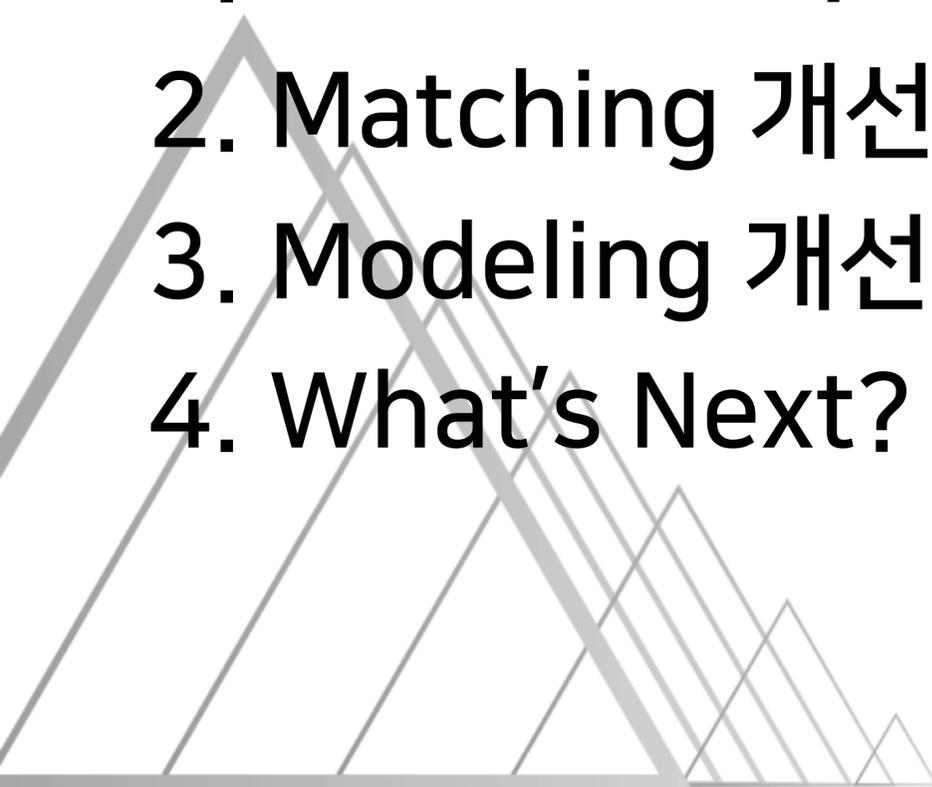


생활의달인에 나온 과배기 맛집 알고싶으세요? 질의 의미를 파악해 찾아주는 Deep Local Search

CONTENTS

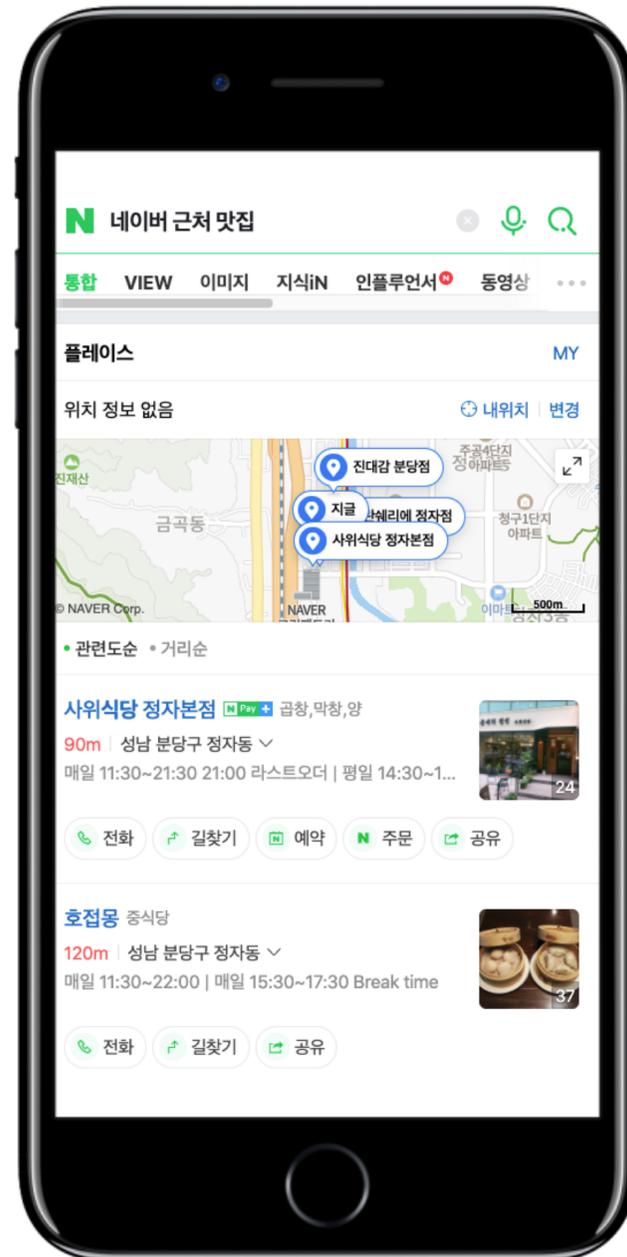
1. What to do?
2. Matching 개선을 통한 정답형 질의 확장
3. Modeling 개선을 통한 정답형 질의 확장
4. What's Next?



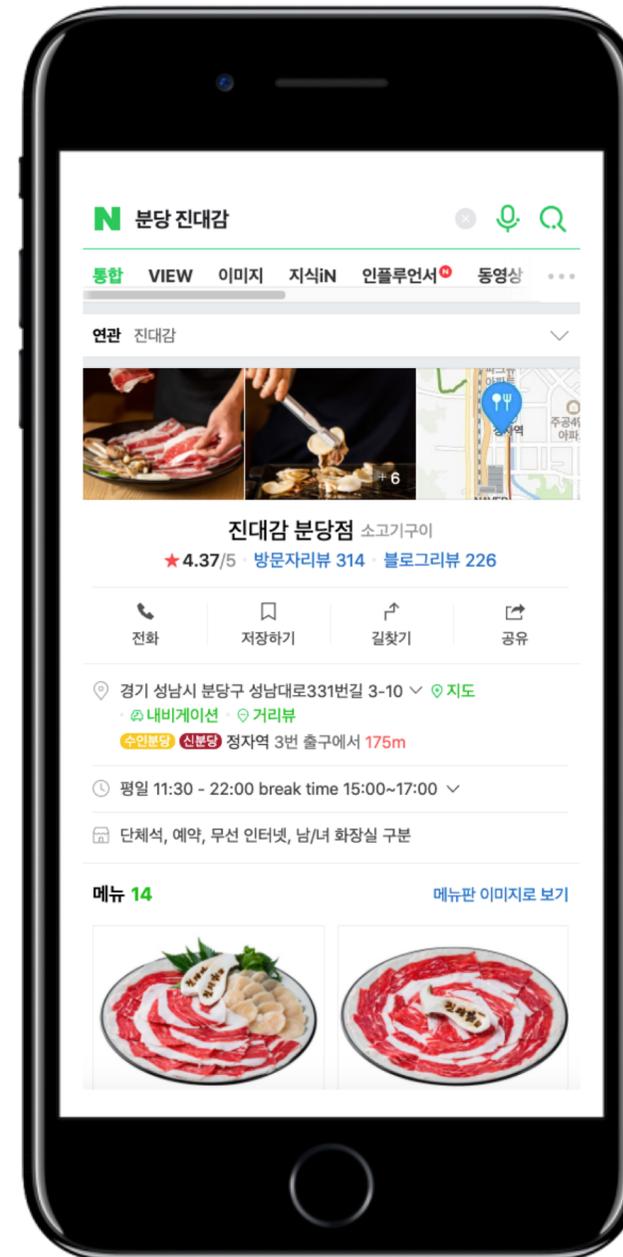
1. What to do?

1.1 지역 검색의 구분

목록형



정답형



1.2 Q&A in 지역 검색

N 미금 강남교자



N 김해 chef1989



N 경성양꼬치 정자점



N 정자 크로와상랩



N 생활의달인 콩고물파베기



베이커리울교p3120 베이커리
★ 4.44/5 · 방문자리뷰 508 · 블로그리뷰 110

전화 저장하기 길찾기 공유

예약 **주문**

🕒 지금은 네이버 주문 준비 중입니다

🏠 **코로나 상생 국민지원금** 사용처입니다. [자세히](#)

📍 인천 남동구 예코중앙로156번길 13-17 지도 내비게이션 거리뷰

🕒 평일 09:00 - 21:00 —

📄 단체석, 주차, 포장, 예약, 무선 인터넷, 남/녀 화장실 구분, 국민지원금

📅 생활의달인 663회, 19.03.25. 콩고물파베기

메뉴	포장	매장
coffee ☑️		2,500원
cafe latte ☑️		3,500원
choco latte		4,500원
tea ☑️		3,500원

1.3 초창기 지역검색 Q&A 대응은?

정답형 질의 수동 선정

- 수동으로 3만개 정답형 질의 선정
- 문제점
 - 업체 변경 대응의 어려움 (개업, 폐업, 업체 이름 변경, ..)
 - 방송 출연, 신생 단어 등 trendy 대응의 어려움

1.4 현재 지역 검색 Q&A 대응 방법

정답형 질의 추출 기준

- 같은 질의를 했을 때 같은 답을 얻길 원하는가?
- 질의에 대한 답이 적절한가?

