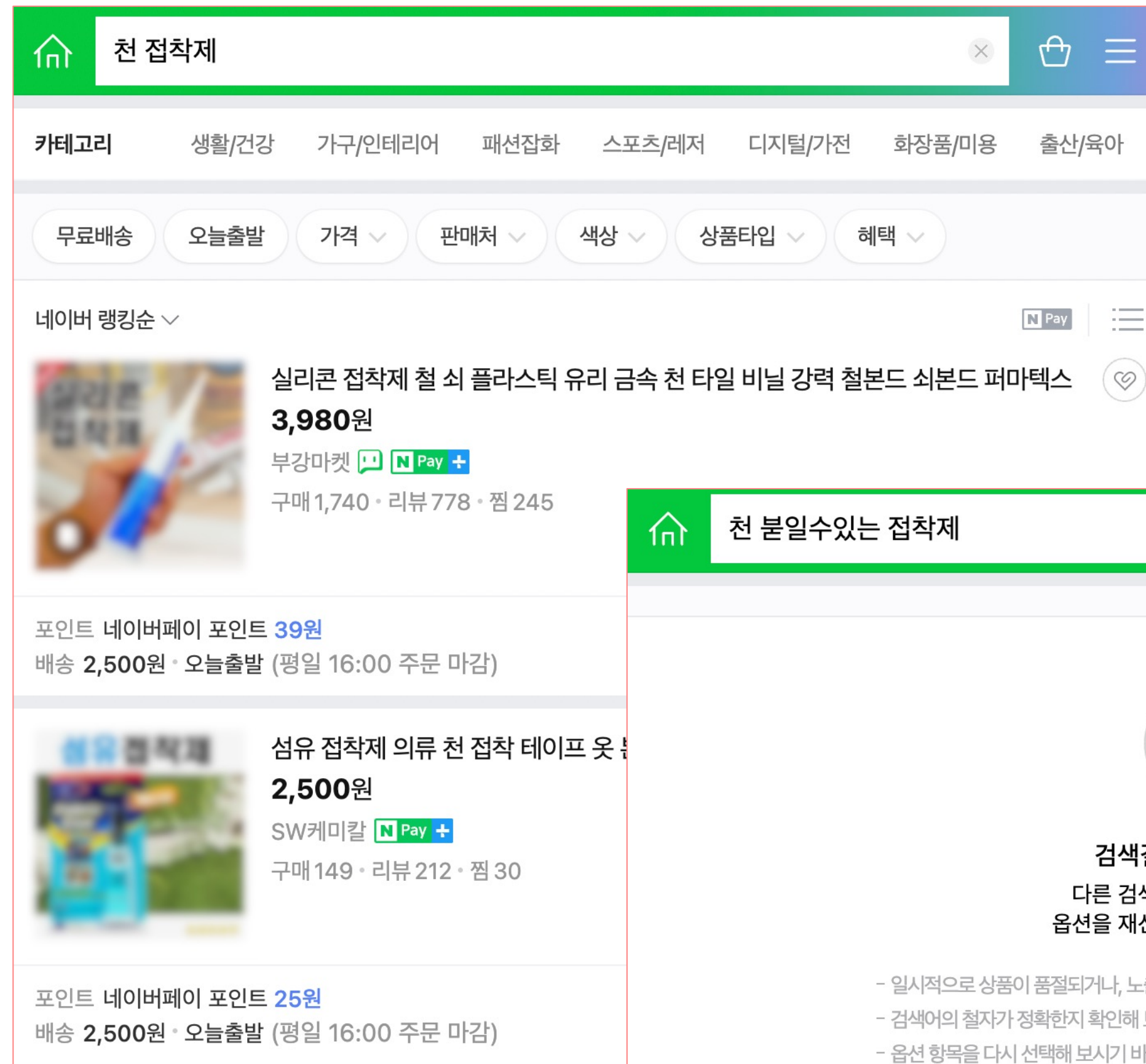
**톰포드 패블러스** 판매하는 곳

- 생략된 상품 코드

**as9371ikt.akor**

- 오타로 noisy term 삽입

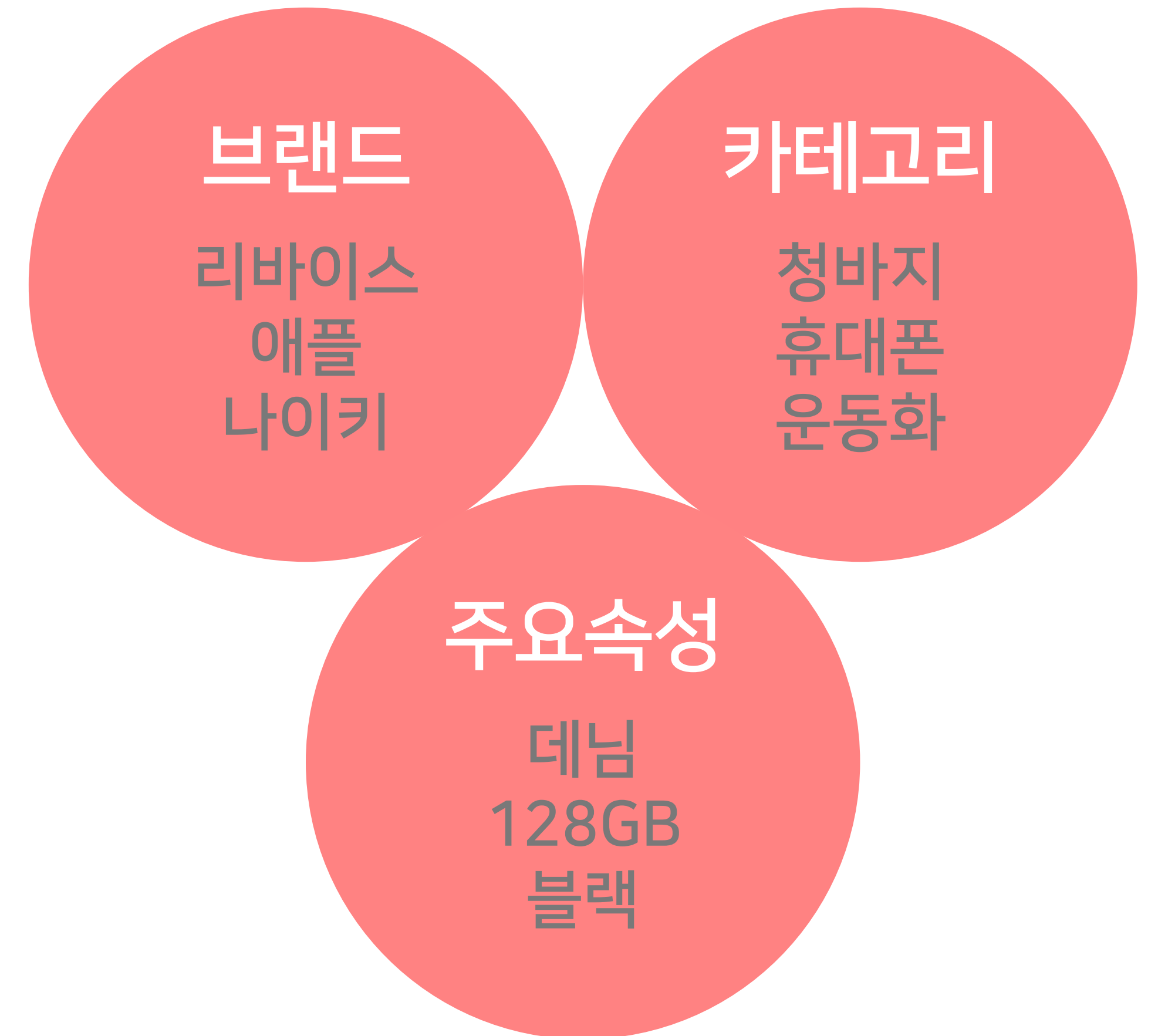
**천**분일수있는**접착제**



## 3.2 사전 기반 매칭의 부분 검색

### 사전 기반 매칭

- 쇼핑 주요 entity를 사전으로 관리  
브랜드, 카테고리, 주요 속성 등
- 검색어에서 사전에 매칭된 entity를  
중요 단어(AND term)로 선택



## 3.3 단어 중요도 예측 모델 (QTI Model)

### Query Term Importance (QTI) Model

- 검색어의 **단어별 중요도** 예측
- 사전 기반 매칭의 **커버리지 보완**

## 3.3 단어 중요도 예측 모델 (QTI Model)

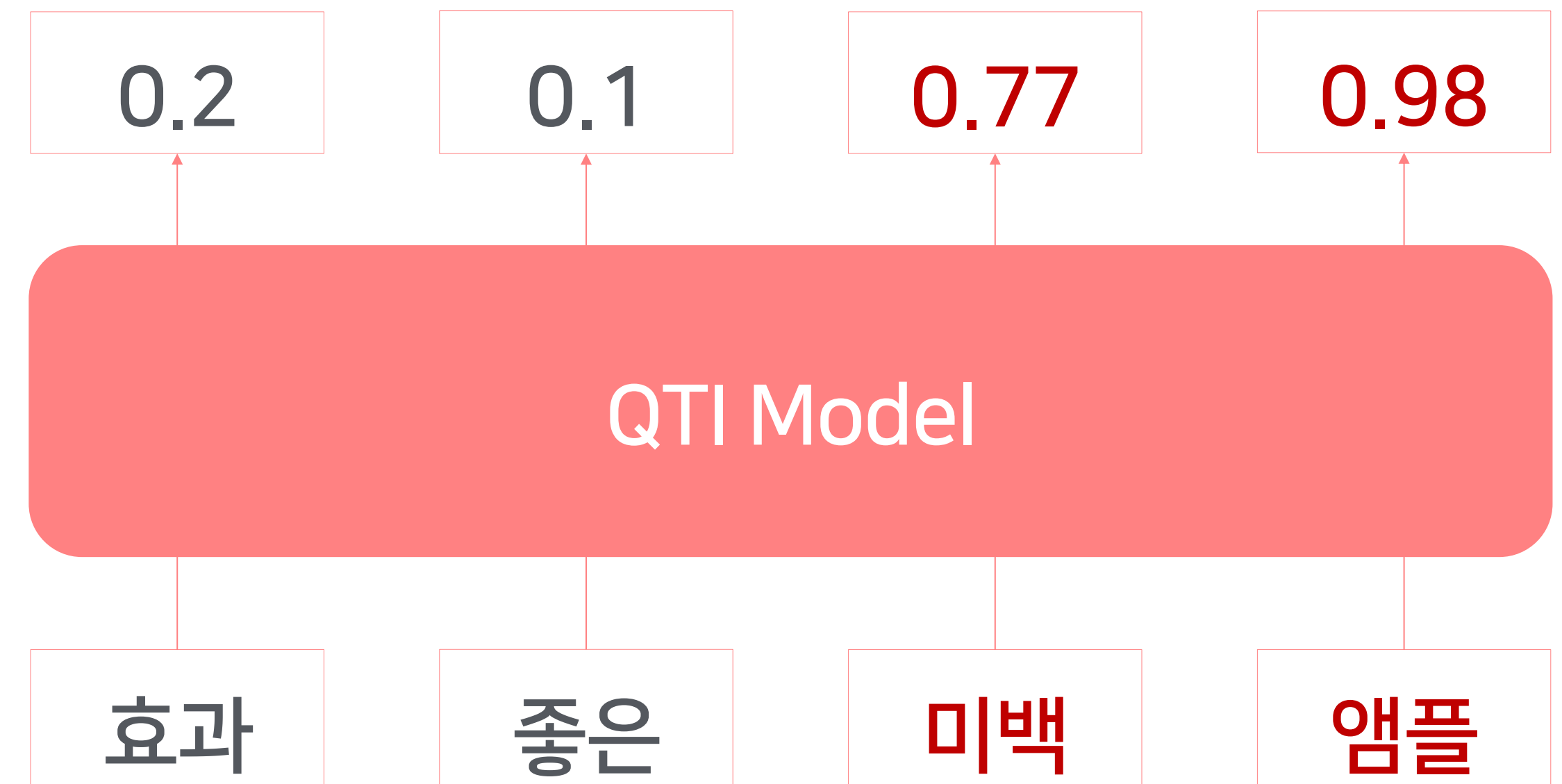
### Query Term Importance (QTI) Model

- 검색어의 **단어별 중요도** 예측
- 사전 기반 매칭의 **커버리지 보완**
- 가중치가 높은 단어를 **중요 단어(AND term)**로 검색
- 중요 단어를 포함한 상품에 **랭킹 가중치** 부여

# 3.3 단어 중요도 예측 모델 (QTI Model)

## Query Term Importance (QTI) Model

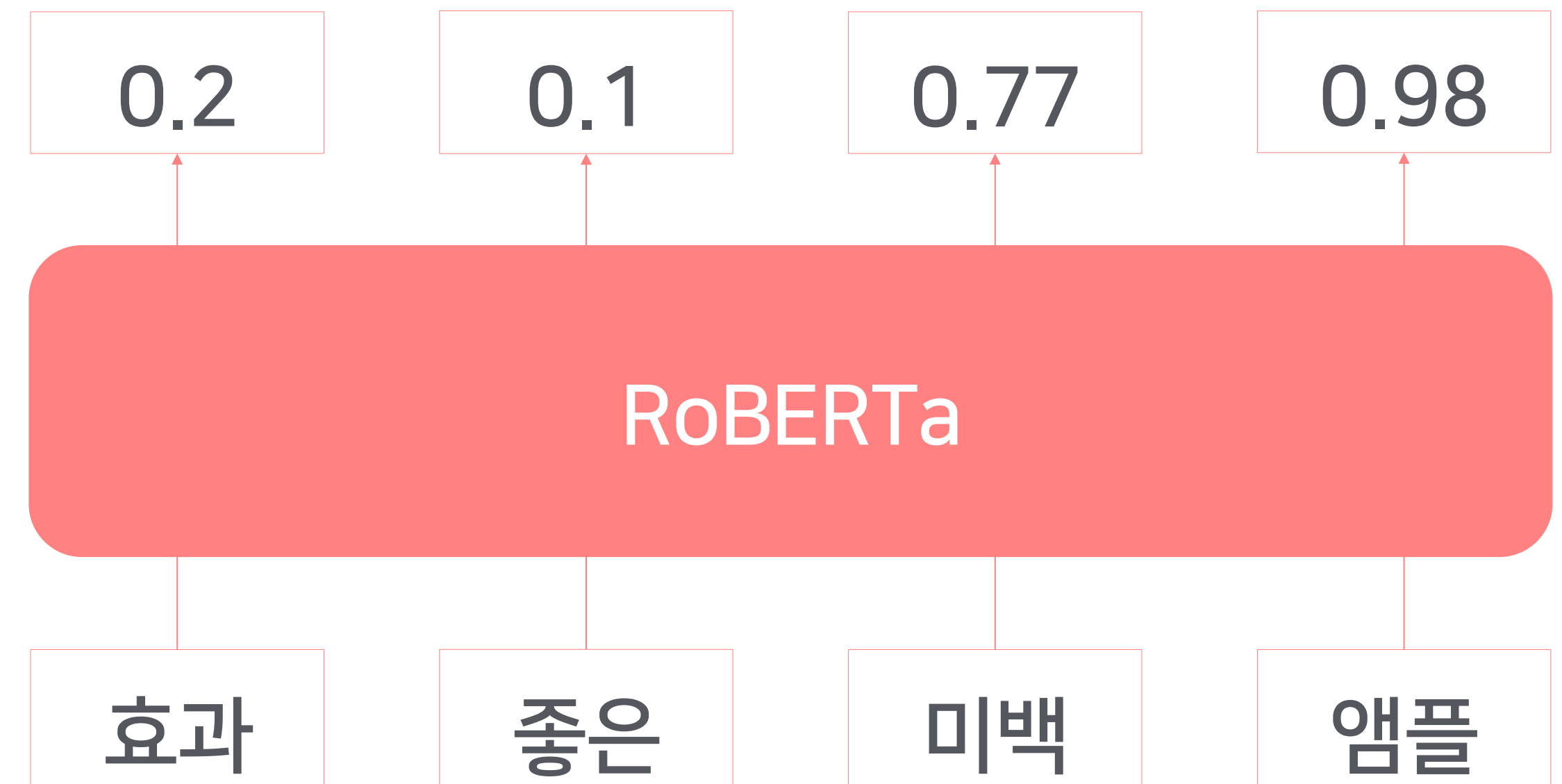
- 검색어의 **단어별 중요도** 예측
- 사전 기반 매칭의 **커버리지 보완**
- 가중치가 높은 단어를 **중요 단어(AND term)**로 검색
- 중요 단어를 포함한 상품에 **랭킹 가중치** 부여



# 3.3 단어 중요도 예측 모델 (QTI Model)

## Query Term Importance (QTI) Model

- RoBERTa: A Robustly Optimized BERT Pretraining Approach



# 3.3 단어 중요도 예측 모델 (QTI Model)

## Query Term Importance (QTI) Model

- RoBERTa: A Robustly Optimized BERT Pretraining Approach

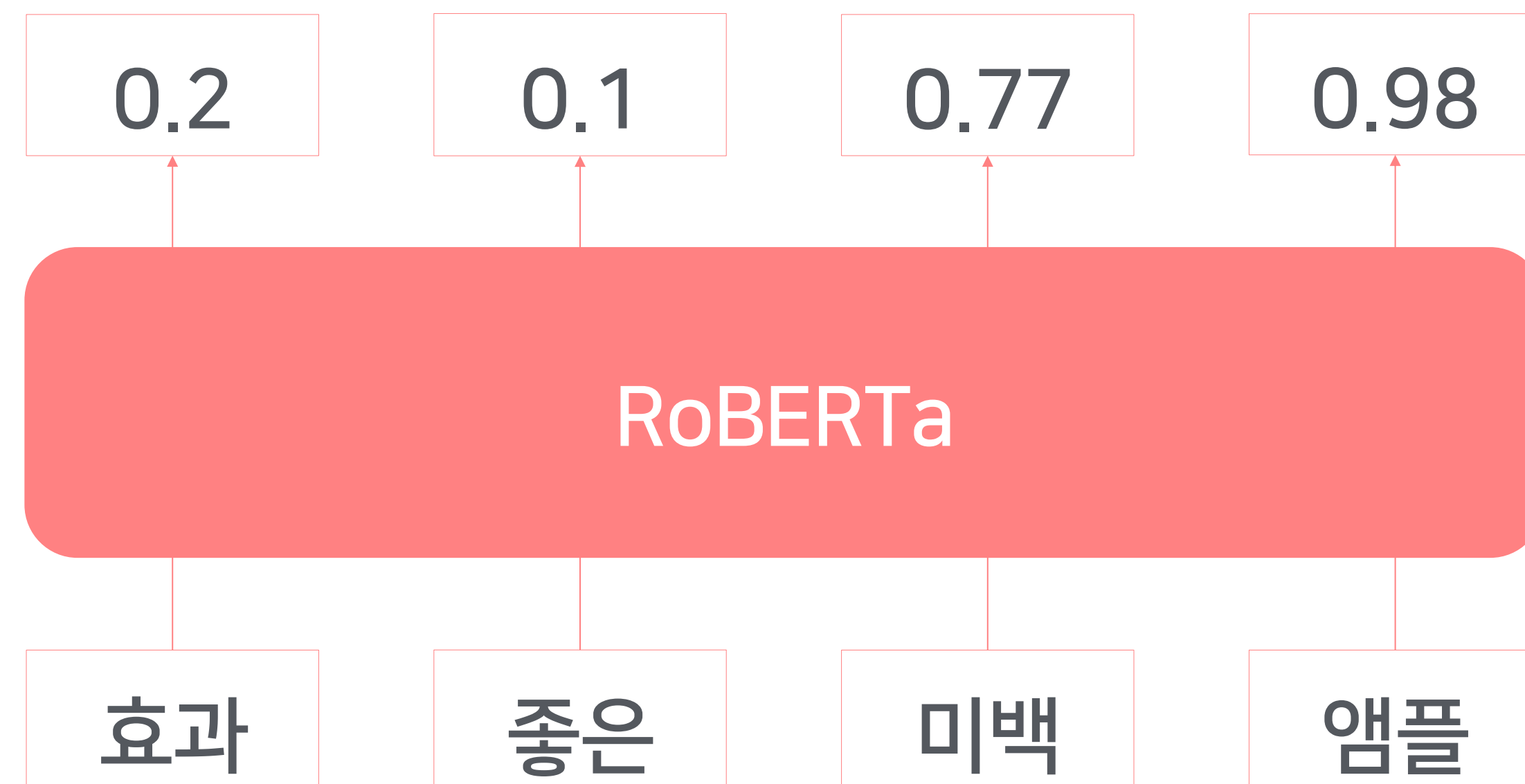
```

from transformers import RobertaForTokenClassification
from torch import nn

class RobertaForTokenRegression(RobertaForTokenClassification):
    def __init__(self, config):
        super().__init__(config)
        self.sigmoid = nn.Sigmoid()

    def forward(self, input_ids=None, attention_mask=None, token_type_ids=None):
        outputs = self.roberta(input_ids=input_ids,
                                attention_mask=attention_mask,
                                token_type_ids=token_type_ids)

        sequence_output = outputs[0]
        sequence_output = self.dropout(sequence_output)
        logits = self.classifier(sequence_output)
        sigmoid = self.sigmoid(logits)
        return sigmoid
    
```



# 3.3 단어 중요도 예측 모델 (QTI Model)

## Query Term Importance (QTI) Model

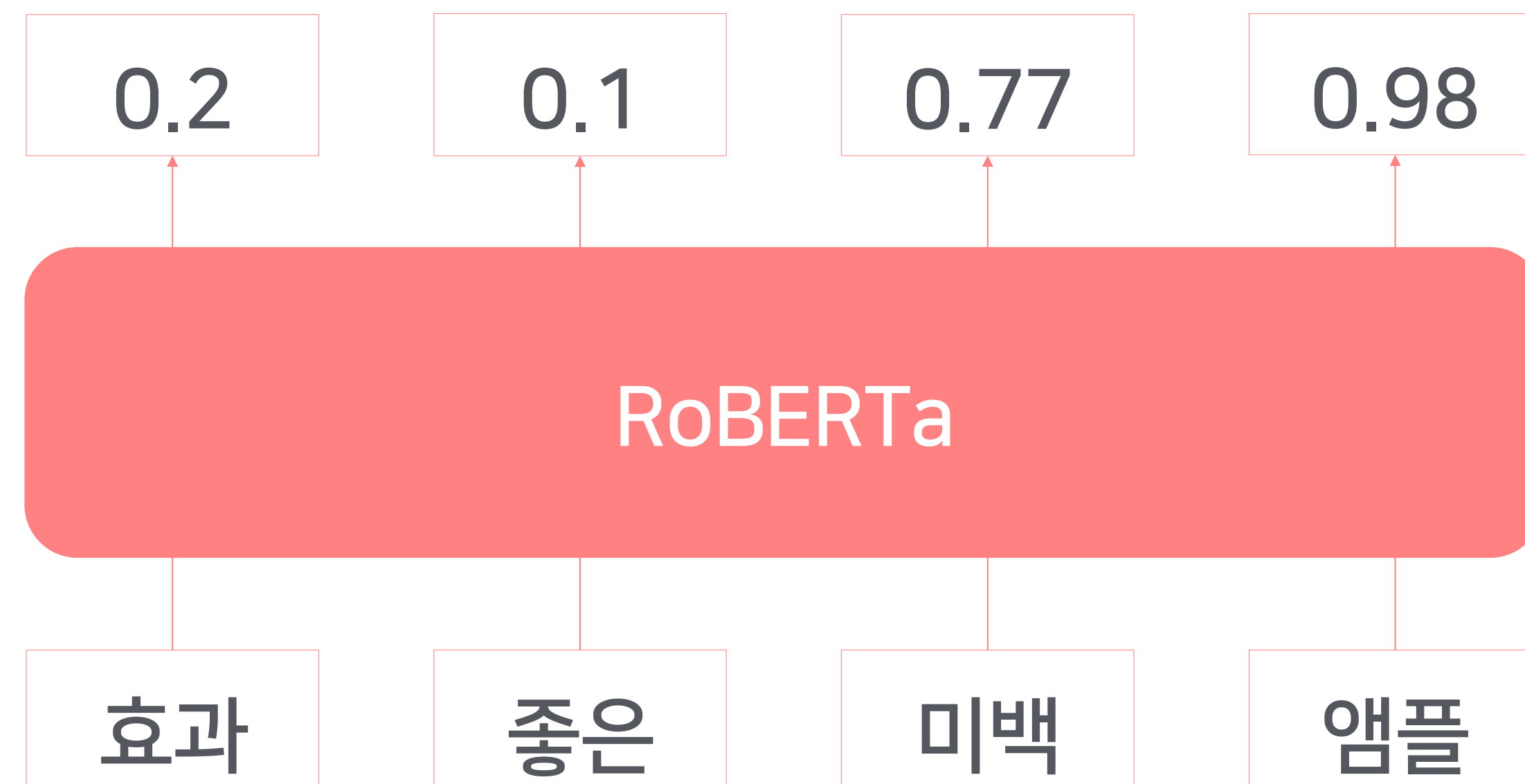
- RoBERTa: A Robustly Optimized BERT Pretraining Approach

```

from transformers import RobertaForTokenClassification
from torch import nn

class RobertaForTokenRegression(RobertaForTokenClassification):
    def __init__(self, config):
        super().__init__(config)
        self.sigmoid = nn.Sigmoid()

    def forward(self, input_ids=None, attention_mask=None, token_type_ids=None):
        outputs = self.roberta(input_ids=input_ids,
                               attention_mask=attention_mask,
                               token_type_ids=token_type_ids)
        sequence_output = outputs[0]
        sequence_output = self.dropout(sequence_output)
        logits = self.classifier(sequence_output)
        sigmoid = self.sigmoid(logits)
        return sigmoid
    
```





# 3.3 단어 중요도 예측 모델 (QTI Model)

## Query Term Importance (QTI) Model

- RoBERTa: A Robustly Optimized BERT Pretraining Approach

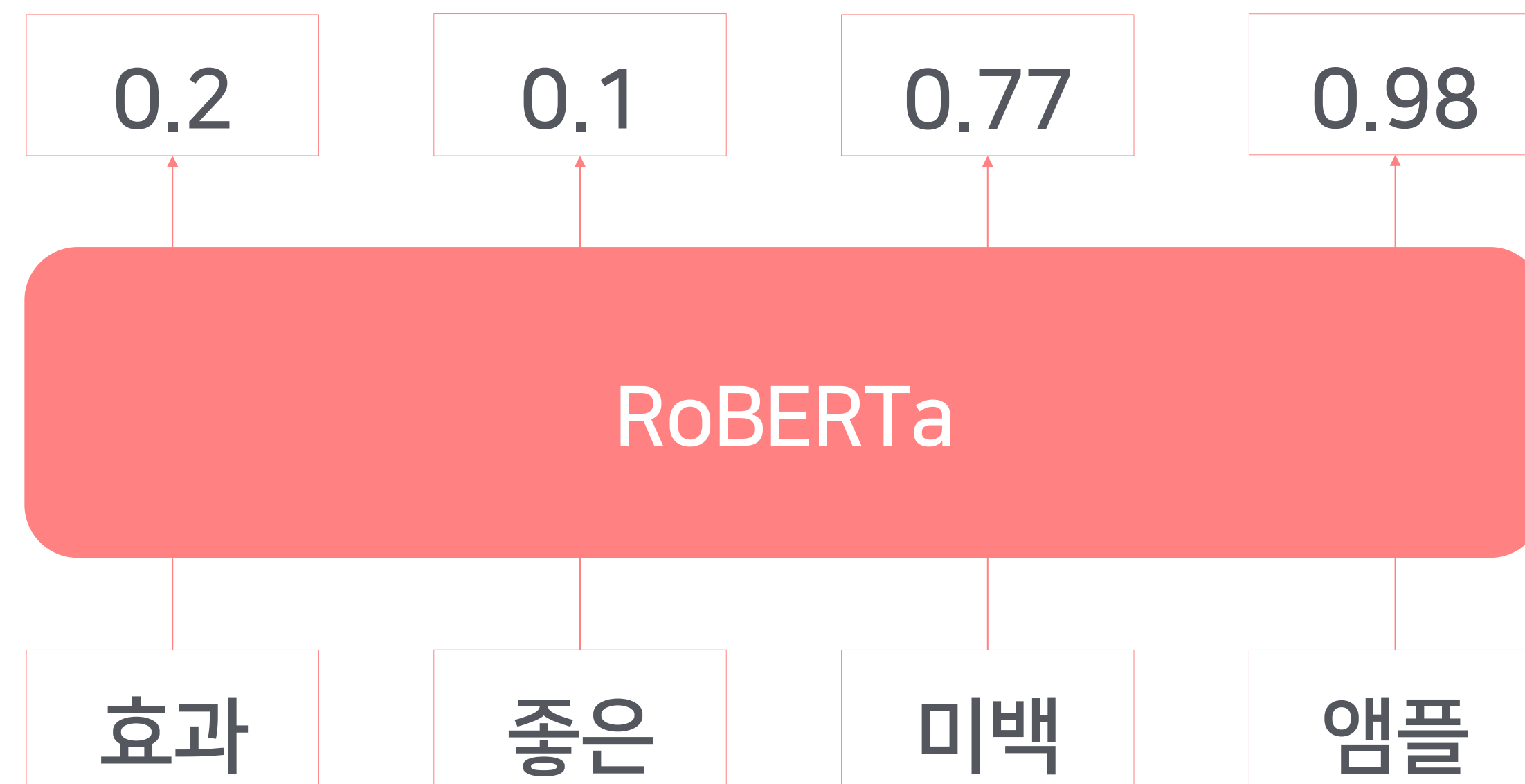
```

from transformers import RobertaForTokenClassification
from torch import nn

class RobertaForTokenRegression(RobertaForTokenClassification):
    def __init__(self, config):
        super().__init__(config)
        self.sigmoid = nn.Sigmoid()

    def forward(self, input_ids=None, attention_mask=None, token_type_ids=None):
        outputs = self.roberta(input_ids=input_ids,
                                attention_mask=attention_mask,
                                token_type_ids=token_type_ids)

        sequence_output = outputs[0]
        sequence_output = self.dropout(sequence_output)
        logits = self.classifier(sequence_output)
        sigmoid = self.sigmoid(logits)
        return sigmoid
    
```



# 3.3 단어 중요도 예측 모델 (QTI Model)

## Query Term Importance (QTI) Model

- RoBERTa: A Robustly Optimized BERT Pretraining Approach

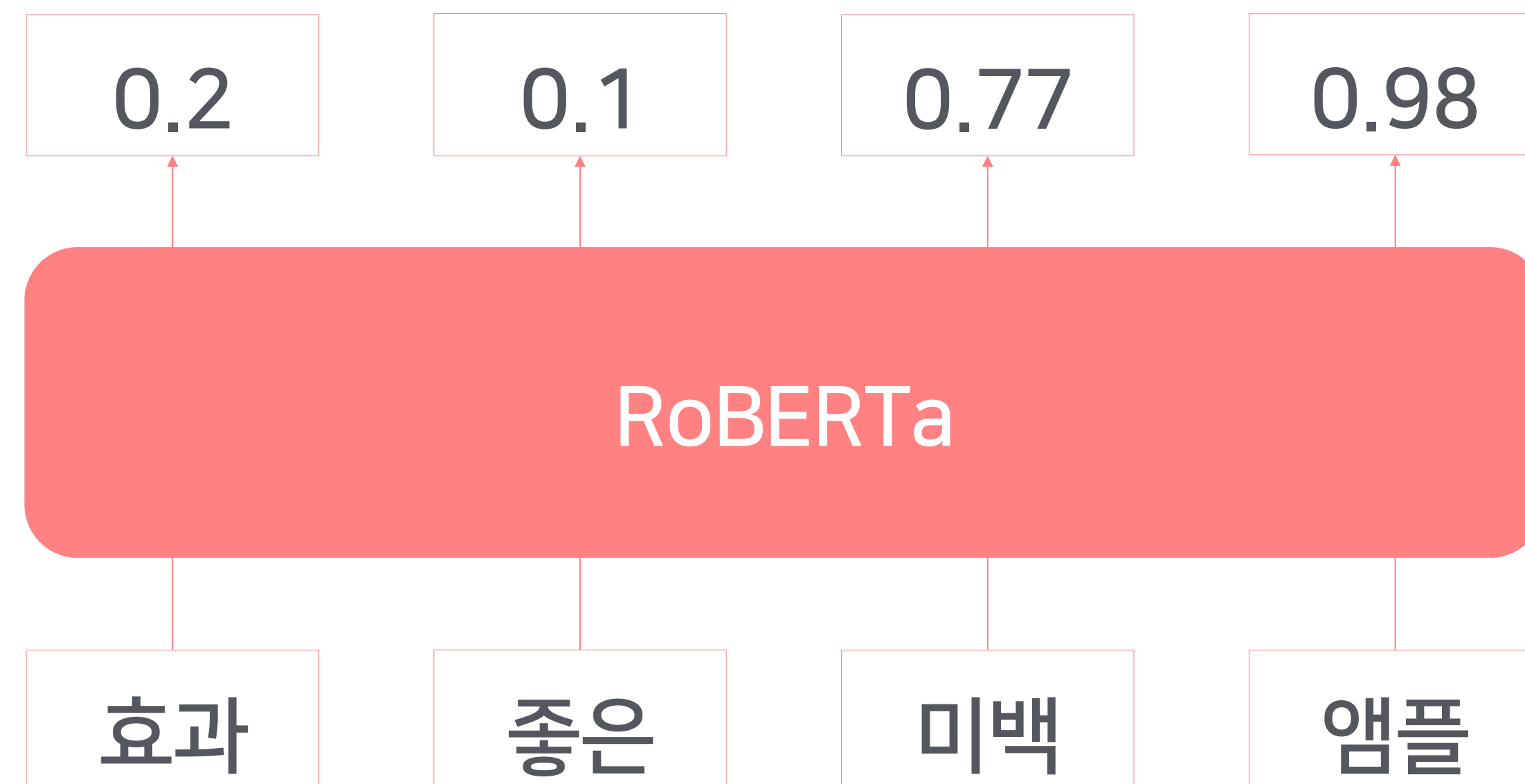
```

from transformers import RobertaForTokenClassification
from torch import nn

class RobertaForTokenRegression(RobertaForTokenClassification):
    def __init__(self, config):
        super().__init__(config)
        self.sigmoid = nn.Sigmoid()

    def forward(self, input_ids=None, attention_mask=None, token_type_ids=None):
        outputs = self.roberta(input_ids=input_ids,
                                attention_mask=attention_mask,
                                token_type_ids=token_type_ids)

        sequence_output = outputs[0]
        sequence_output = self.dropout(sequence_output)
        logits = self.classifier(sequence_output)
        sigmoid = self.sigmoid(logits)
        return sigmoid
    
```



# 3.4 QTI Model Pretrain

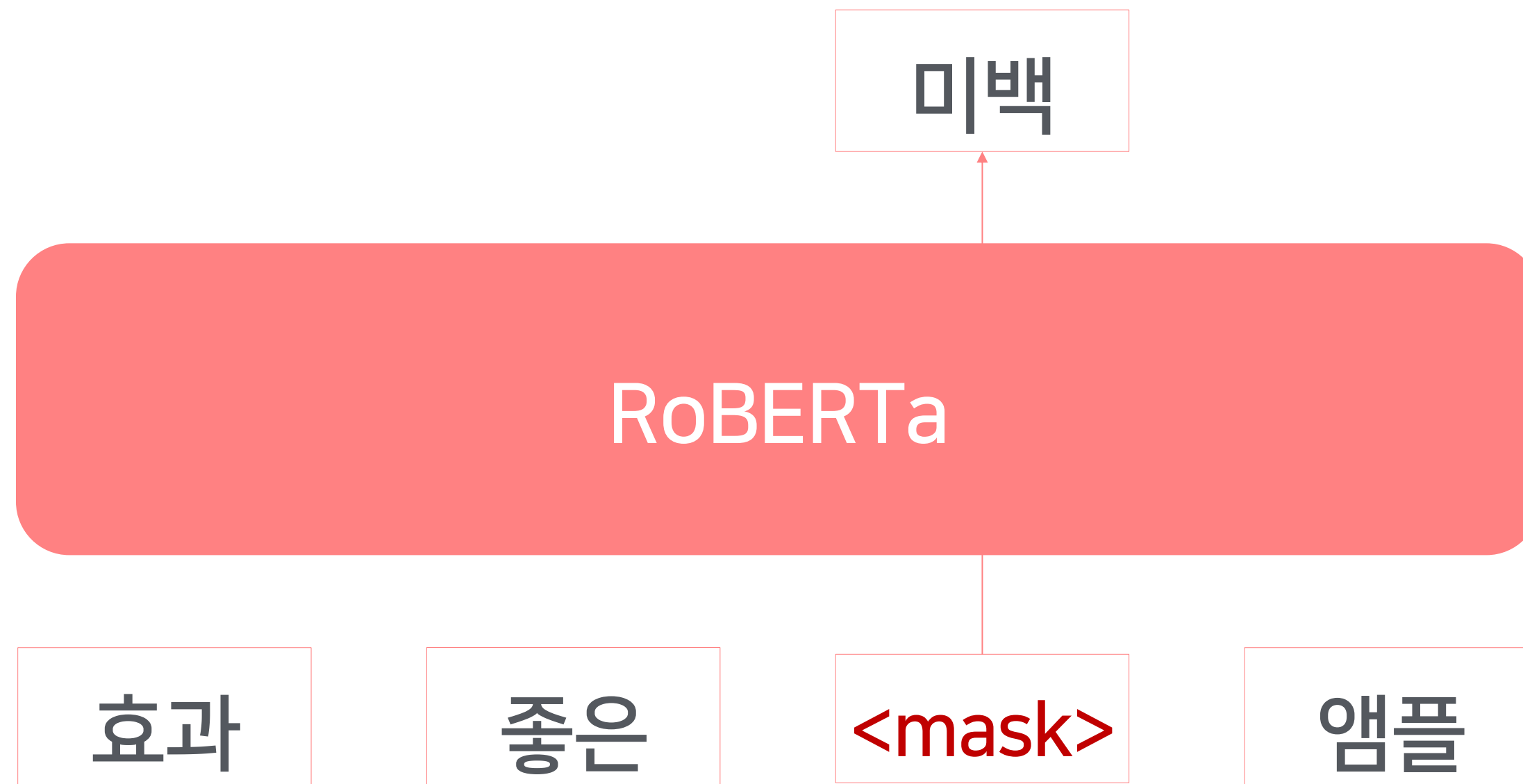
쇼핑 검색어와 상품명으로  
masked language model pretraining