1.5 같은 질의를 했을 때 같은 답을 얻길 원하는가? N DEVIEW 2021



101 남산



A 101번지 남산돈까스 본점 101번지남산돈까스
1.3km | 서울 중구 남산동2가
매일 10:30~22:00 연중무휴

공이미식회 | 인플루언서 | 2021.02.15.
[남산]101번지 남산돈까스, 경양식돈까스의 베이직함.(남산데이트,케이블카,오락실)
101번지 남산돈까스는 다른 지점도 생겼고.. 그런다고 다른 돈까스 집 가는것도 좀..... ㅎㅎ 101 남산 돈까스는 큰 기대 없이 방문하면 맛있게 먹을수 있습니다. 고 급스러움보다는...

새라의 일상블로그 | 2021.02.20.
일산 101번지 남산돈까스
일산 101번지 남산돈까스에서 배터지게 먹었는데요~ 아쉬운 부분도 있었지만 돈까스는 맛있게 잘 먹고 왔어요. #101번지남산돈까스 #남산돈까스 #일산돈까스 #일산줄면

101번지 남산돈까스 본점



101번지 남산돈까스 본점 101번지남산돈까스
★ 4.17/5 · 방문자리뷰 2,199 · 블로그리뷰 967
남산에서 먹는 추억의 왕돈까스

☎ 전화 📌 저장하기 📍 길찾기 📄 공유

홈 소식 메뉴 리뷰 사진 지도 주변

☞ 서울 중구 소파로 101 근 주변 📍 복사 📍 지도 📍 내비게이션 📍 거리뷰

📍 명동역 3번 출구에서 501m

🕒 매일 10:30 - 22:00 연중무휴

📄 단체석, 주차, 발렛파킹, 포장, 예약

101번지 남산돈까스 일산가좌점



101번지남산돈까스 일산가좌점 101번지남산돈까스
★ 4.23/5 · 방문자리뷰 79 · 블로그리뷰 18
📄 농림축산식품부 제공 안심식당 📄

☎ 전화 📌 저장하기 📍 길찾기 📄 공유

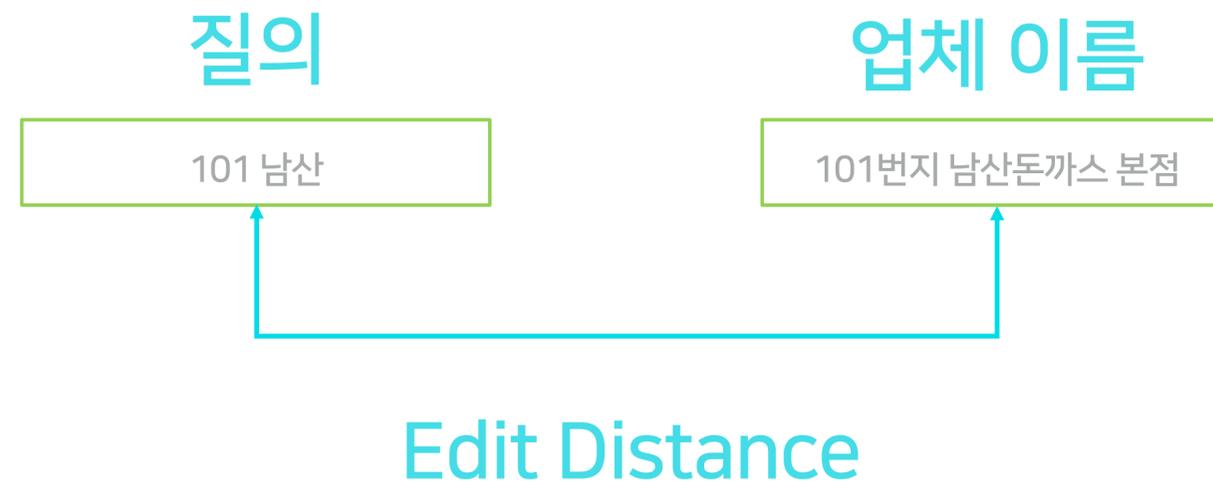
홈 메뉴 리뷰 사진 지도 주변

☞ 경기 고양시 일산서구 송파로 182 101번지남산돈까스 근 주변 📍 복사 📍 지도 📍 내비게이션 📍 거리뷰

🕒 매일 11:00 - 15:30 코로나19 2.5단계 연장으로 인한 1월 31일까지 영 업시간변경, 주문가능!!

📄 단체석, 주차, 포장, 무선 인터넷, 남/녀 화장실 구분, 지역화폐(카드형)

1.6 질의에 대한 답이 적절한가?



1.7 지역 Q&A 질의 자동 추출 과정



101번지 남산돈까스 본점

101번지 남산돈까스 본점 101번지남산돈까스
★ 4.17/5 · 방문자리뷰 2,199 · 블로그리뷰 967
남산에서 먹는 추억의 왕돈까스

전화 저장하기 길찾기 공유

소식 메뉴 리뷰 사진 지도 주변

서울 중구 소파로 101 근 주변 · ④ 거리뷰 · ④ 내비게이션

④ 명동역 3번 출구에서 501m

매일 10:30 - 22:00 연중무휴

단체석, 주차, 발렛파킹, 포장, 예약

미운우리새끼 122회, 19.01.20. 돈가스/생선가스 <

http://www.namsan101.co.kr
페이스북 · 인스타그램

101번지 남산돈까스는 90년대 생긴 남산 최초의 한국식 수제왕돈까스 전문점으로 남산 소파로 101번지에 개업하여 각종 매스컴에 소개되고 입소문을 타 현재 연간 70만명이 넘는 고객이 찾아주시는 국내 1위 ... <

101번지 남산돈까스 본점

101번지남산돈까스 본점 101번지남산돈까스
★ 4.17/5 · 방문자리뷰 2,199 · 블로그리뷰 967
남산에서 먹는 추억의 왕돈까스

전화 저장하기 길찾기 공유

소식 메뉴 리뷰 사진 지도 주변

서울 중구 소파로 101 근 주변 · ④ 거리뷰 · ④ 내비게이션

④ 명동역 3번 출구에서 501m

매일 10:30 - 22:00 연중무휴

단체석, 주차, 발렛파킹, 포장, 예약

101번지남산돈까스 일산가좌점 101번지남산돈까스
★ 4.23/5 · 방문자리뷰 79 · 블로그리뷰 18
농림축산식품부 제공 안심식당 >

전화 저장하기 길찾기 공유

메뉴 리뷰 사진 지도 주변

경기 고양시 일산서구 송파로 182 101번지남산돈까스 근 주변 · ④ 복사 · ④ 지도 · ④ 내비게이션 · ④ 거리뷰

매일 11:00 - 15:30 코로나19 2.5단계 연장으로 인한 1월 31일까지 영업시간변경, 주문가능!!

단체석, 주차, 포장, 무선 인터넷, 남/녀 화장실 구분, 지역화폐(카드형)

101번지 남산돈까스 본점 101번지남산돈까스
1.3km | 서울 중구 남산동2가 <
매일 10:30~22:00 연중무휴

전화 길찾기 공유

퐁이마식회 | 인플루언서 | 2021.02.15.
[남산]101번지 남산돈까스, 경양식돈까스의 베이직함.(남산데이트,케이볼카,오락실)
101번지 남산돈까스는 다른 지점도 생겼고.. 그러다그 다른 돈까스 집 가는것도 좀..... ㅎㅎ 101 남산 돈까스는 큰 기대 없이 방문하면 맛있게 먹을수 있습니다. 고급스러움보다는...

#101번지남산돈까스 #남산돈까스 #왕돈까스 #남산 #남산데이트

새라의 일상블로그 > | 2021.02.20.
일산 101번지 남산돈까스
일산 101번지 남산돈까스에서 배터지게 먹었는데요~ 아쉬운 부분도 있었지만 돈까스는 맛있게 잘 먹고 왔어요. #101번지남산돈까스 #남산돈까스 #일산돈까스 #일산풀면