# 3.4 QTI Model Pretrain

쇼핑 검색어와 상품명으로  
masked language model pretraining

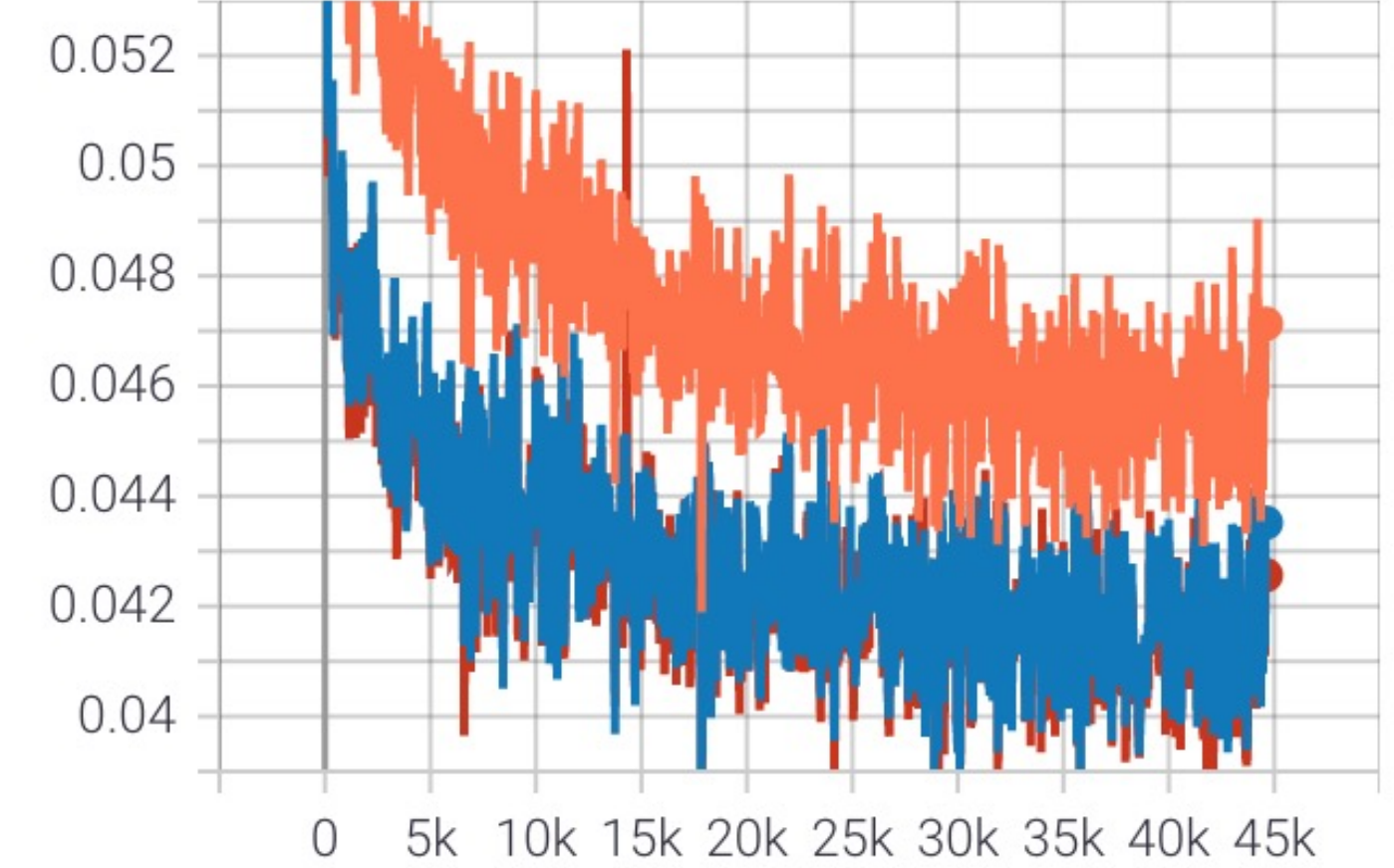
- version\_0 : no pretrain (random 초기화)
- version\_1 : **1M** step pretrain
- version\_2 : **2M** step pretrain



train\_loss

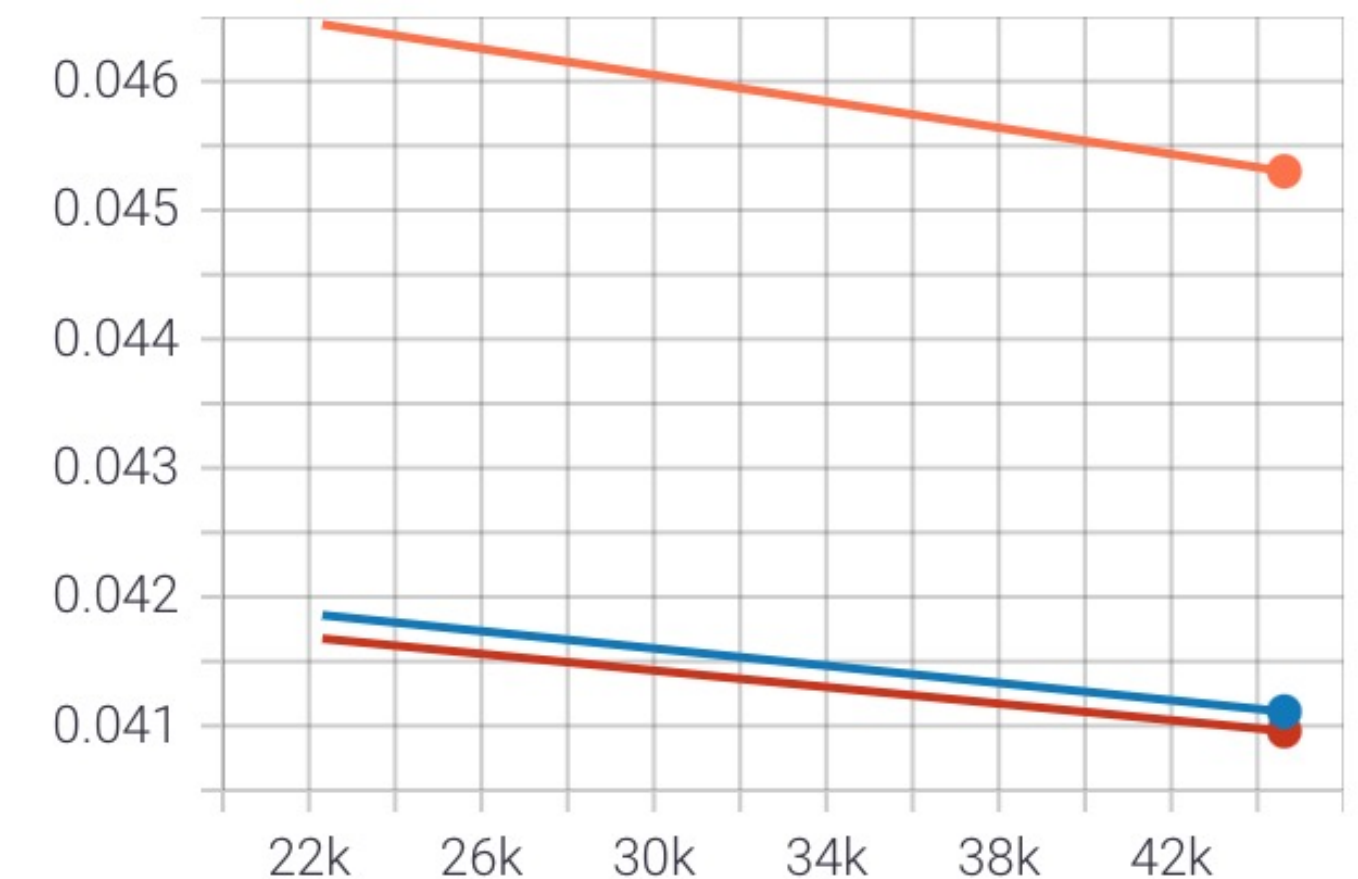
**N** DEVIEW  
2021

train\_loss



Name	Smoothed	Value	Step
default/version_0	0.0453	0.0453	44.62k
default/version_1	0.04111	0.04111	44.62k
default/version_2	0.04096	0.04096	44.62k

valid\_loss



# 3.5 QTI Model 학습 데이터 구축

## 1. 사용자 검색어 입력 시퀀스에서 추출

Session Id	Query	Expose items	Click items	Date Time
u6MQ0qL8	긴 테이블 1350	3	0	2021-11-11 13:00:02
u6MQ0qL8	테이블 1350	1,057	5	2021-11-11 13:00:05

- $\text{Set}(Q1) \supset \text{Set}(Q2)$  검색어 쌍
- Q1, Q2 에 같이 출현한 단어에 가중치 부여

# 3.5 QTI Model 학습 데이터 구축

## 1. 사용자 검색어 입력 시퀀스에서 추출

Session Id	Query	Expose items	Click items	Date Time
u6MQ0qL8	긴 테이블 1350	3	0	2021-11-11 13:00:02
u6MQ0qL8	테이블 1350	1,057	5	2021-11-11 13:00:05

- Set(Q1) ⊃ Set(Q2) 검색어 쌍
- Q1, Q2 에 같이 출현한 단어에 가중치 부여  
**긴 테이블 1350 → 테이블 1350**  
 ⇒ [ ("긴", 0.0), ("테이블", 1.0), ("1350", 1.0) ]

The image shows two overlapping screenshots of a Naver shopping search results page for the query '테이블 1350'. The left screenshot is titled '긴 테이블 1350' and shows three product listings. The right screenshot is titled '테이블 1350' and shows a list of 1,057 items, with the first few items visible. The products include '베스티아가구 복도 엔틱 현탁 높은 유럽풍 서랍형콘솔' (620,000원), '해찬솔원목이야기 통원목 탁(1350긴의자2개)' (1,391,000원), '해외 24V 전기 선형 액추에이터 N 1350LBS 긴 스트로크' (302,540원), '한샘 바흐 701 인칸토 4인 식탁 1350 고급형' (1,374,000원), and '동서가구 네오 장미원목 우드슬랩 테이블1350 로지아 NEO482' (158,000원).

# 3.5 QTI Model 학습 데이터 구축

## 2. 검색어 - 상품 클릭 데이터에서 추출

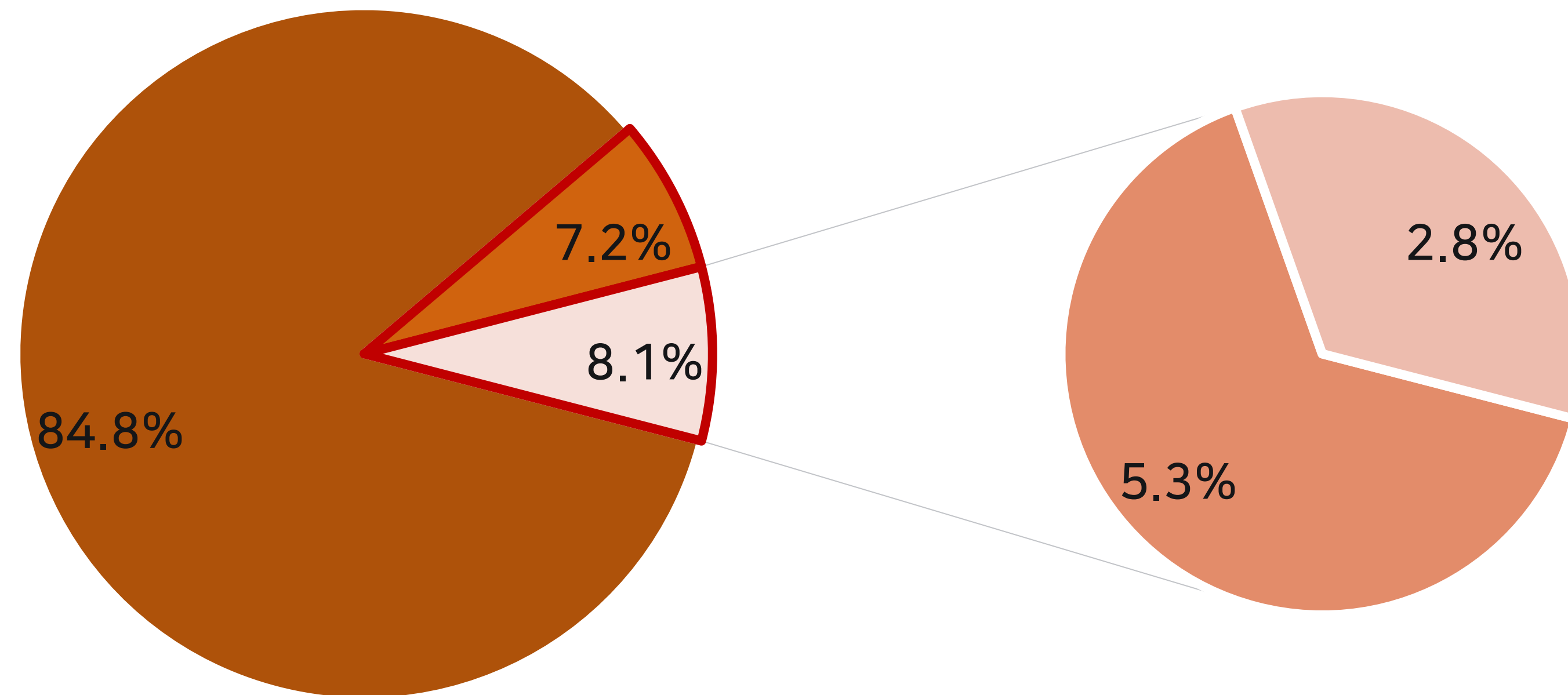
- Set(검색어)  $\cap$  Set(클릭 된 상품명) 에 해당하는 단어에 가중치 부여



# 3.6 QTI 효과

## 쇼핑 클릭 질의 분포 (UQC)

- 부분 일치 검색 미동작
- 부분 일치 검색 동작 (OFF시 검색 결과 있음)
- 부분 일치 검색 동작 (OFF시 검색 결과 0건) 사전기반
- 부분 일치 검색 동작 (OFF 시 검색 결과 0건) QTI

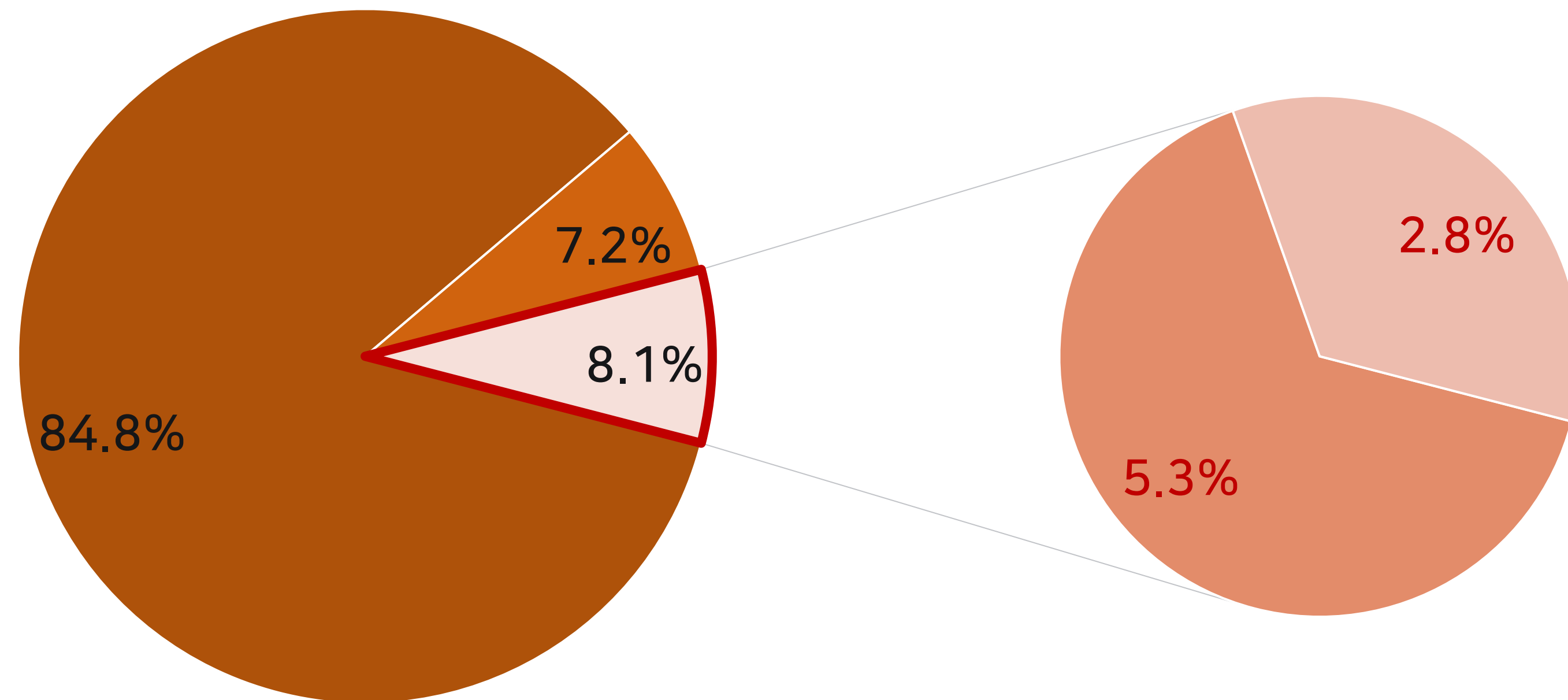




# 3.6 QTI 효과

## 쇼핑 클릭 질의 분포 (UQC)

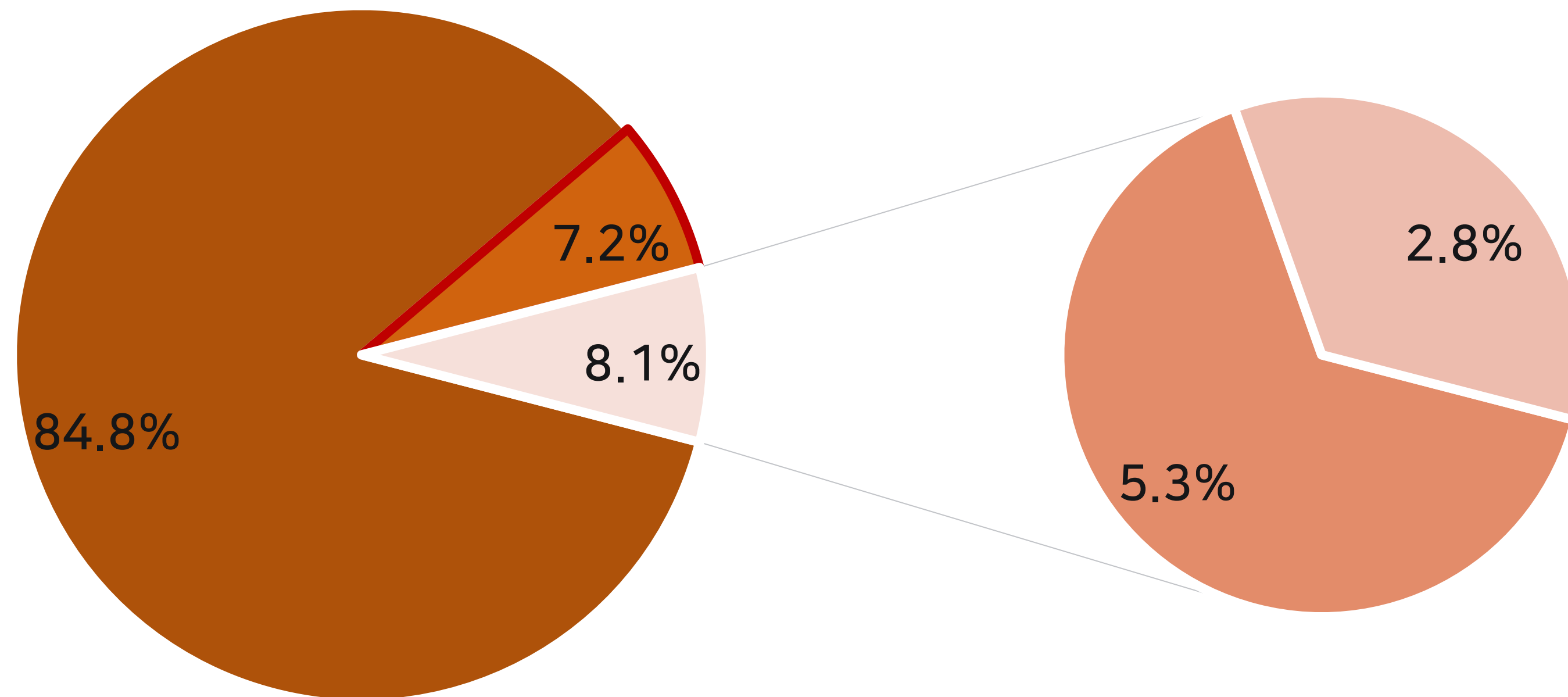
- 부분 일치 검색 미동작
- 부분 일치 검색 동작  
(OFF시 검색 결과 있음)
- 부분 일치 검색 동작  
(OFF시 검색 결과 0건)  
사전기반
- 부분 일치 검색 동작  
(OFF 시 검색 결과 0건)  
QTI



# 3.6 QTI 효과

## 쇼핑 클릭 질의 분포 (UQC)

- 부분 일치 검색 미동작
- 부분 일치 검색 동작  
(OFF시 검색 결과 있음)
- 부분 일치 검색 동작  
(OFF시 검색 결과 0건)  
사전기반
- 부분 일치 검색 동작  
(OFF 시 검색 결과 0건)  
QTI

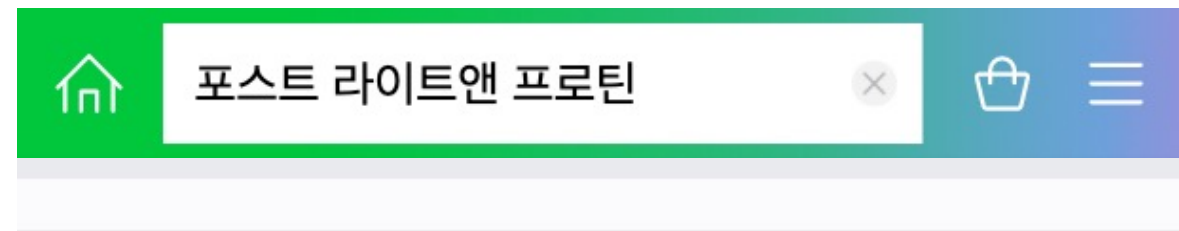


# 4. BERT 기반 검색 : 쇼핑 딥서치

# 4.1 BERT 기반 쇼핑 검색

## 키워드 매칭이 어려운 검색어

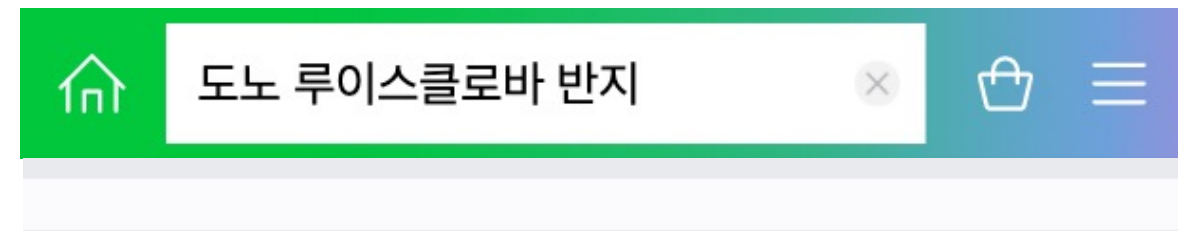
상품명 표기 오류



검색결과가 없습니다.  
다른 검색어를 입력해 보거나,  
옵션을 재선택해 주시기 바랍니다.

- 일시적으로 상품이 품절되거나, 노출이 제한된 상품일 수 있습니다.
- 검색어의 철자가 정확한지 확인해 주시기 바랍니다.
- 옵션 항목을 다시 선택해 주시기 바랍니다.

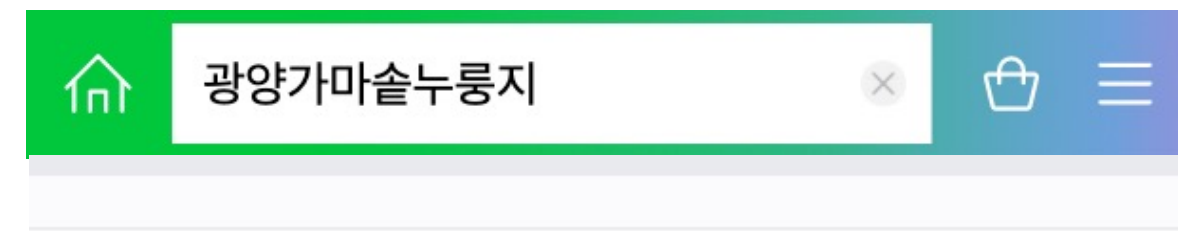
외래어 표기 불일치



검색결과가 없습니다.  
다른 검색어를 입력해 보거나,  
옵션을 재선택해 주시기 바랍니다.

- 일시적으로 상품이 품절되거나, 노출이 제한된 상품일 수 있습니다.
- 검색어의 철자가 정확한지 확인해 주시기 바랍니다.
- 옵션 항목을 다시 선택해 주시기 바랍니다.

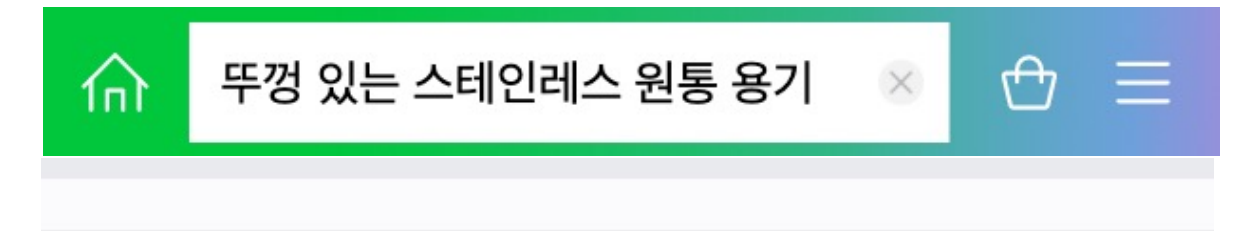
수급되지 않은 상품



검색결과가 없습니다.  
다른 검색어를 입력해 보거나,  
옵션을 재선택해 주시기 바랍니다.

- 일시적으로 상품이 품절되거나, 노출이 제한된 상품일 수 있습니다.
- 검색어의 철자가 정확한지 확인해 주시기 바랍니다.
- 옵션 항목을 다시 선택해 주시기 바랍니다.

자연어 형태의 긴 검색어

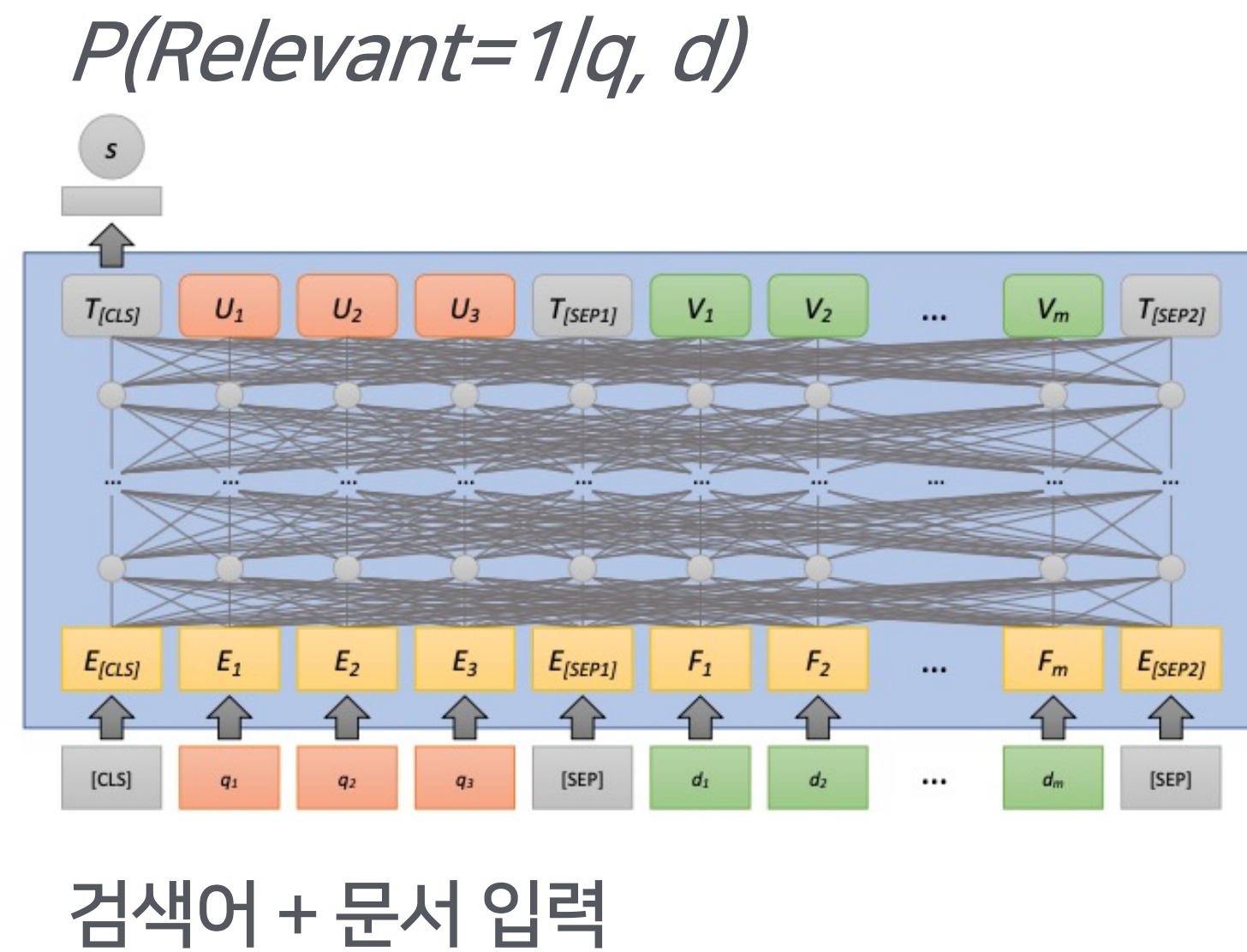


검색결과가 없습니다.  
다른 검색어를 입력해 보거나,  
옵션을 재선택해 주시기 바랍니다.

- 일시적으로 상품이 품절되거나, 노출이 제한된 상품일 수 있습니다.
- 검색어의 철자가 정확한지 확인해 주시기 바랍니다.
- 옵션 항목을 다시 선택해 주시기 바랍니다.