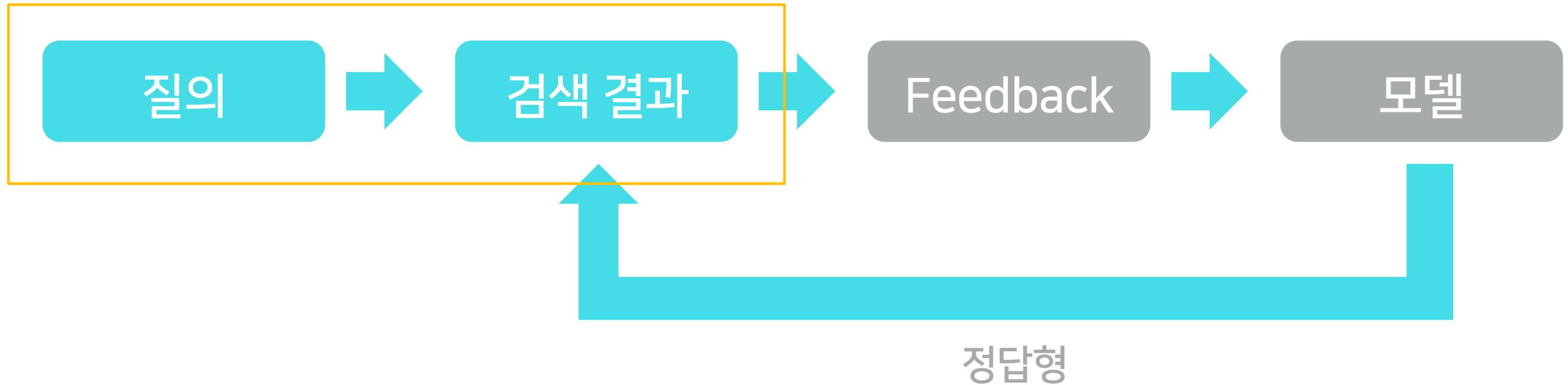
#101번지남산돈까스 #남산돈까스 #일산돈까스 #일산풀면

1.8 지역 정답형 질의 자동 추출 과정

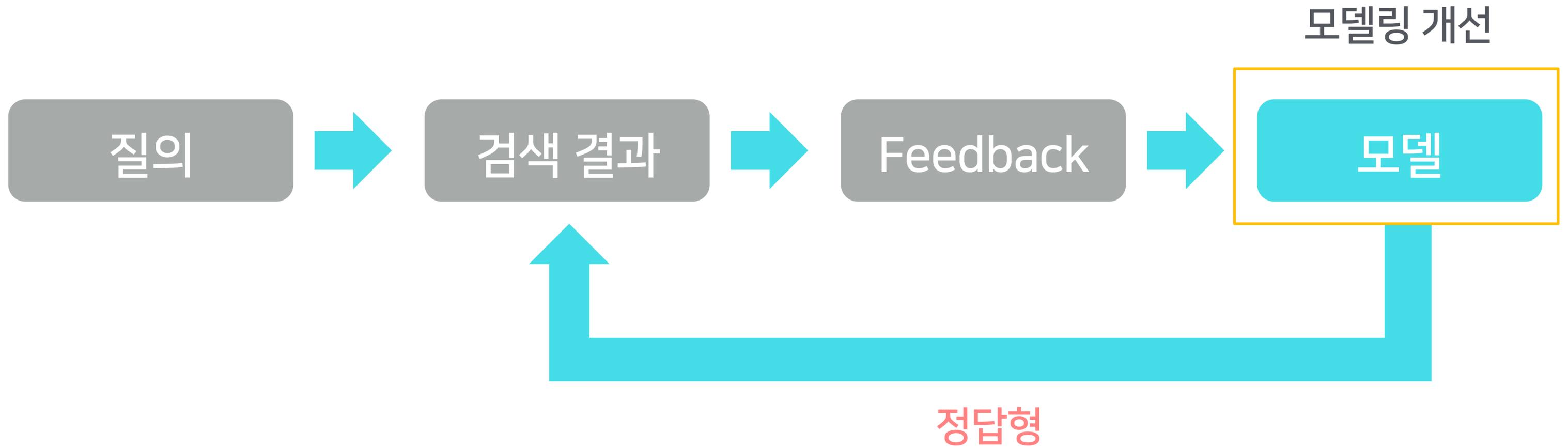


1.9 매칭 개선

질의 이해를 통한 매칭 개선



1.10 모델링 개선



2. Matching 개선을 통한 정답형 질의 확장

2.1 매칭 개선 방법

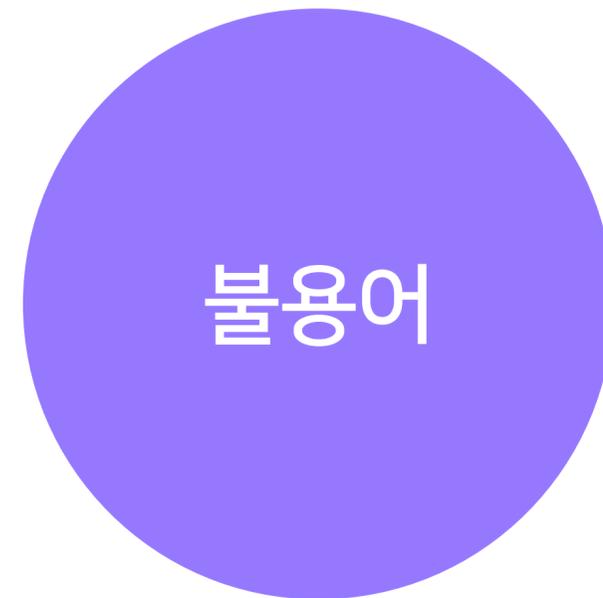
질의의 의미를 이해해 질의를 변경

- 질의에서 의미적으로 불필요하고 검색에 방해되는 단어를 찾아 제거
- 댕댕이 병원으로 검색하면 의미적으로 같은 강아지 병원으로도 검색

2.2 질의 변경 (불용어 제거)

불용어 제거를 통한 매칭 개선

- 질의에서 불용어를 어떻게 판별할까?



2.3 인덱싱 시 제거될 단어 판별

검색 질의의 일부 단어 제거 후 검색 & 클릭

- 제거된 단어 : 불용어

🔍 | ✕

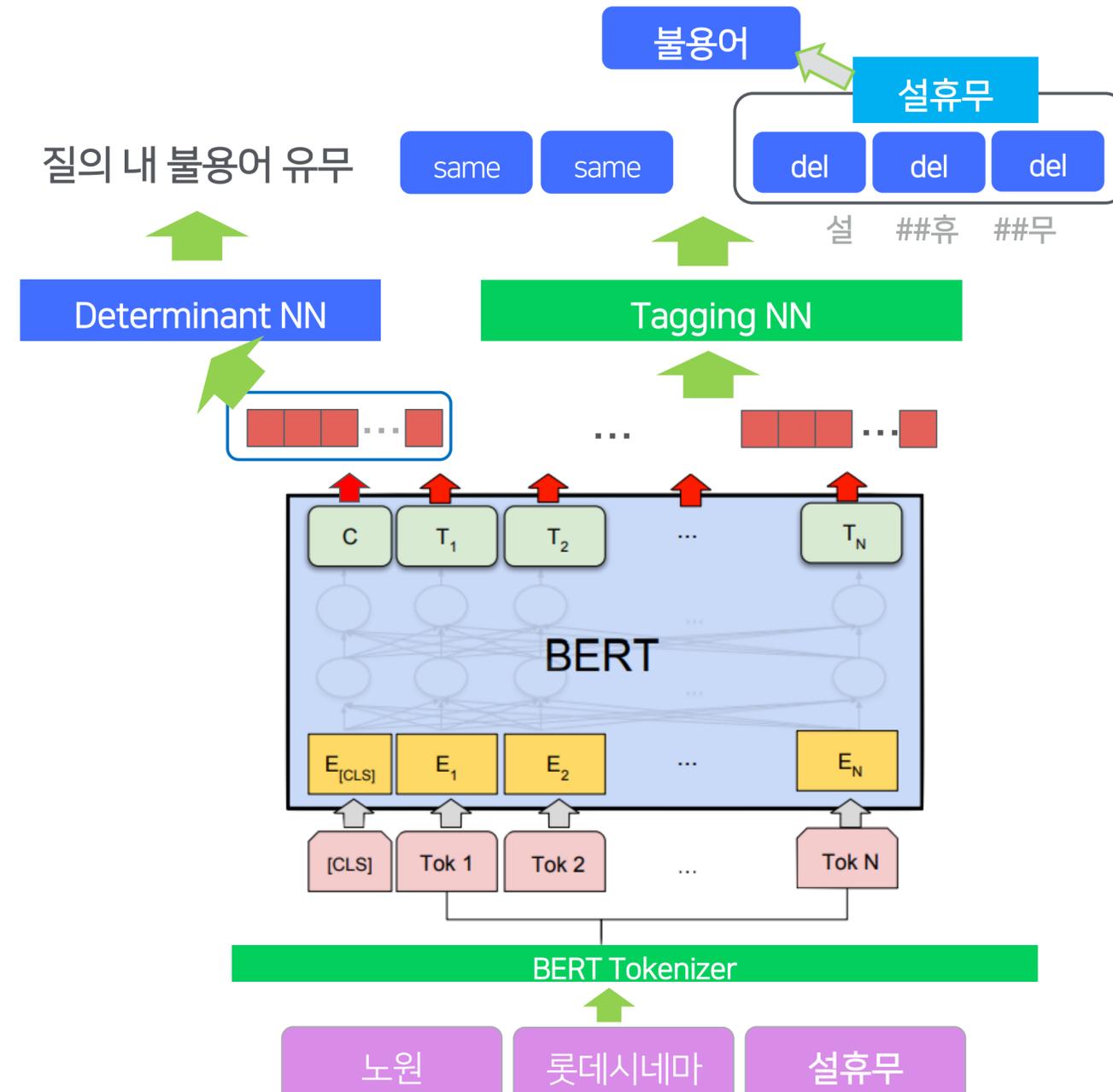


🔍 | ✕

ⓘ
조건에 맞는 업체가 없습니다.