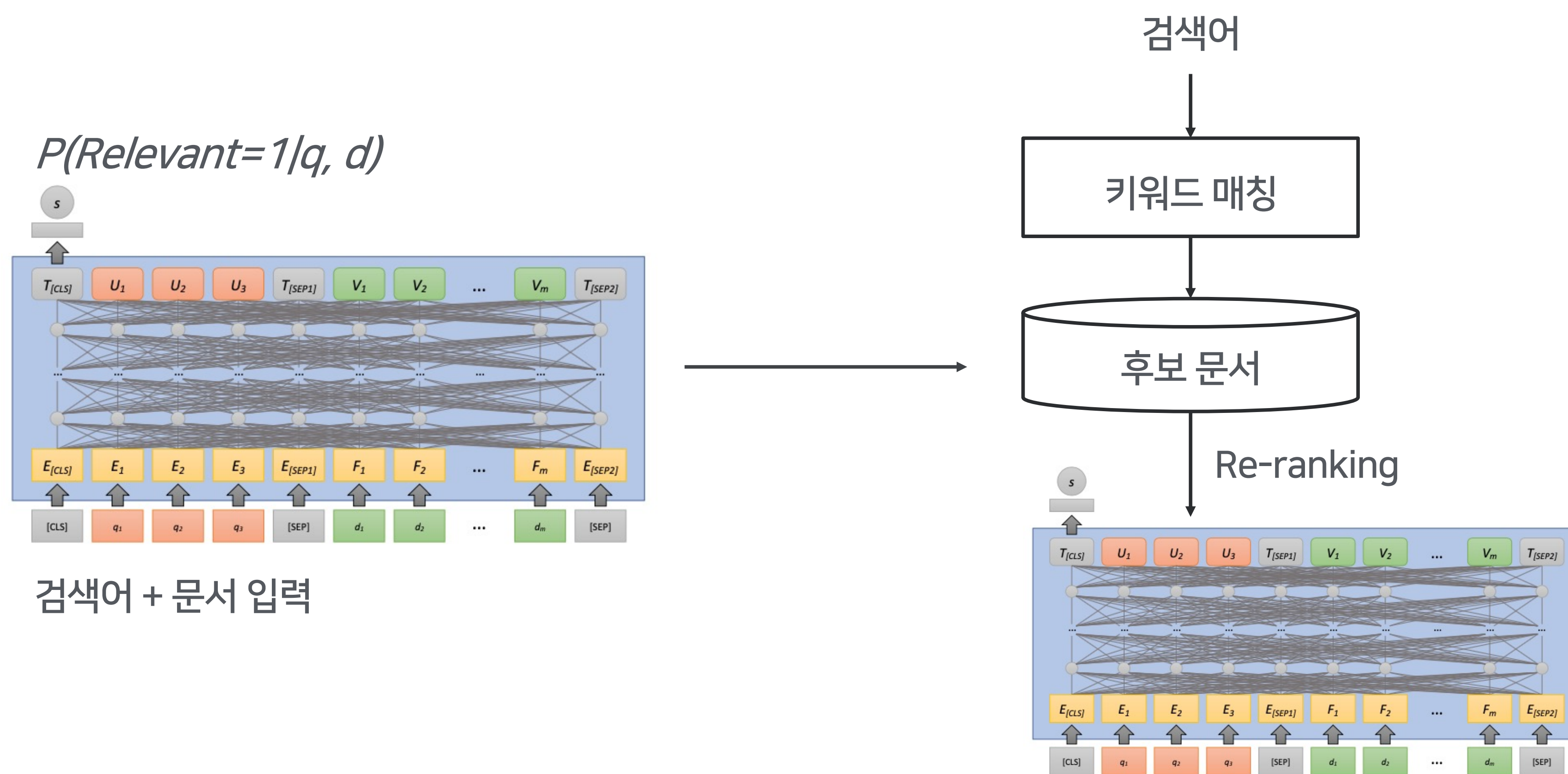
# 4.1 BERT 기반 쇼핑 검색

BERT 기반 검색 모델: cross-encoder 방식



# 4.1 BERT 기반 쇼핑 검색

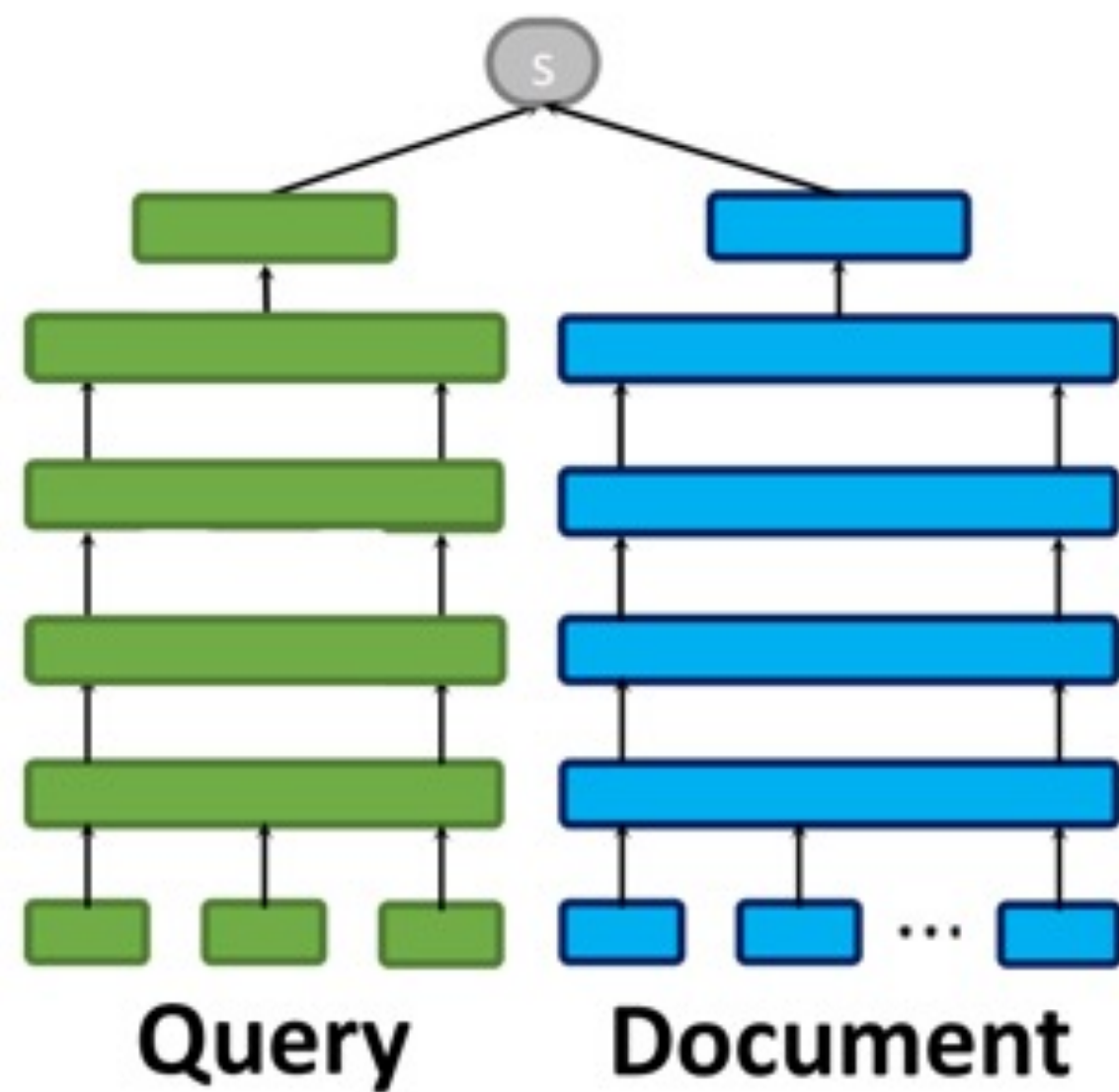
BERT 기반 검색 모델: cross-encoder 방식



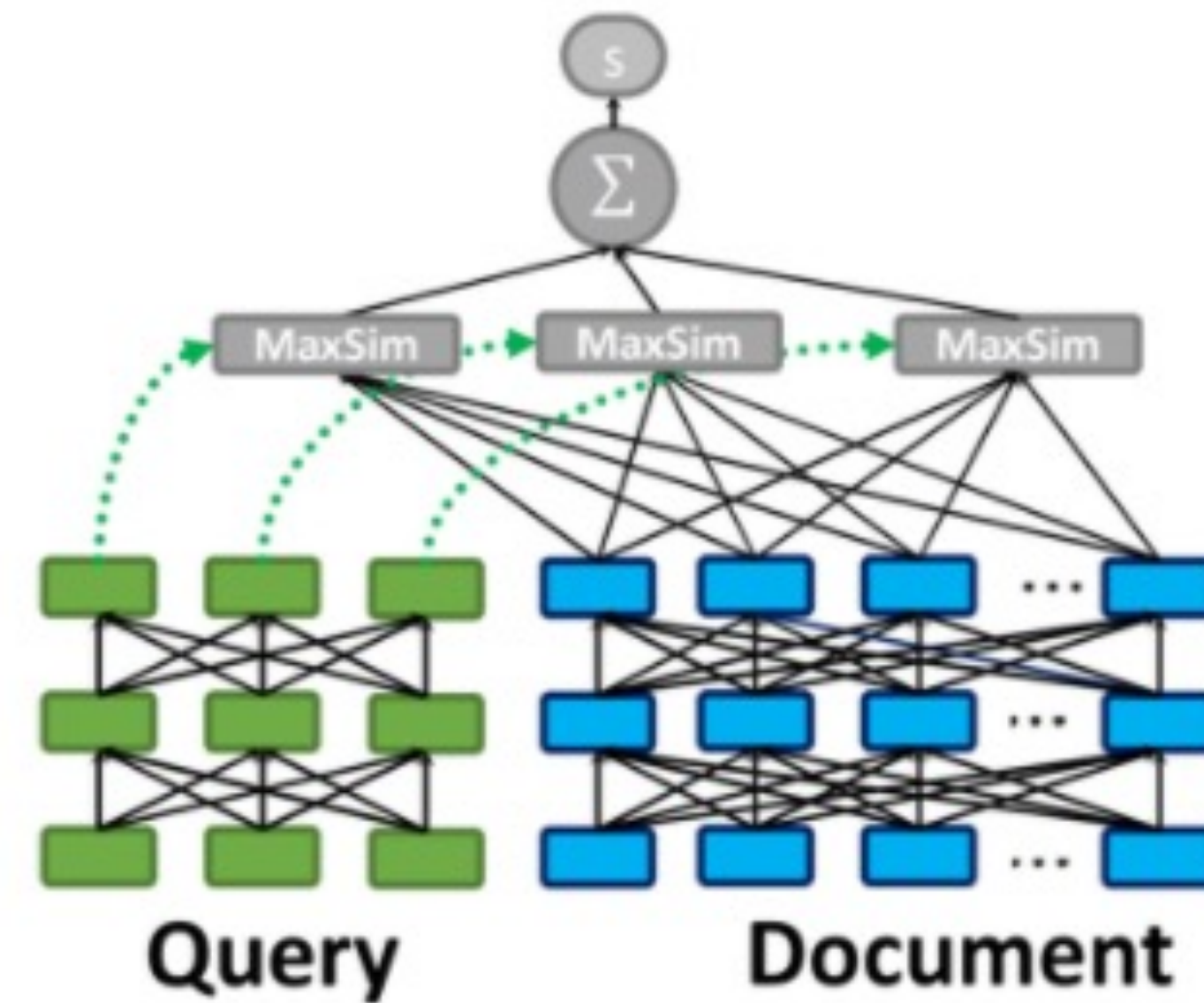
# 4.1 BERT 기반 쇼핑 검색

BERT 기반 검색 모델: bi-encoder 방식

$$P(\text{Relevant}=1|d,q) \leftarrow \phi(\eta(q), \eta(d))$$



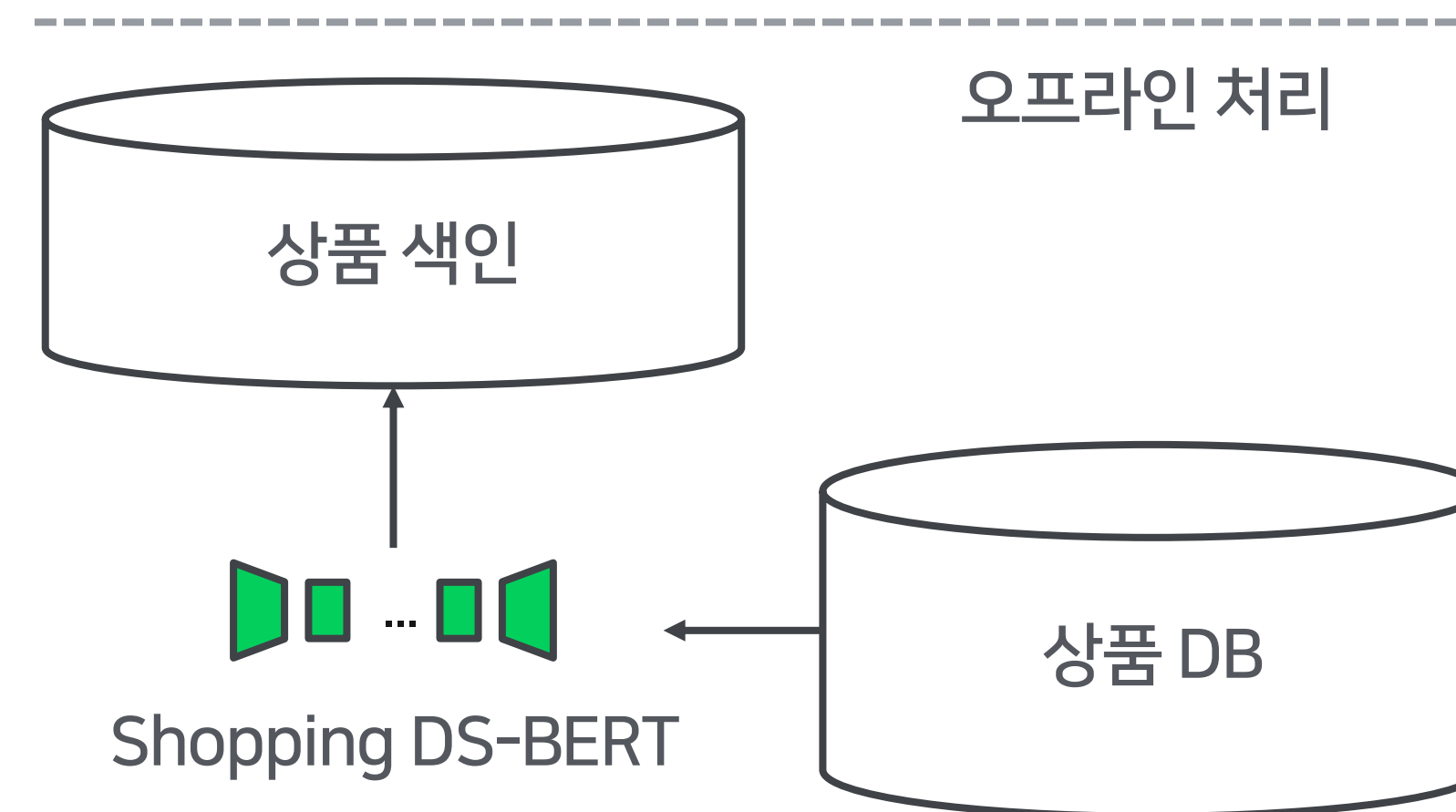
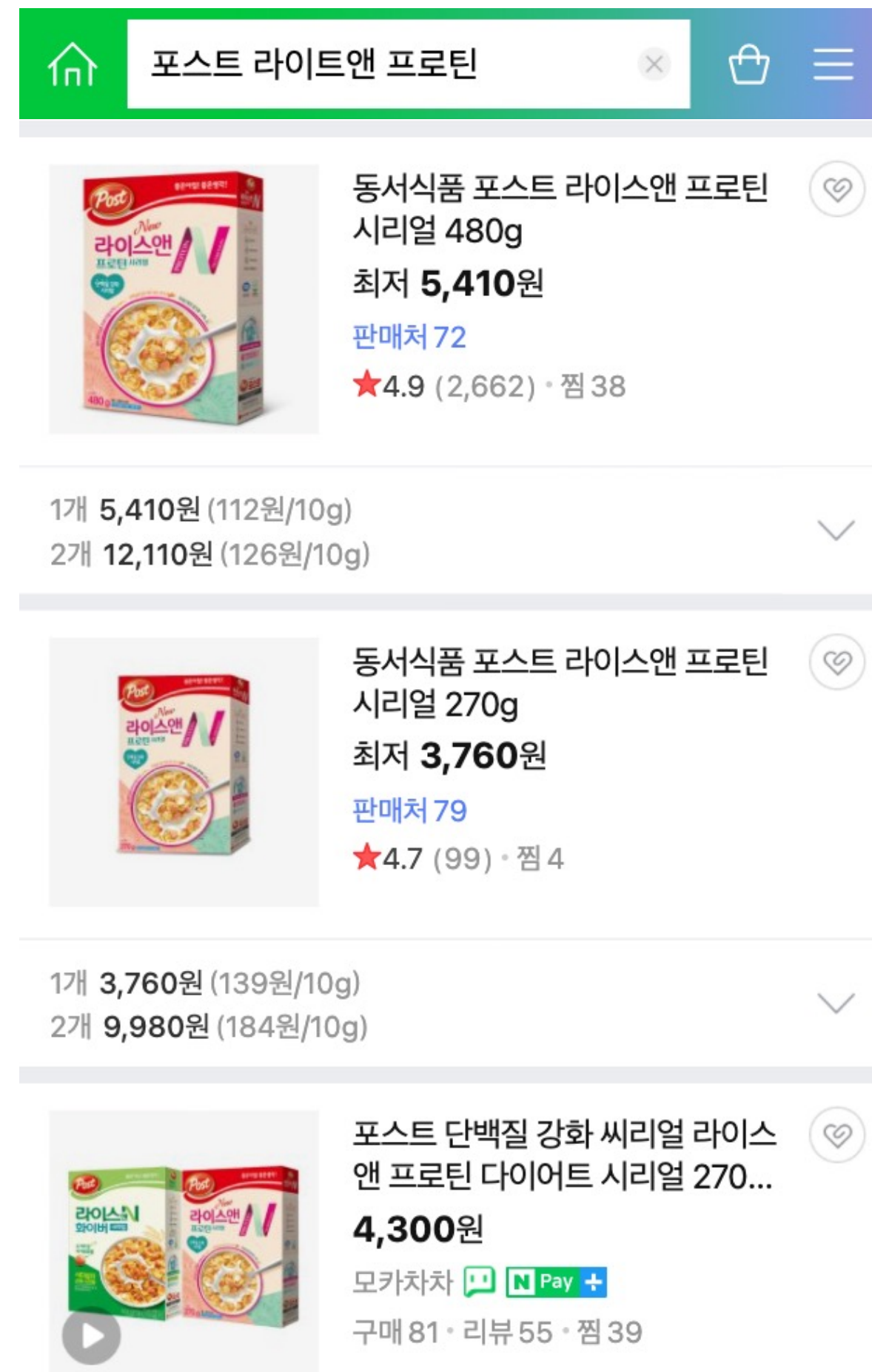
SentenceBERT



ColBERT

# 4.2 쇼핑 딥서치

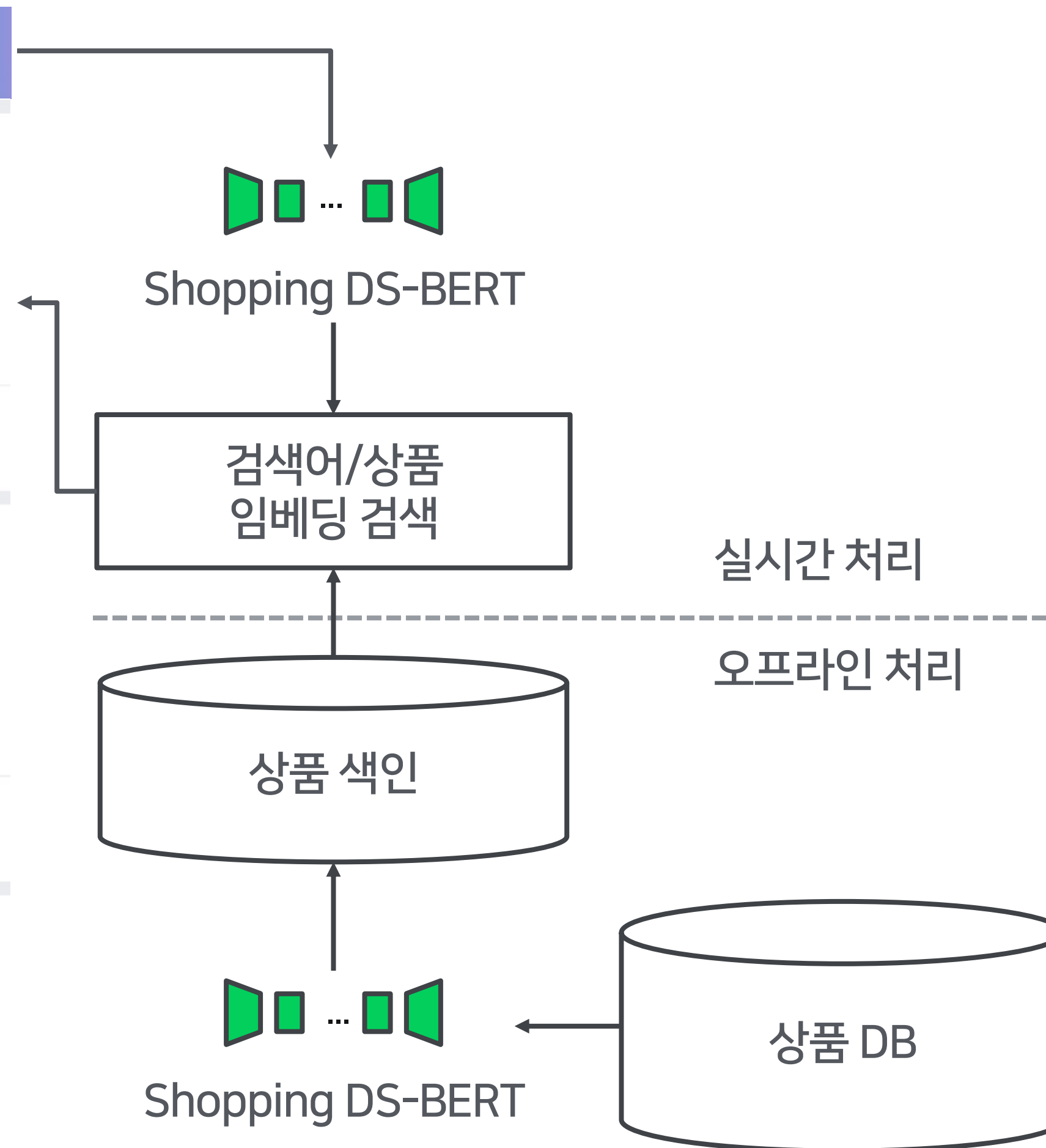
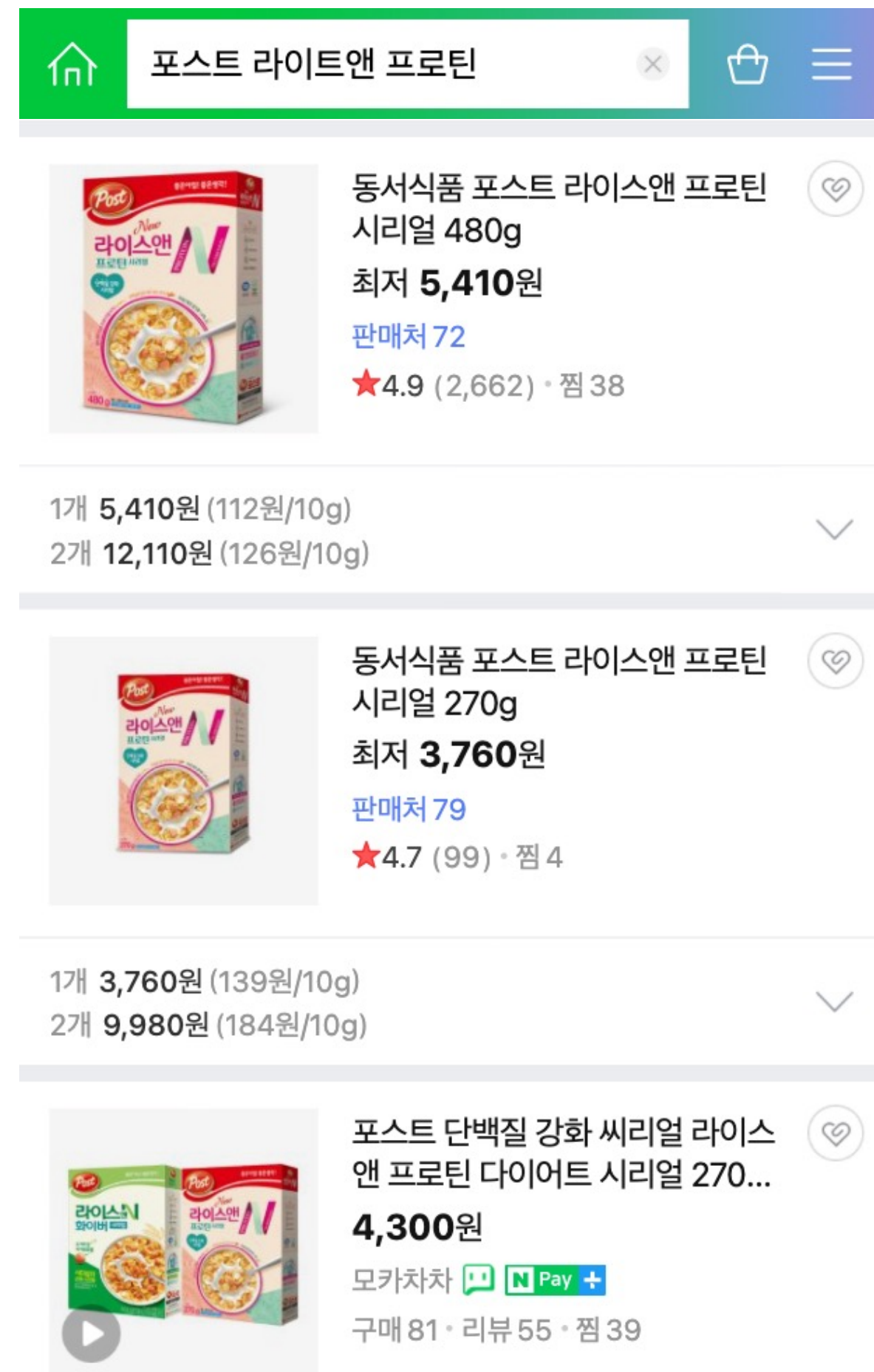
## 쇼핑 딥서치 모델 개요





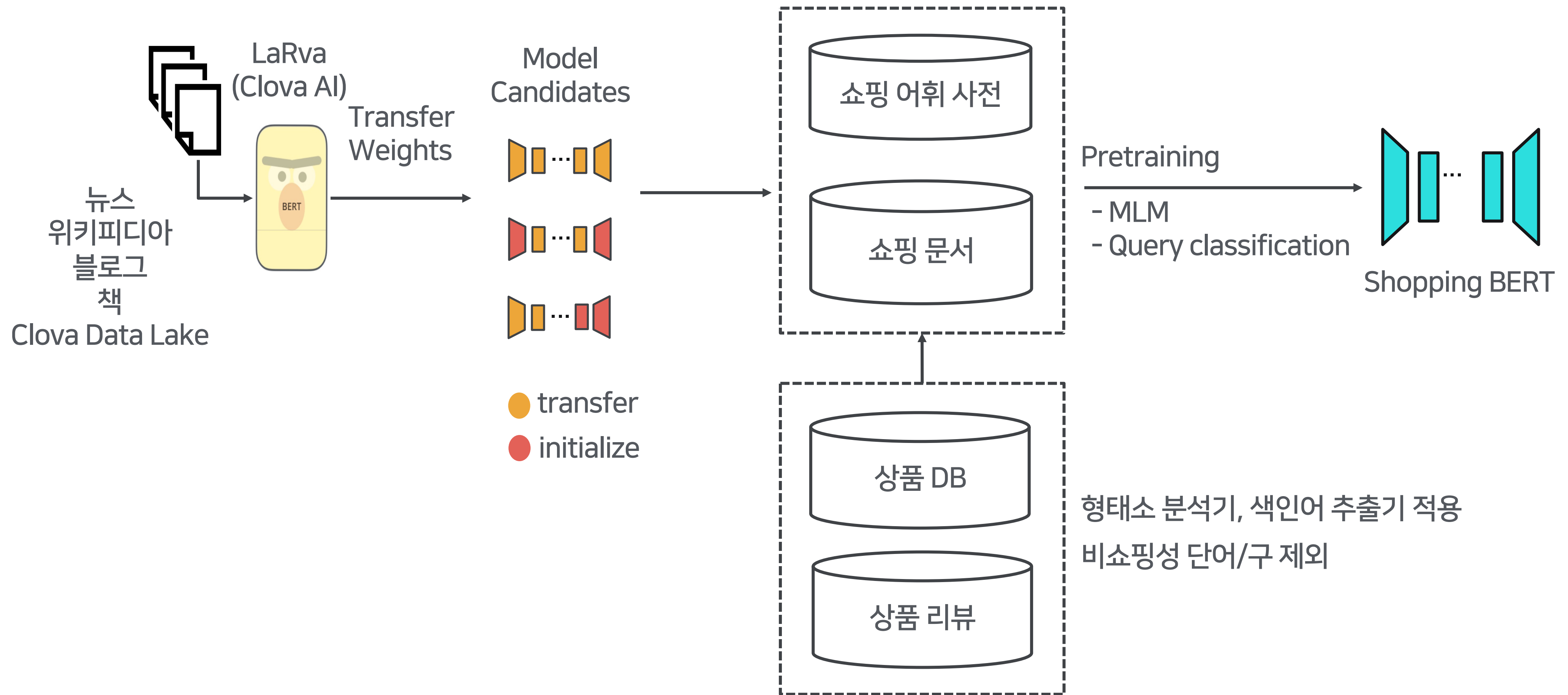
# 4.2 쇼핑 딥서치

## 쇼핑 딥서치 모델 개요



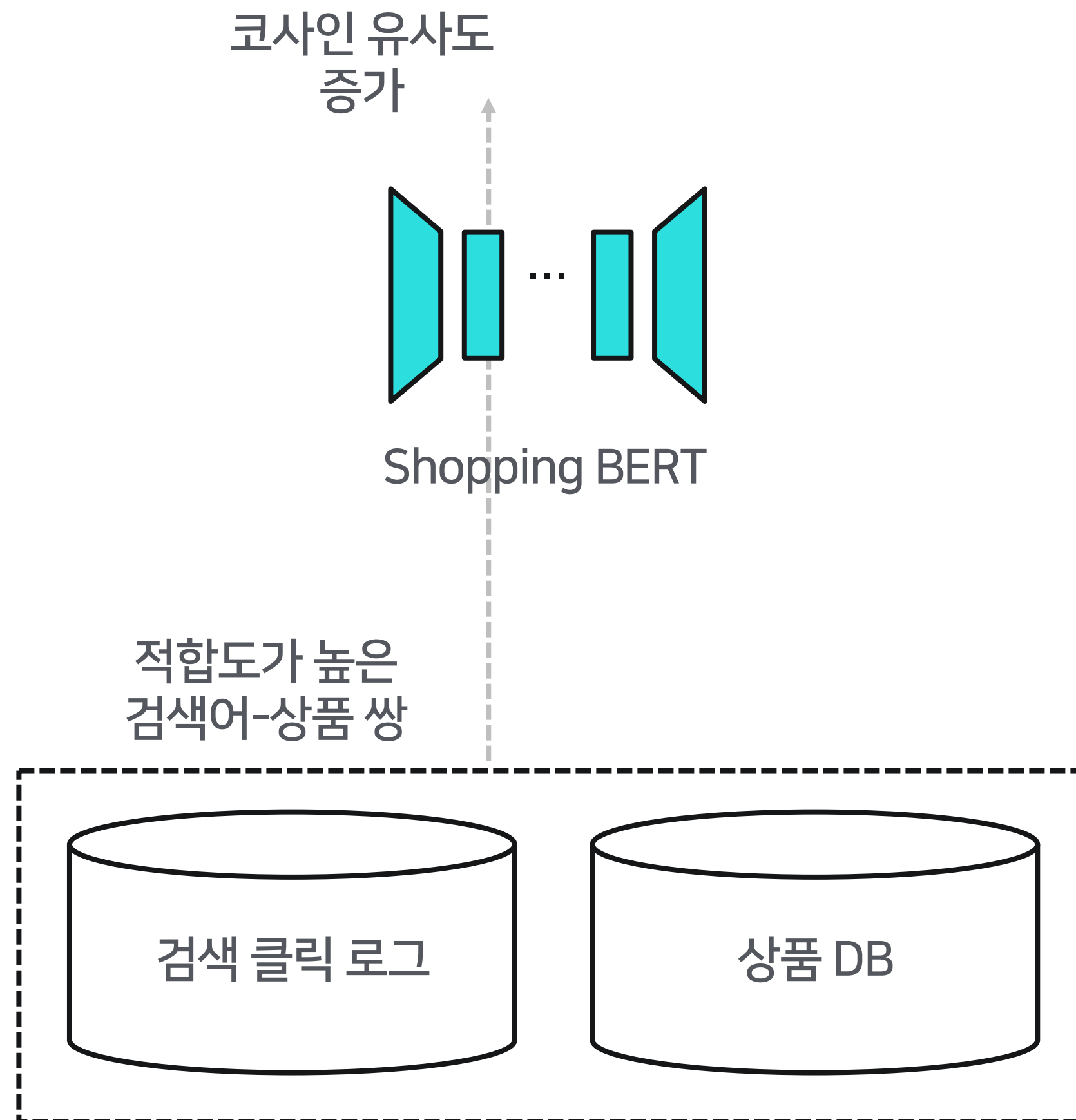
# 4.2 쇼핑 딥서치

## Shopping BERT: 쇼핑 도메인 텍스트 학습



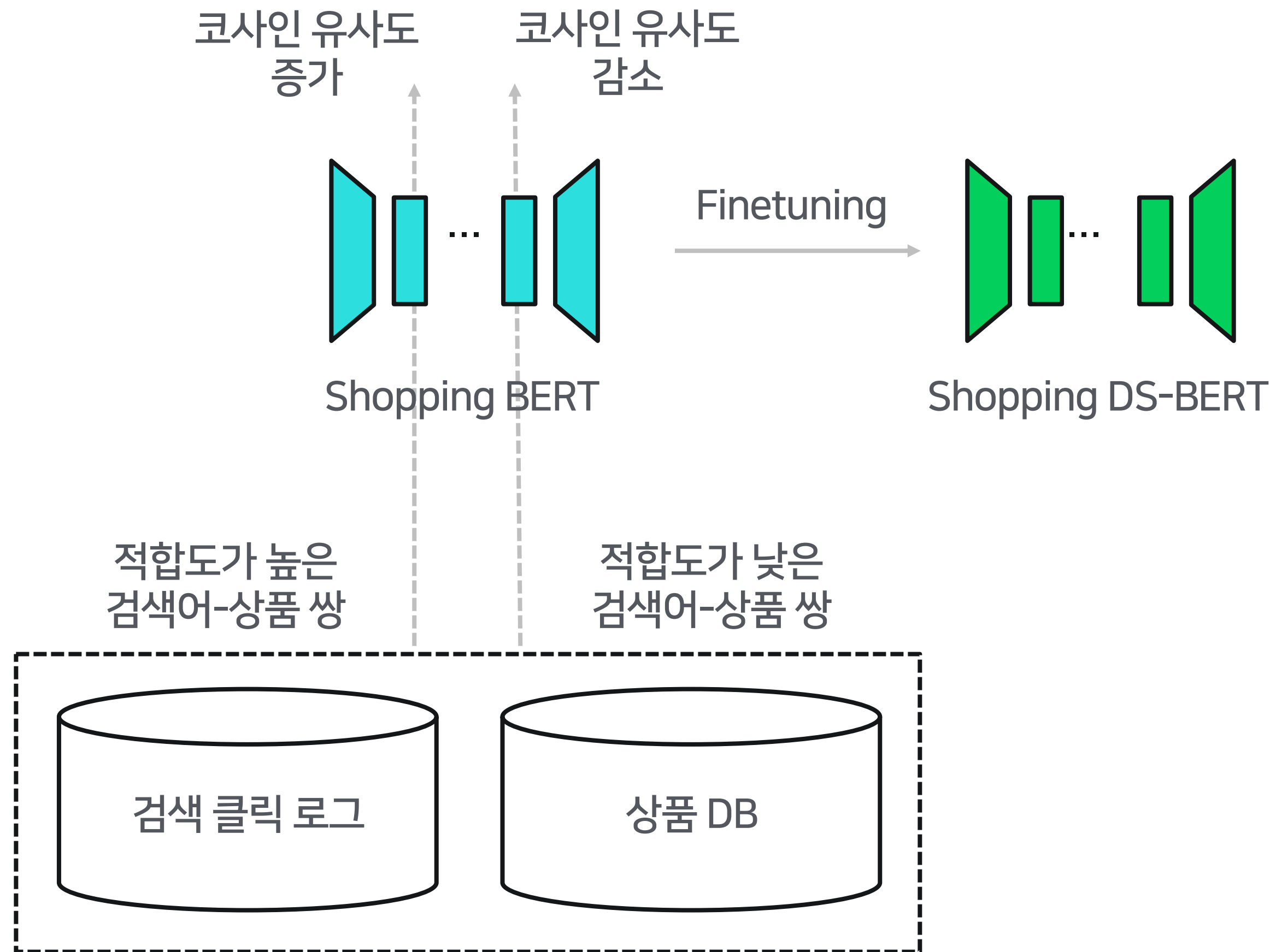
# 4.2 쇼핑 딥서치

Shopping DS-BERT: 검색어/상품 임베딩 모델 학습



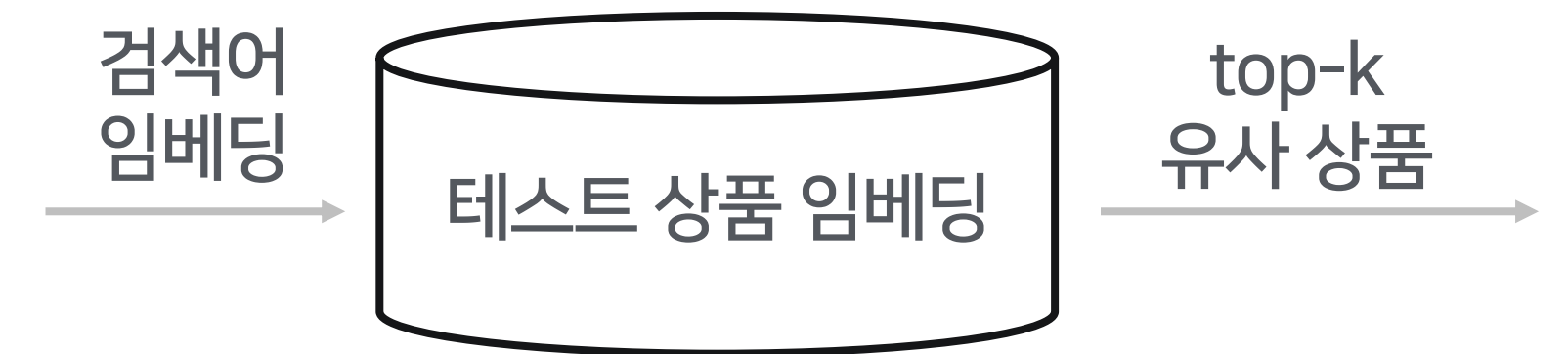
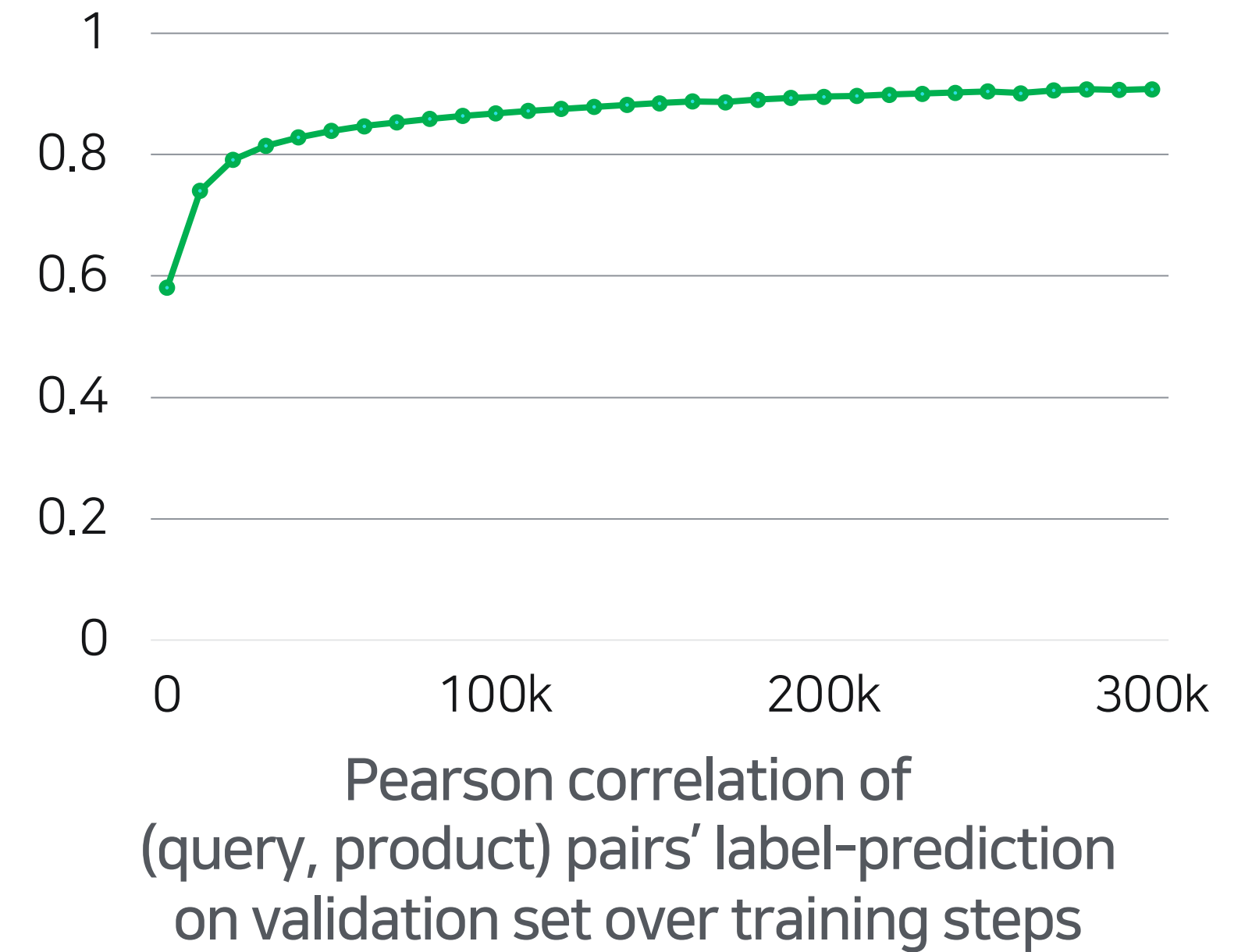
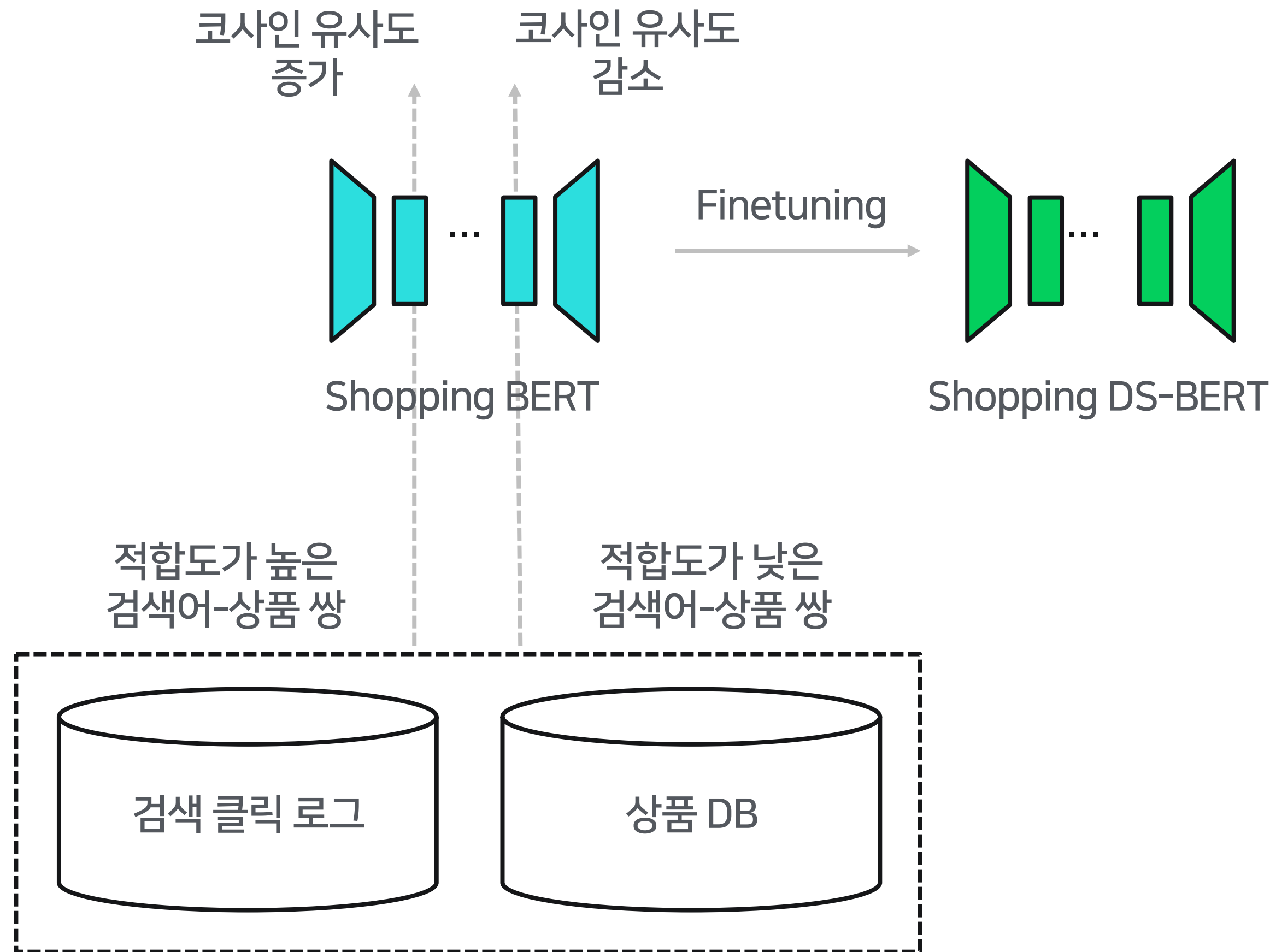
# 4.2 쇼핑 딥서치