우리은행
우리은행 성남중앙지점 은행
방문자리뷰 2 · 블로그리뷰 1
출발 도착
저장하기 거리뷰 공유
홈 리뷰 사진
☎ 031-751-4811 📄 복사

2.4 BERT 기반의 질의 불용어 판별 모델



2.5 학습 데이터

단어 제거 예시

제거 전 질의	제거 후 질의
수원 간사이등심 번호	수원 간사이등심
우리은행 성남중앙지점 번호	우리은행 성남중앙지점
노원 롯데백화점 설휴무	노원 롯데백화점
제주 옷 롯데마트	제주 롯데마트



제거 전 질의	제거 후 질의	Tag
수원 간사이등심 번호	수원 간사이등심	S S D
우리은행 성남중앙지점 번호	우리은행 성남중앙지점	S S D
노원 롯데백화점 설휴무	노원 롯데백화점	S S D
제주 옷 롯데마트	제주 롯데마트	S D S

S : Same, D : Delete



Insert 태그 제거 및 tokenizing

제거 전 질의	Tag
수원 간 ##사이 ##등심 번호	S S S S D
우리 ##은행 성남 ##중 ##앙 ##지 ##점 번호	S S S S S S S D
노원 롯데 ##백 ##화 ##점 설 ##휴 ##무	S S S S S D D D
제주 옷 롯데 ##마트	S D S S

2.6 질의 불용어 판별 모델 실험 결과

Model	Test
Determinant	93.75%
Tagging	91.42%

질의	추출 불용어	추론 태그
수원 간사이등심 번호	번호	same same del
춘천 노브랜드 현금	현금	same same del
제주 옷 롯데마트 이마트	옷	same del same same
경주 아리랑양꼬치	X	same same
안성스타필드주차	X	same
김포 아이손컴퍼니	X	same same

2.7 질의 변경 (동의어 확대)

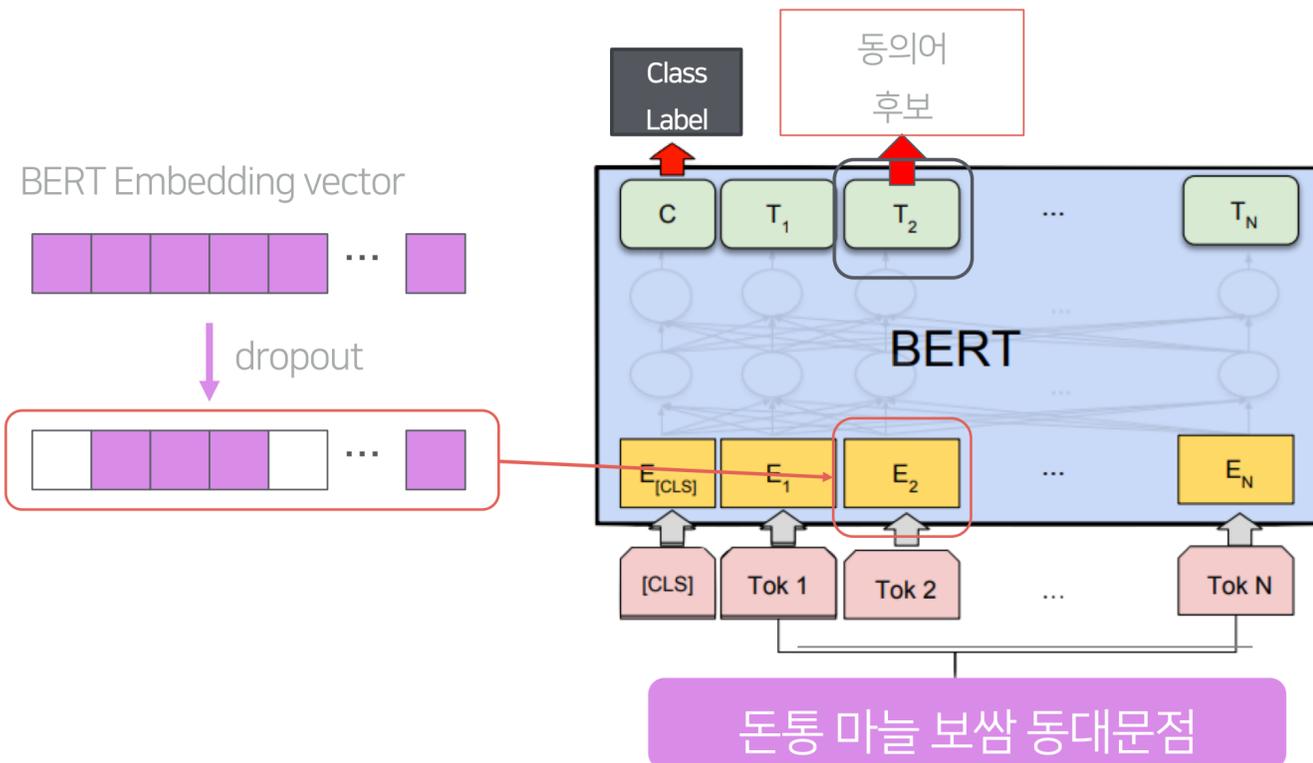
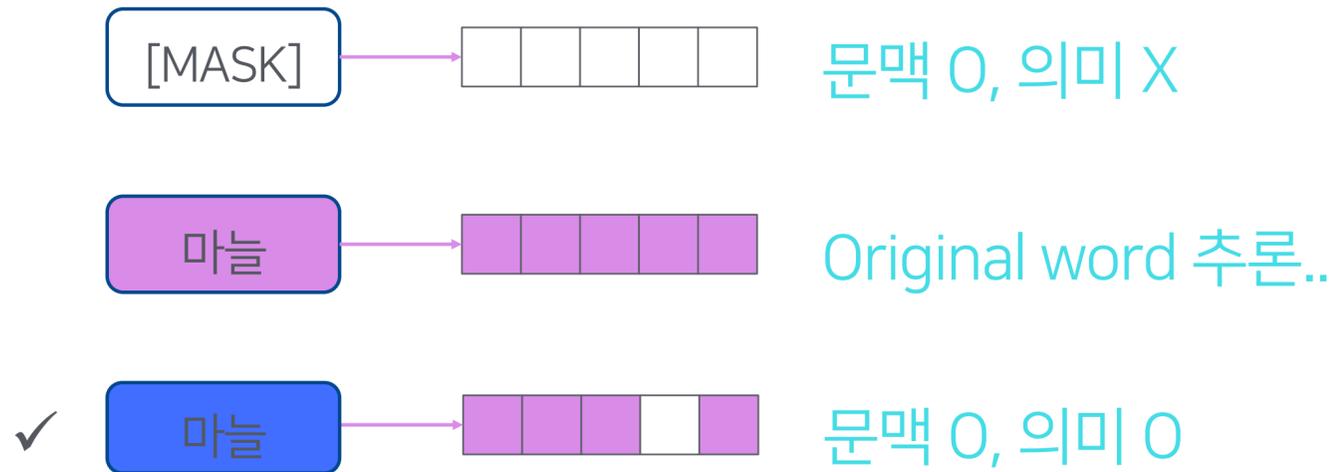
동의어 확대를 통한 매칭 개선

- 두 단어가 동의어인지 어떻게 판단할까?



2.8 BERT-based lexical substitution

- Dropout input encoding

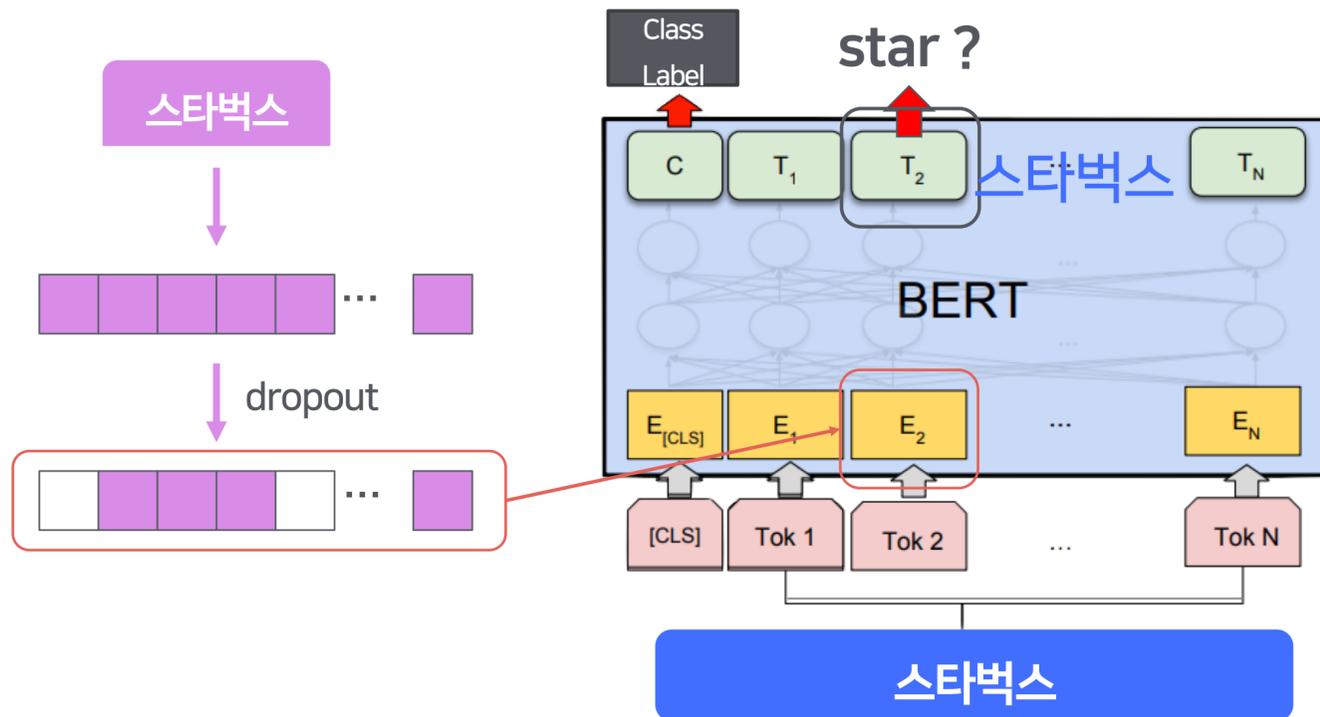


동의어 사전

컷	컷트
농장	농원
쇼파	쇼파
삼겹	삼겹살
마늘	갈릭
	...

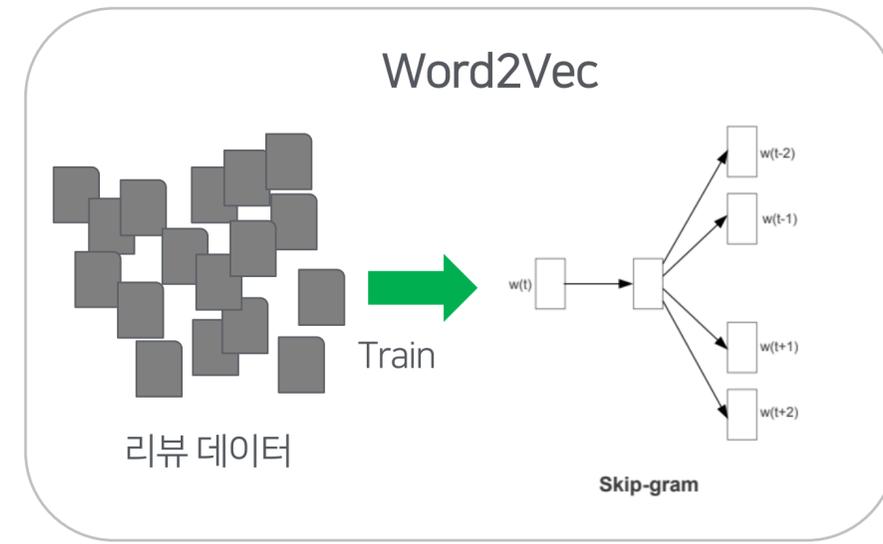
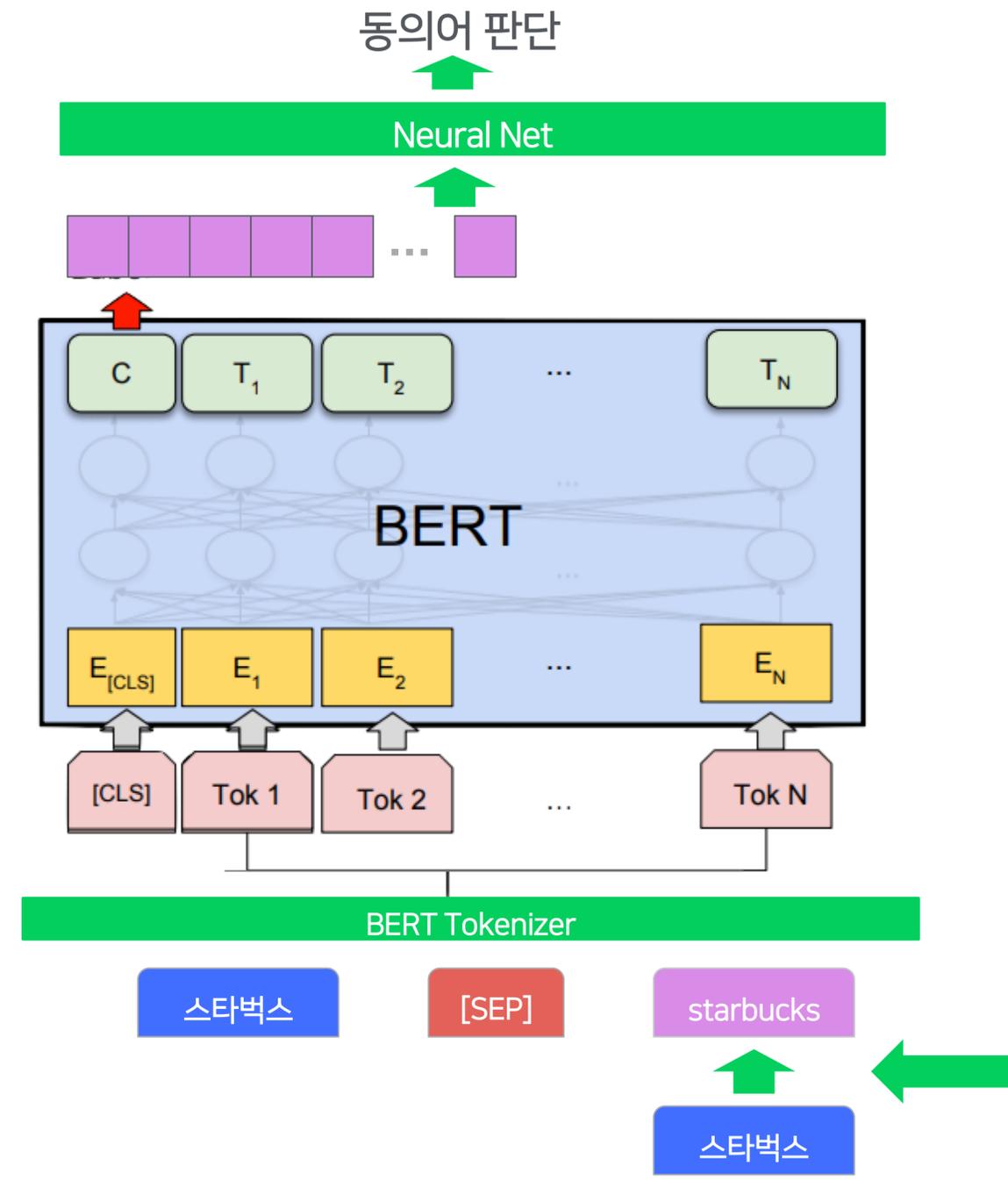
2.9 문제점

- Tokenizing problem!!