## Shopping DS-BERT: 검색어/상품 임베딩 모델 학습



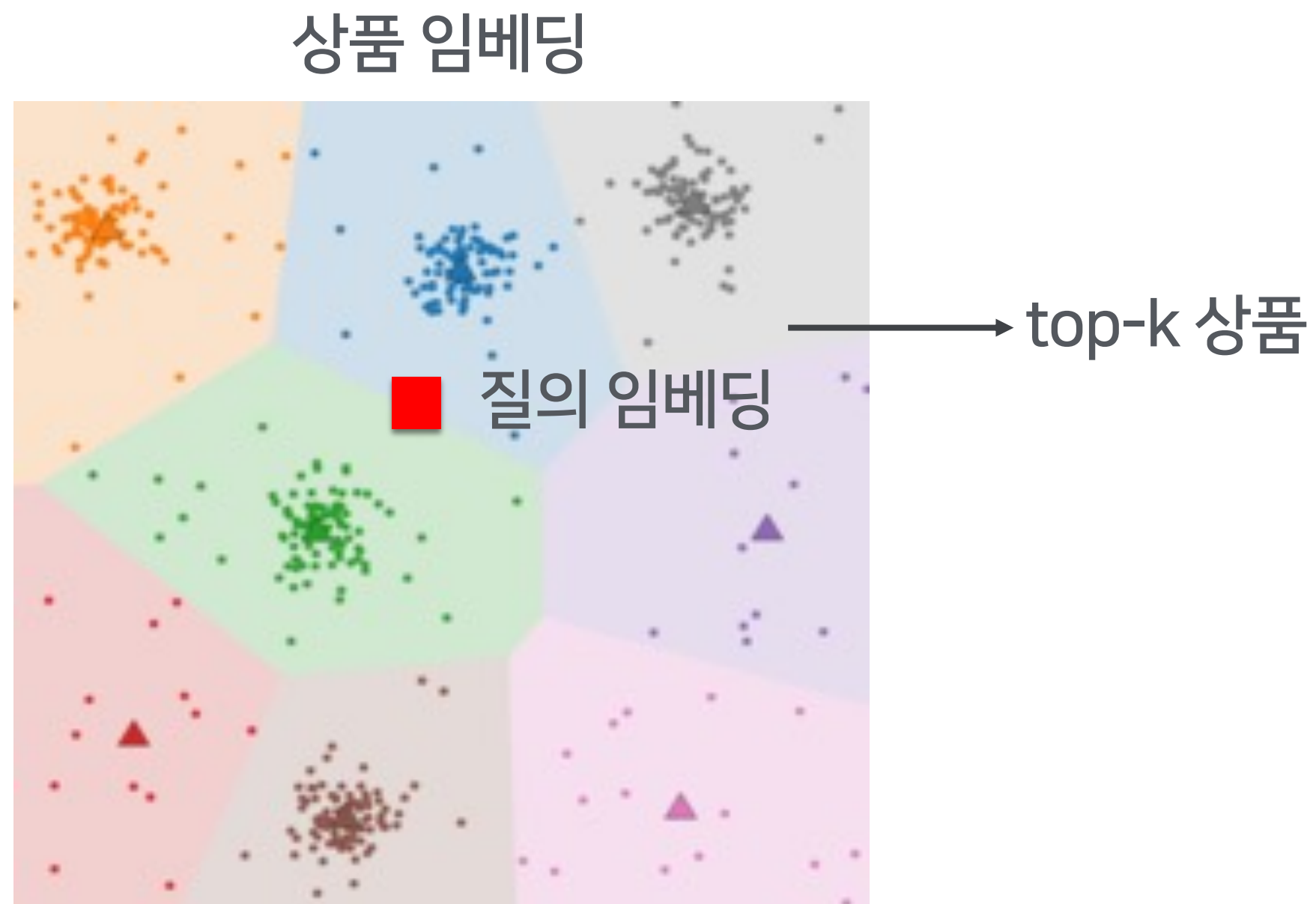
# 4.2 쇼핑 딥서치

## Shopping DS-BERT: 검색어/상품 임베딩 모델 학습



# 4.2 쇼핑 딥서치

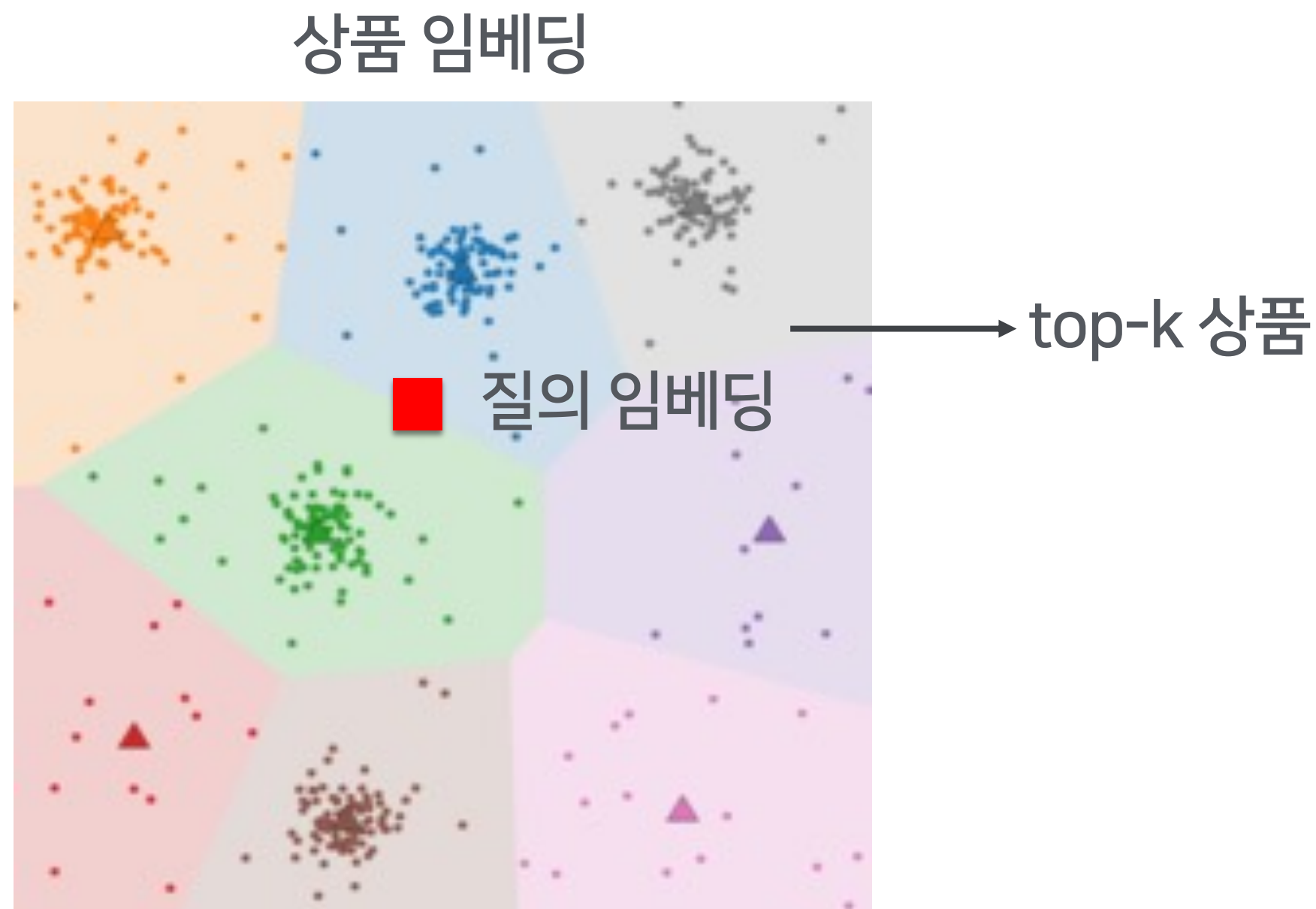
ANN: Approximate Nearest Neighbor 검색



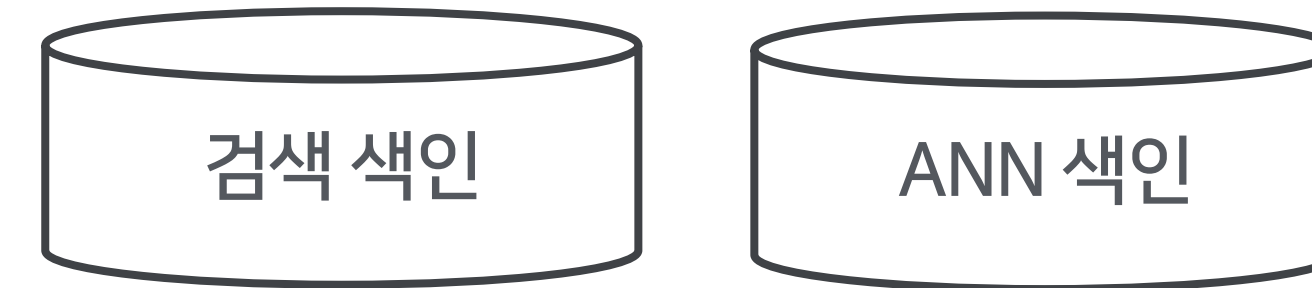
- 부분 탐색 (예: k-means clustering)
- 유사도 근사치 계산 (예: PQ)

# 4.2 쇼핑 딥서치

ANN: Approximate Nearest Neighbor 검색



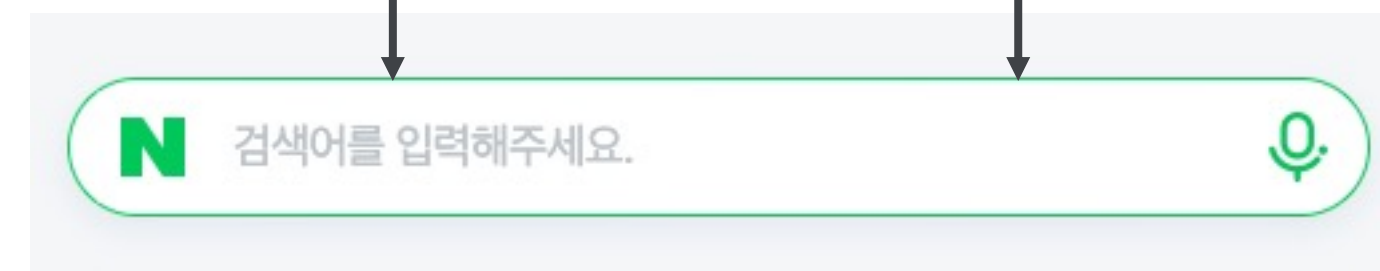
- 부분 탐색 (예: k-means clustering)
- 유사도 근사치 계산 (예: PQ)