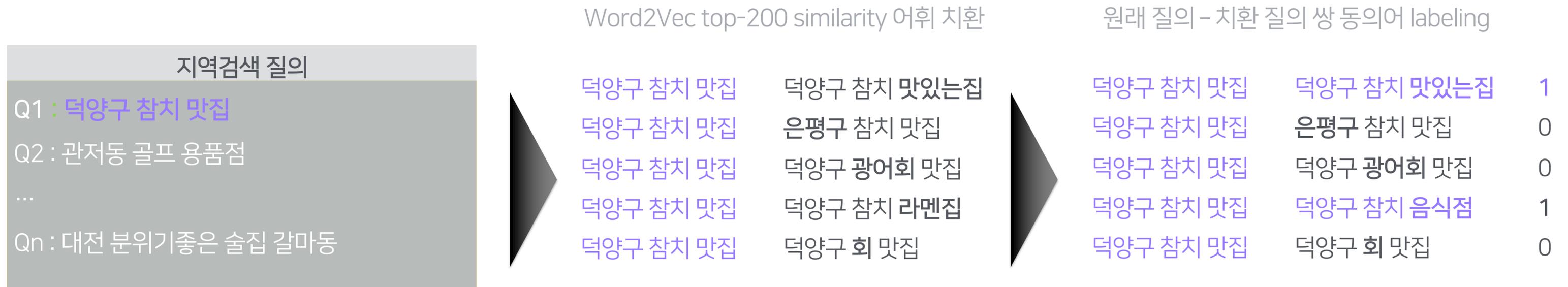
스타벅스는 star와 동의어가 될 수 없음

2.10 동의어 판단 모델 디자인



2.11 학습 데이터 구성

1. word2vec를 이용해 원래 질의 - 치환 질의 pair 구축
2. 동의어 사전을 이용해 labeling



어휘	동의어
맛집	맛있는곳
	맛있는집
	음식점

2.12 동의어 추출 모델 실험 결과

Model	Test
동의어 판단	98.47%

떡볶이	떡볶이
체크	첼크
강아지	댕댕이
샐러드	샐러드
커피샵	커피숍에서
xi	XS
순댓국	순대국집
메타세콰이어	메타세콰이아

사전에 없는 신규 동의어 후보



떡볶이	떡볶이
체크	첼크
강아지	댕댕이
샐러드	샐러드
메타세콰이어	메타세콰이아

신규 동의어

2.13 서비스 예시

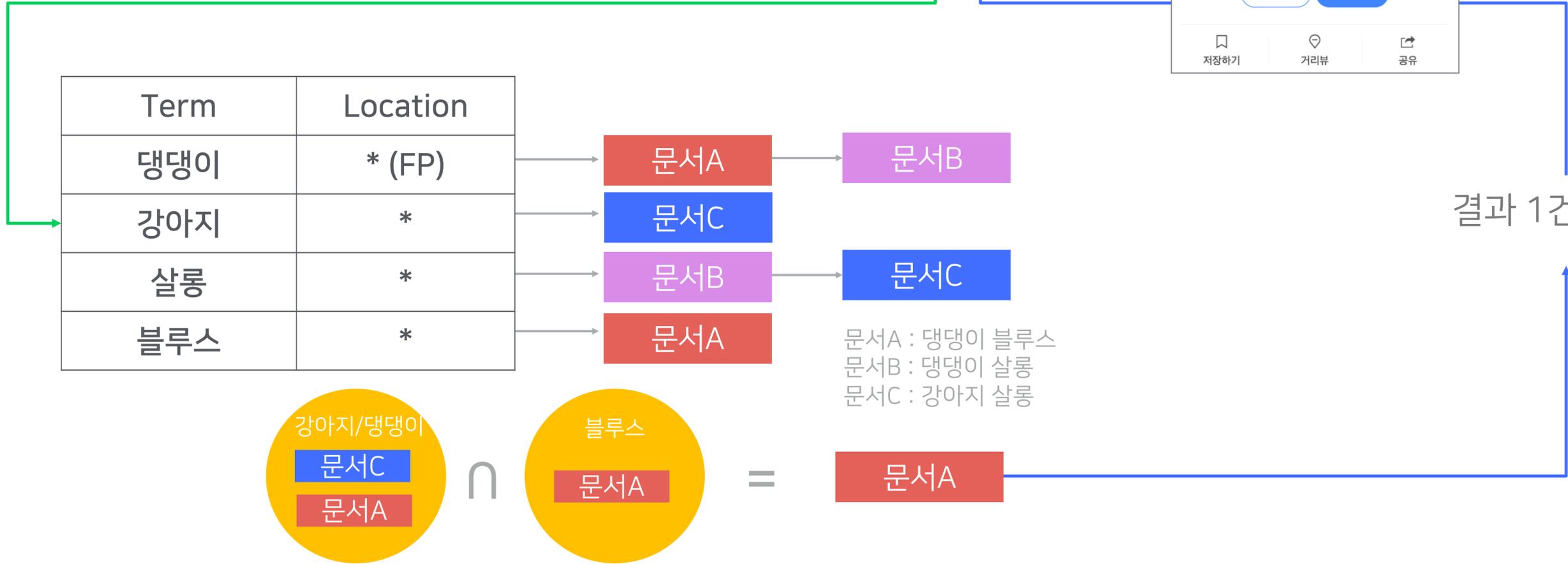
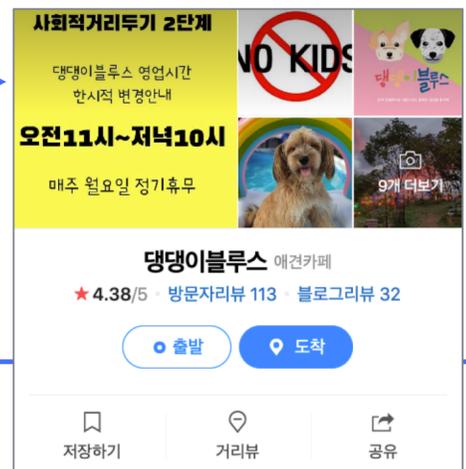


강아지 블루스

강아지	댕댕이
셀러드	셀러드
메타세콰이어	메타세콰이아

(강아지 OR 댕댕이)
AND 블루스

Term matching



3. Modeling 개선을 통한 정답형 질의 확장

3.1 정답형 추출 모델 문제점 / 개선 방안

* 질의-업체 유사도

$$\frac{EditDist_{Qry, Title}}{Max(Len_{Qry}, Len_{Title})} \leq 0.4$$

질의와 업체명의 글자 차이만으로 유사도 측정

↳ 질의의 의미를 기반으로한 유사도 측정

* 클릭 집중비율

$$\frac{Clk_{SID, Qry}}{Clk_{Qry}} \geq 0.8$$

업체 클릭 집중도만으로 파악

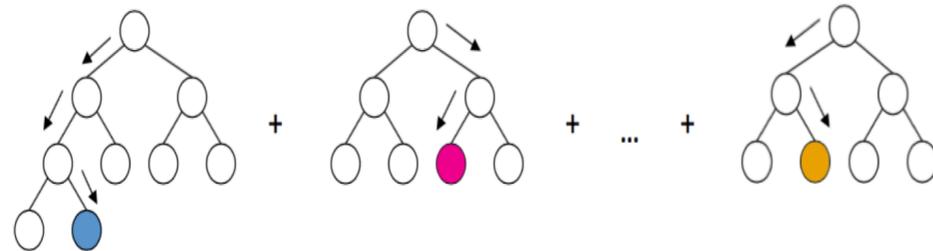
↳ 다양한 user-feedback feature 활용

정답형 질의의 기준을 사람이 정함

↳ Machine Learning으로 기계가 데이터를 보고 정답형 질의 기준 학습

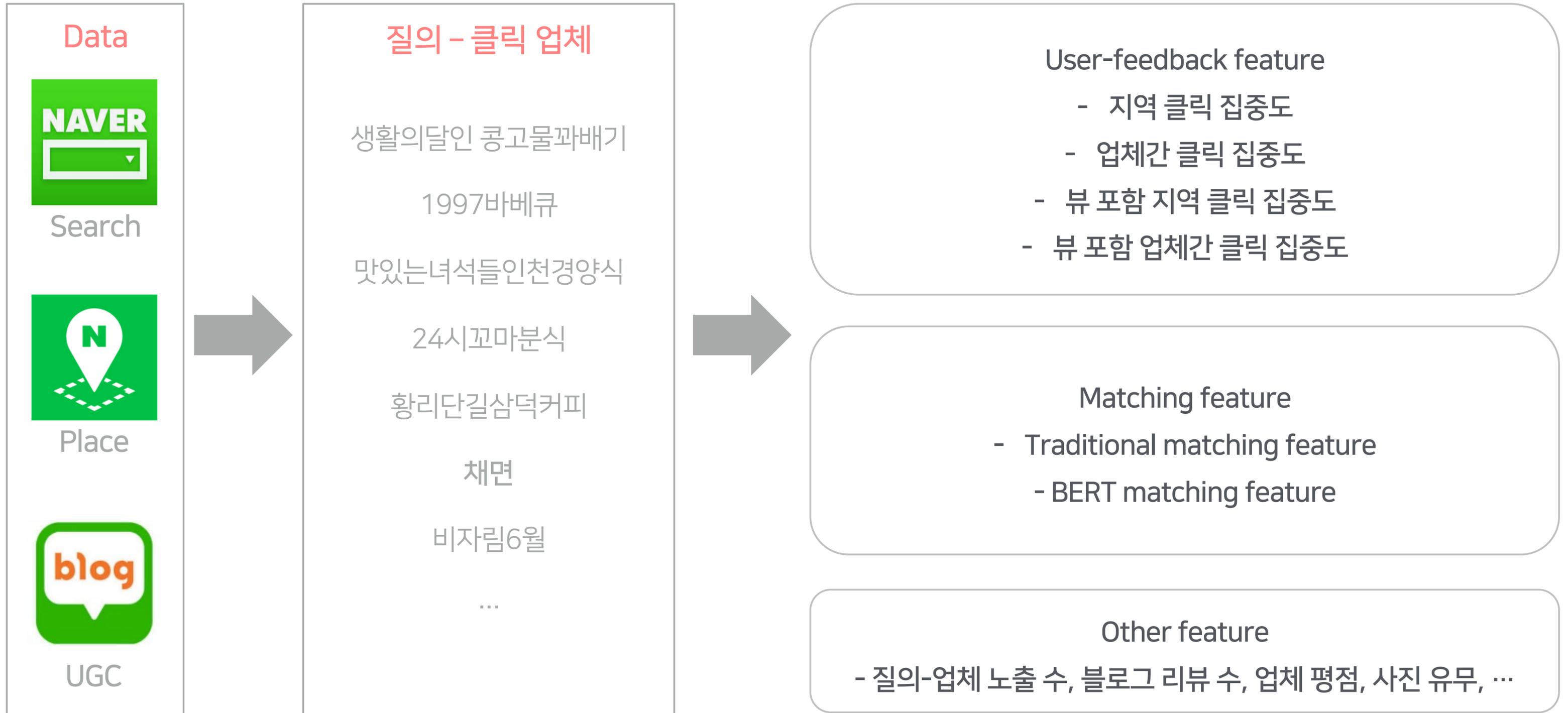


XGBoost Classification Model

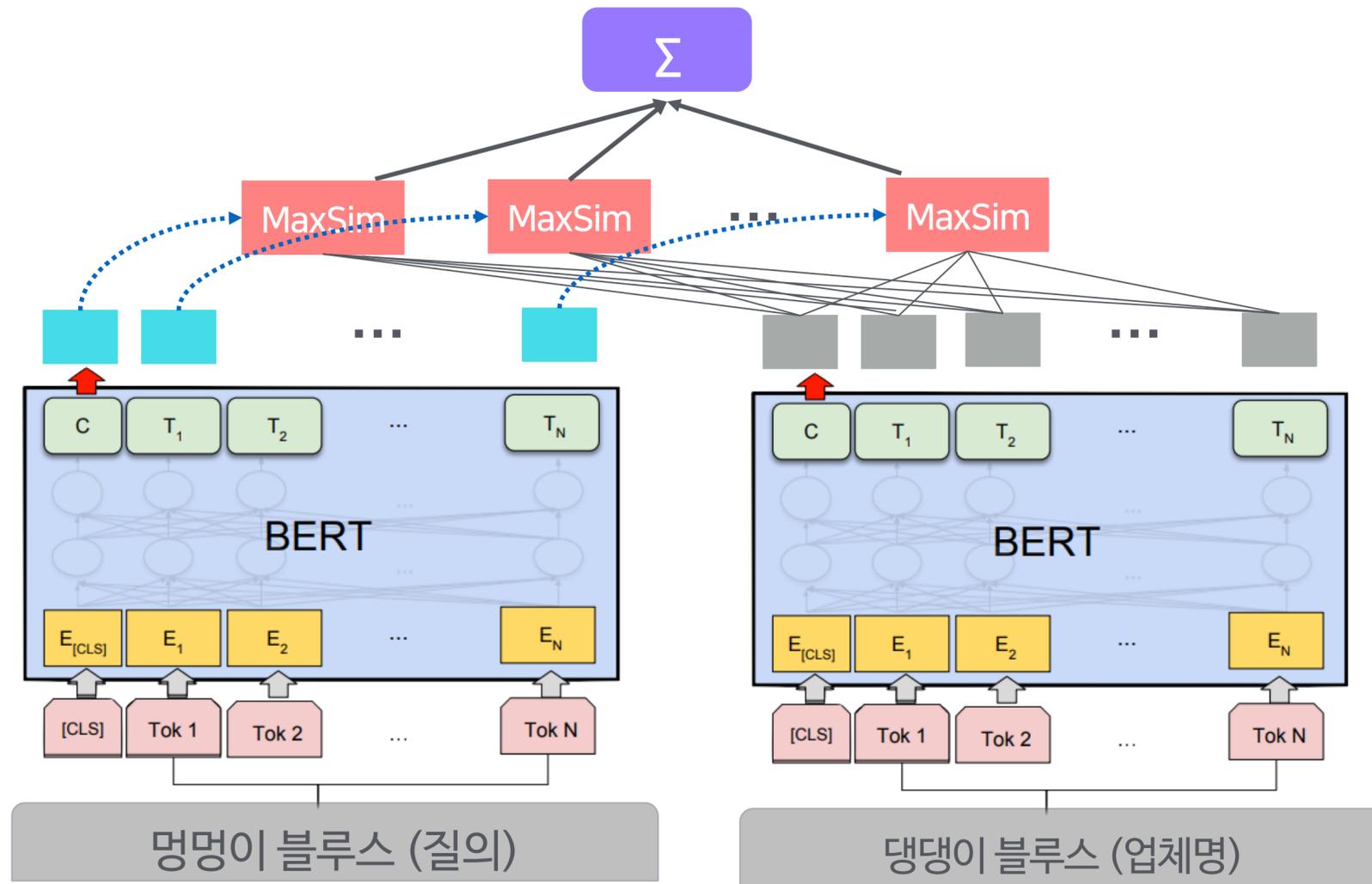


- Decision Tree Gradient Boosting 알고리즘
- 분산환경 실행 가능
- 성능과 자원 효율이 좋음