ANN 색인의 최신성 유지에 많은 비용 발생

쇼핑성 질의

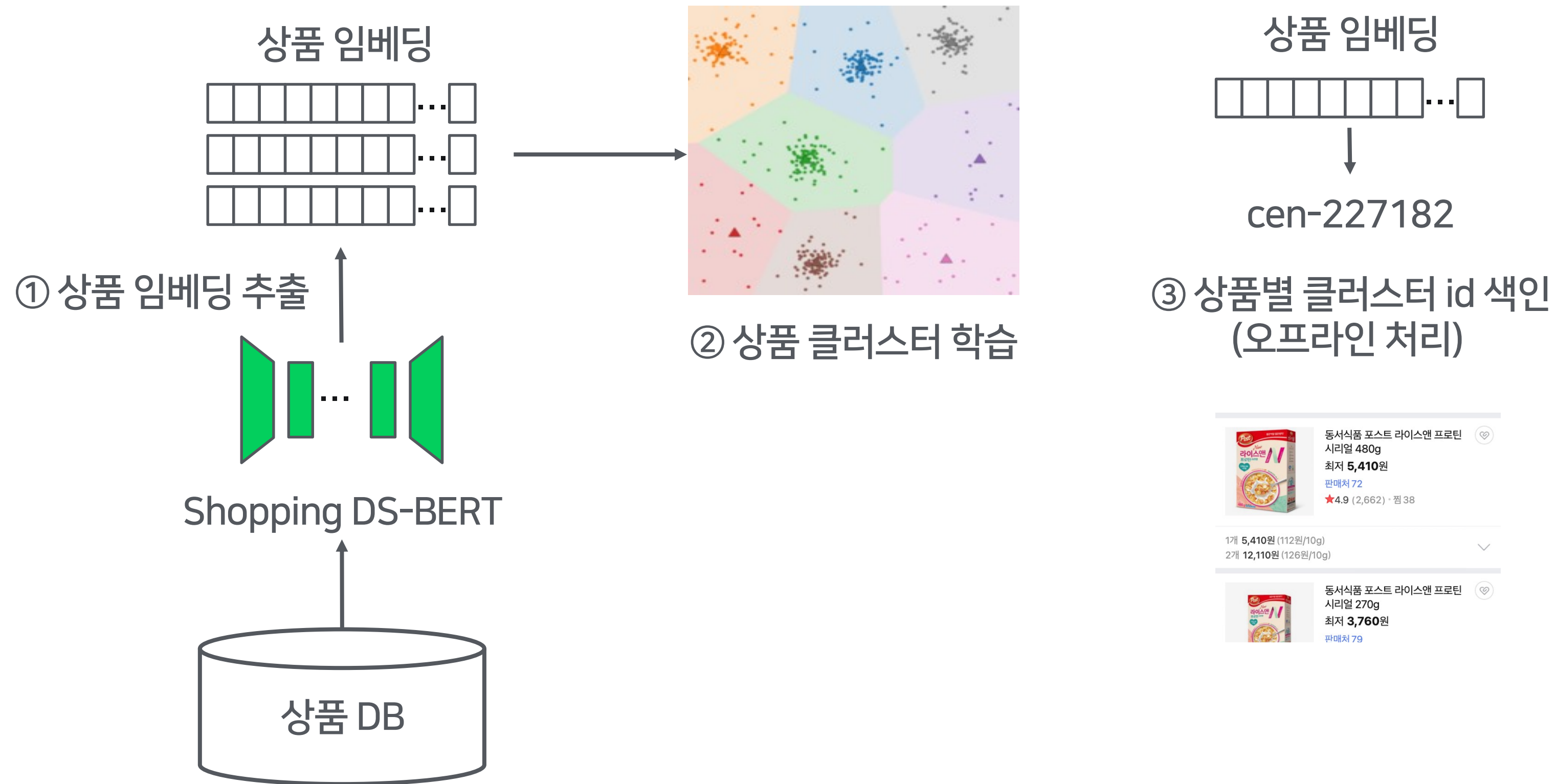
비쇼핑성 질의



단순히 가장 유사한 k개의 상품을 제공할 경우 품질 이슈 발생 가능

# 4.2 쇼핑 딥서치

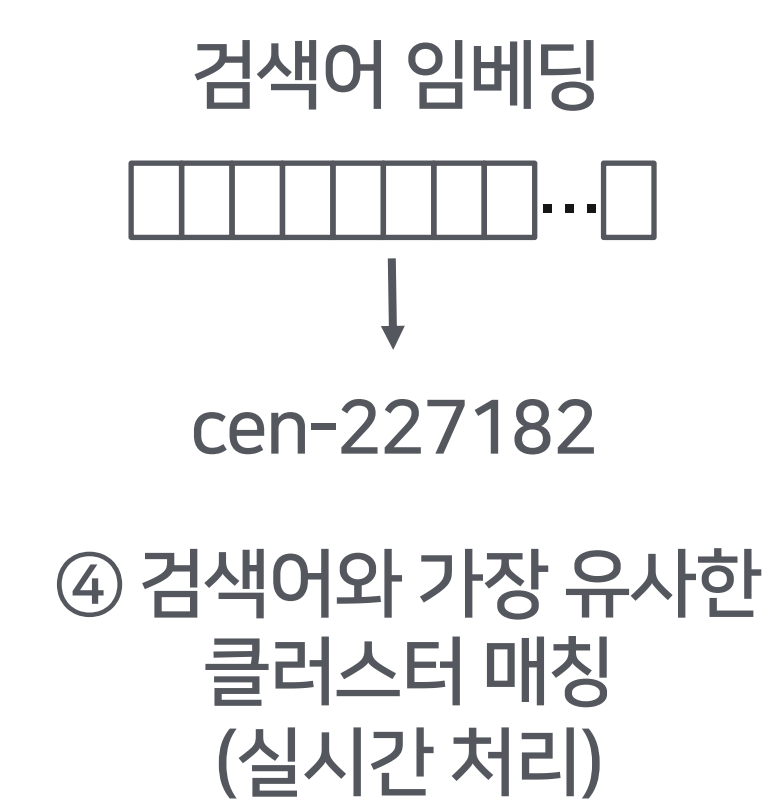
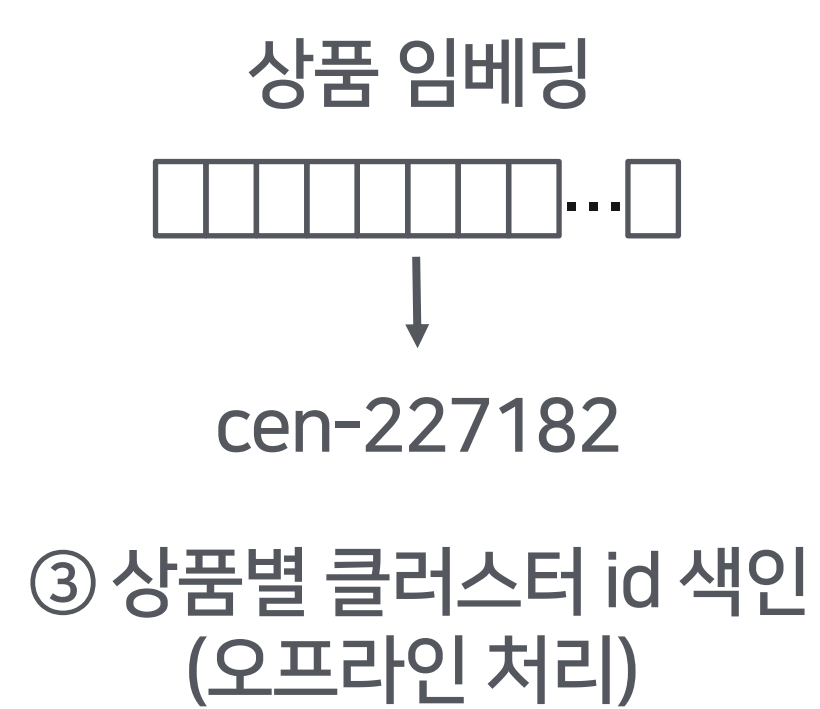
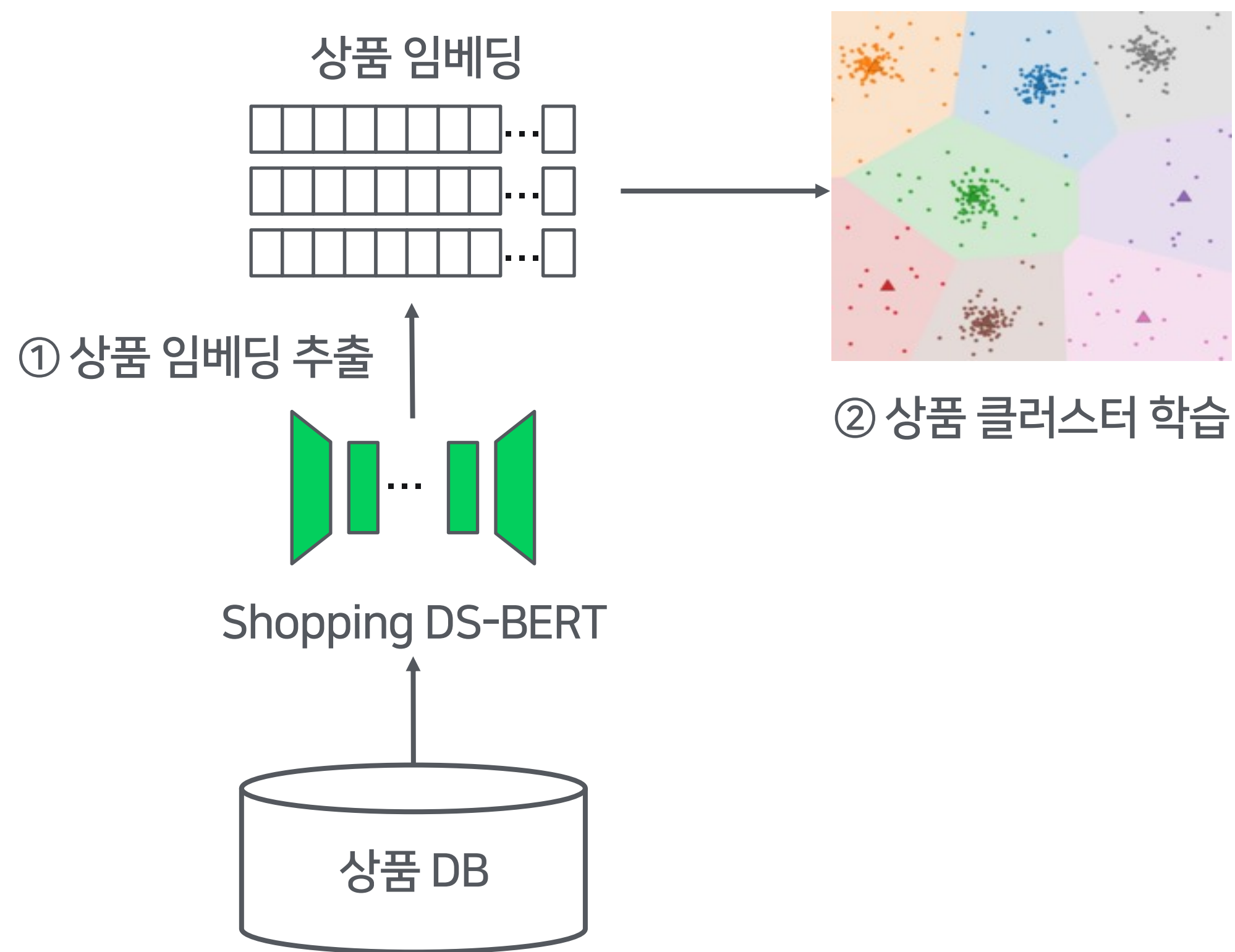
## 검색어-상품 매칭 및 쇼핑 딥서치 품질 개선





# 4.2 쇼핑 딥서치

## 검색어-상품 매칭 및 쇼핑 딥서치 품질 개선



# 4.2 쇼핑 딥서치

## 검색어-상품 매칭 및 쇼핑 딥서치 품질 개선



# 4.2 쇼핑 딥서치

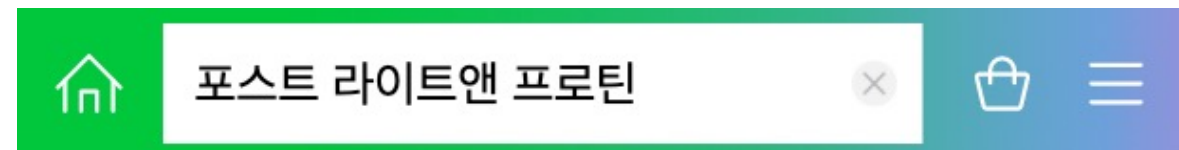
## 검색어-상품 매칭 및 쇼핑 딥서치 품질 개선



# 4.2 쇼핑 딥서치

## A/B 테스트 진행 및 연내 서비스 적용 예정

상품명 표기 오류



검색어와 일치하는 상품이 없어 연관 상품을 노출합니다.

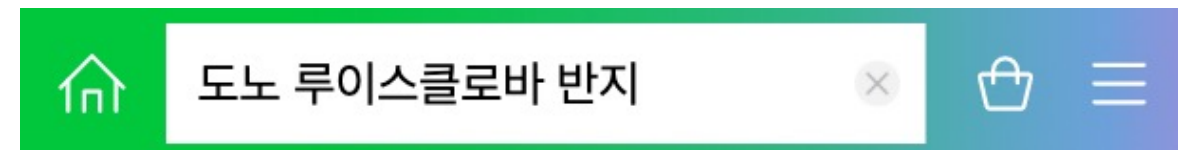
 동서식품 포스트 라이트앤 프로틴 시리얼 480g  
 최저 **5,410원**  
 판매처 72  
 ★4.9 (2,662) · 점 38

1개 5,410원 (112원/10g)  
 2개 12,110원 (126원/10g)

 동서식품 포스트 라이트앤 프로틴 시리얼 270g  
 최저 **3,760원**  
 판매처 79  
 ★4.7 (99) · 점 4

1개 3,760원 (139원/10g)  
 2개 9,980원 (184원/10g)

외래어 표기 불일치



검색어와 일치하는 상품이 없어 연관 상품을 노출합니다.

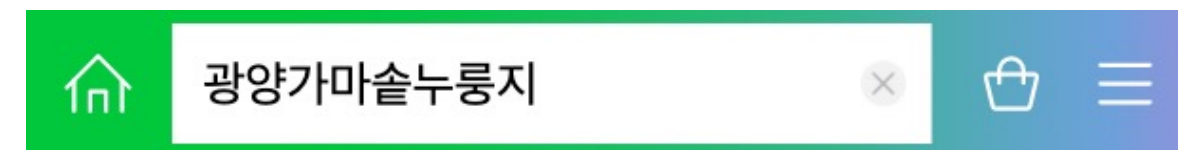
 도노골드 루이스 반지 DONO G  
 최저 **457,000원**  
 판매처 2

성별 여성용 · 주요형태 플랫폼  
 장식형태 기타

 도노골드 이베미스 루이스 반지 DONO G  
 최저 **498,000원**  
 판매처 2  
 ★5 (3) · 점 20

 도노골드 플라베 청다이아 클로버 반지 DONO G

수급되지 않은 상품



검색어와 일치하는 상품이 없어 연관 상품을 노출합니다.

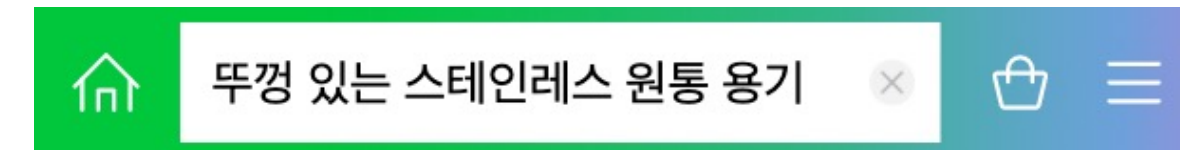
 구가네 국내산쌀100 가마솥맛 햅쌀 누룽지 3kg  
 최저 **1,600원**  
 판매처 55  
 ★4.6 (1,517) · 점 142

맛 다른 누룽지와 달리 부스러기가 적어서 더욱 좋았습니다  
 유통기한 유통 기한도 길고 쌀 외엔 다른 것이 안 섞인 점이 더...

 참좋은식품 가마솥의 누룽지 150g  
 최저 **1,540원**  
 판매처 93  
 ★4.8 (2,347) · 점 27

맛 귀찮을때 누룽지 과자처럼 가지고 다니면서 먹기도 좋구 딱...  
 품질 쌀만큼 안떨어지게 해놔요

자연어 형태의 긴 검색어



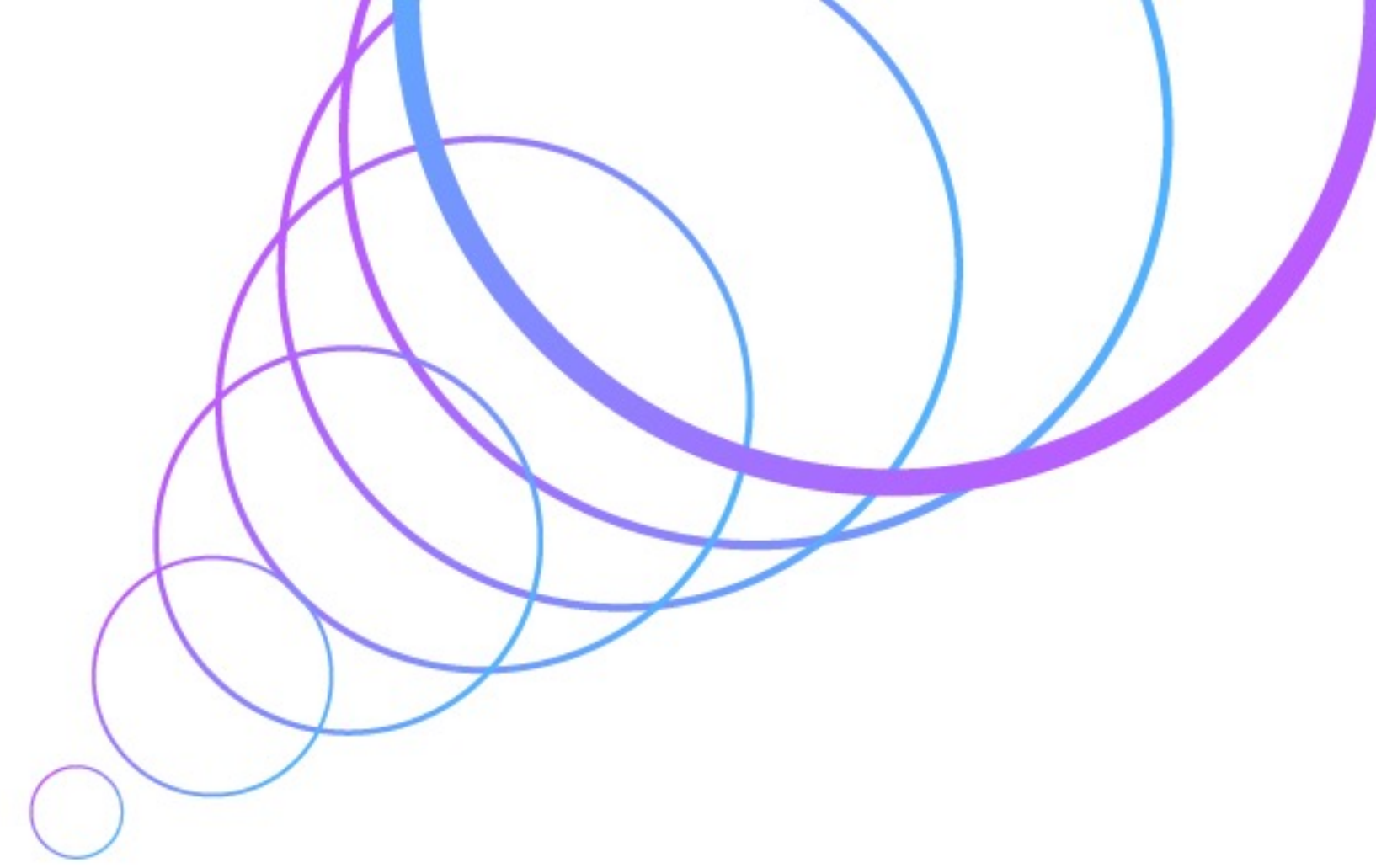
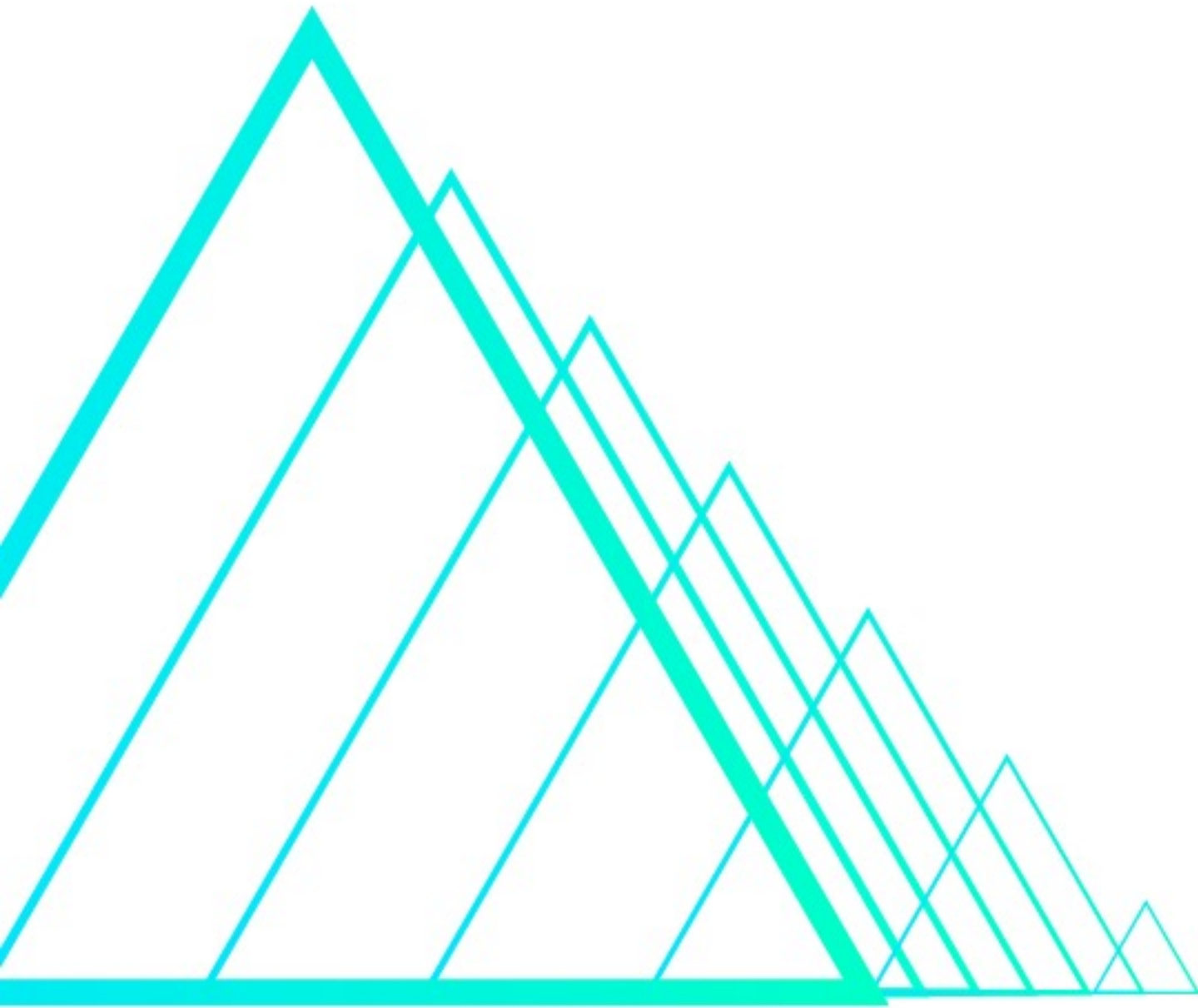
검색어와 일치하는 상품이 없어 연관 상품을 노출합니다.

 그릭요거트 용기 스텐 스텐레스 원형 반찬통 찬통 원통 밀폐 보관...  
**3,950원**  
 집순이스토어 N Pay +  
 구매 6 · 리뷰 1 · 점 2

포인트 네이버페이 포인트 **39원**  
 배송 3,000원

 원형 락스텐 밀폐용기 1.0ℓ 김치통 반찬통 도시락 원통 뚜껑 스텐용...  
**7,100원**  
 씨에스동양상사 N Pay +  
 점 5

포인트 네이버페이 포인트 **81원**  
 배송 2,500원 · 오늘출발 (평일 12:30 주문 마감)



**Thank You**