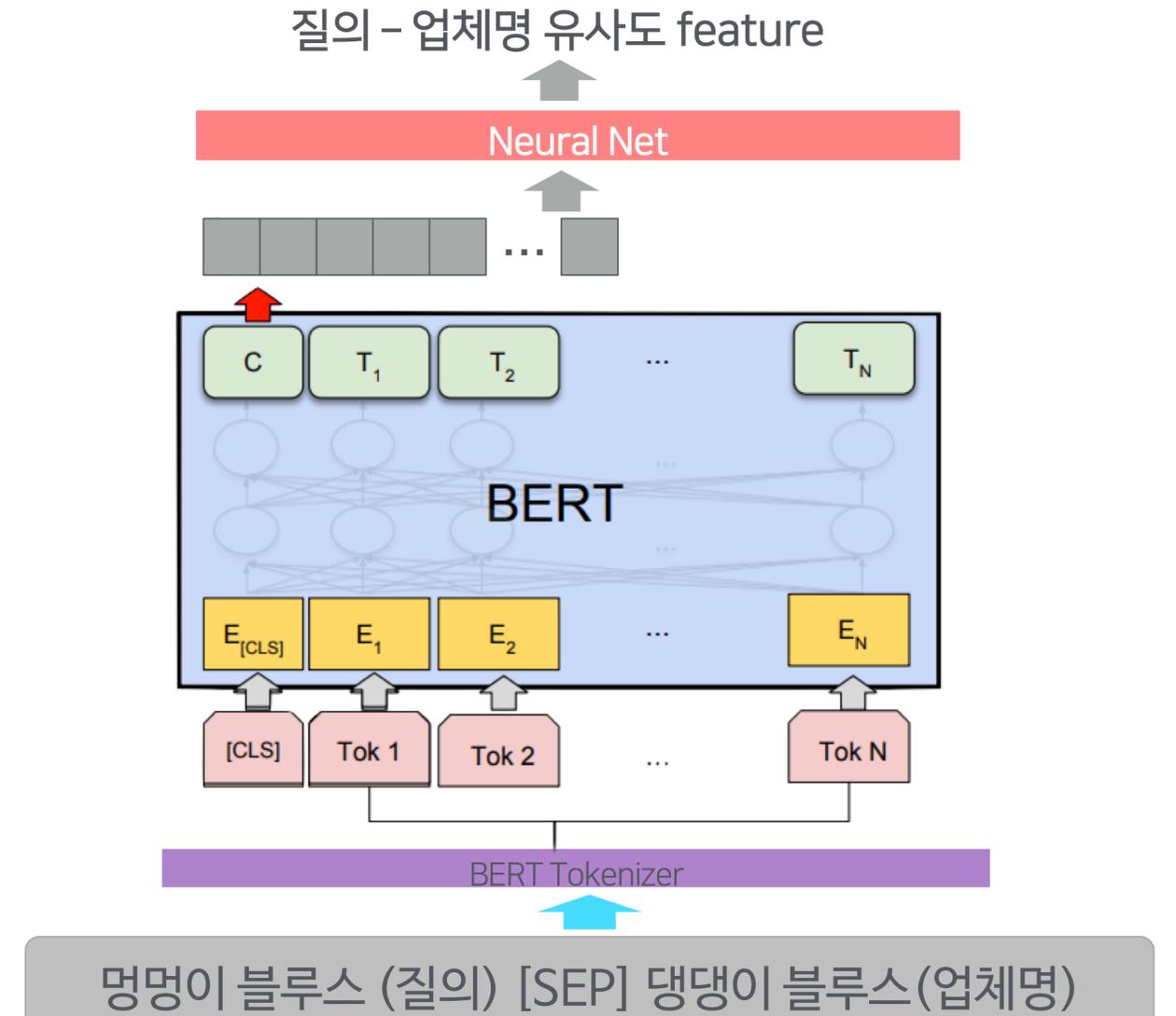
3.2 학습 데이터 구성



3.3 BERT 기반 질의-업체 유사도 feature

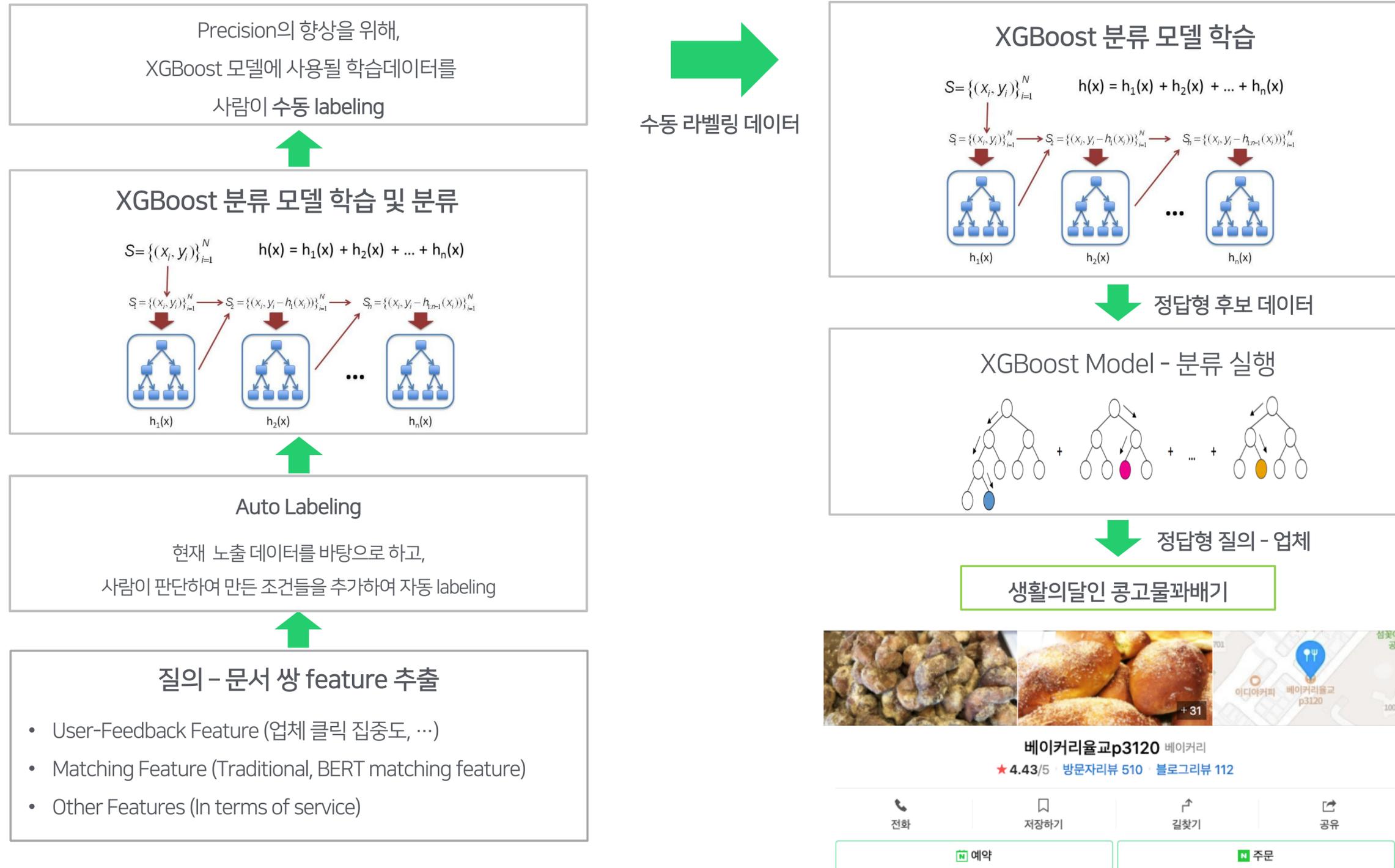


Pretrained BERT 모델



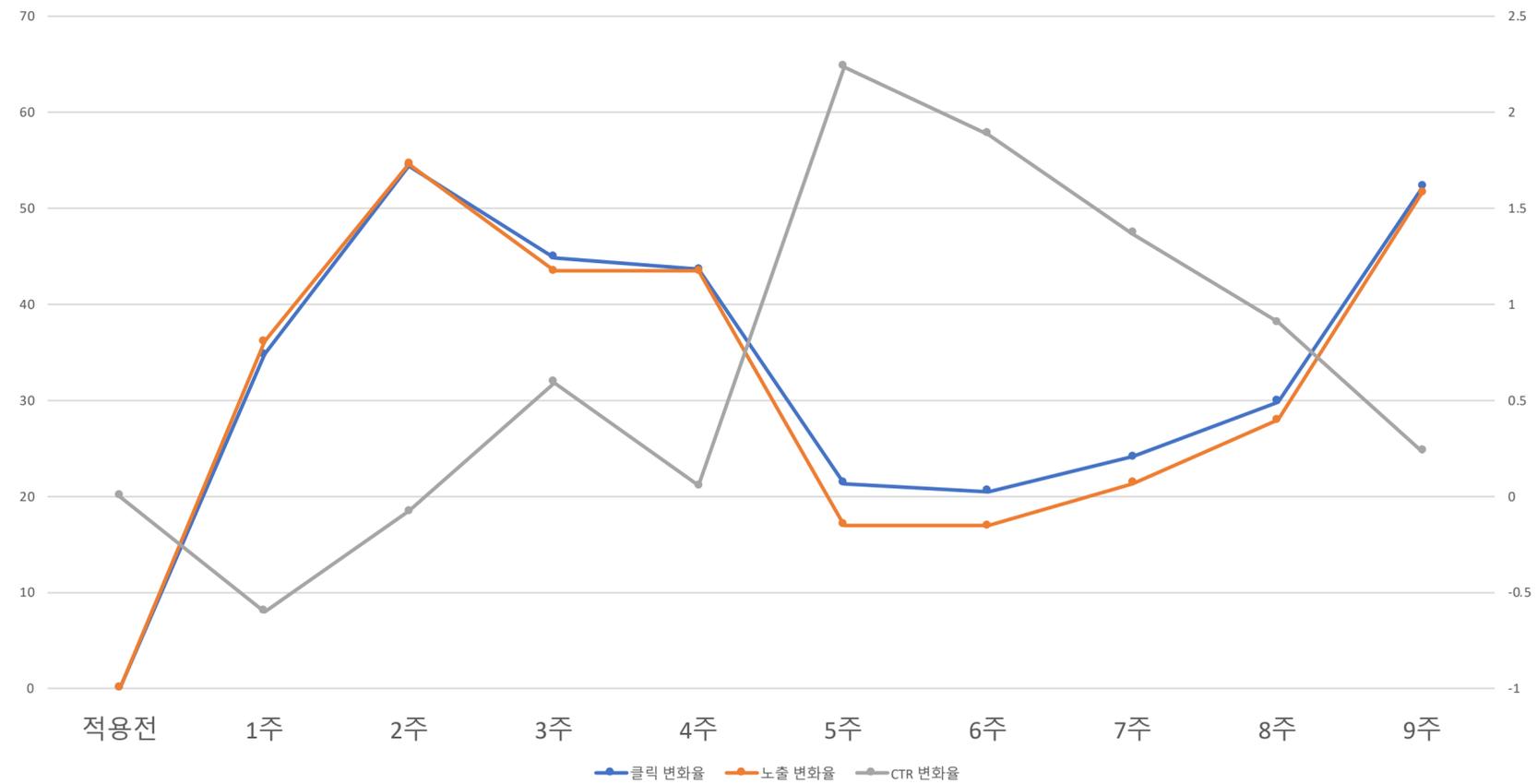
업체명동의어 - 업체명 데이터 학습 기반 BERT 모델

3.4 모델 학습 및 정답형 질의 추출 과정



3.5 서비스 적용 결과

정답형 지표 (적용 전 대비)



증가율
 cc **+36.35%** qc **+34.55%** ctr **+0.82%**

3.6 XGBoost 모델의 개선해야할 점

Long-tail 질의들에 대한 정답형 추출 불가

N 강릉불고기 초당점 ✕ 🔊 🔍

플레이스
MY

위치 정보 없음
📍 내위치 | 변경

초당 중심으로 강릉불고기 검색하기 >

• 관련도순 • 거리순

강릉불고기 초당점 육류,고기요리

172km | 강릉 초당동 ▾

수요일 휴무 공휴일 정상영업 | 매일 10:00~21:00 Last oder 19:30

☎ 전화
📍 길찾기
🔗 공유

2

VIEW
☰ 📄 📺

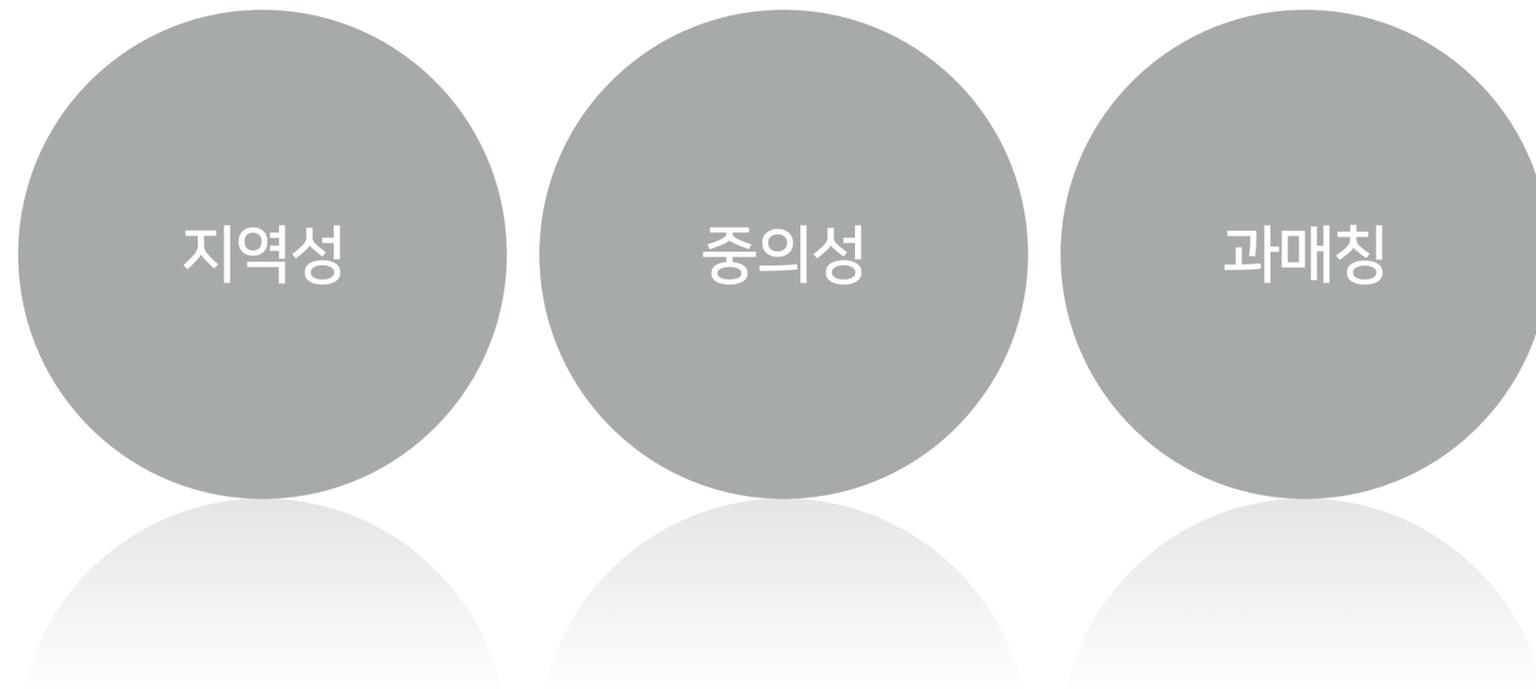
👤 dr.장 이야기 | 2021.04.11.

강릉 불고기 초당점 / 메뉴, 가격

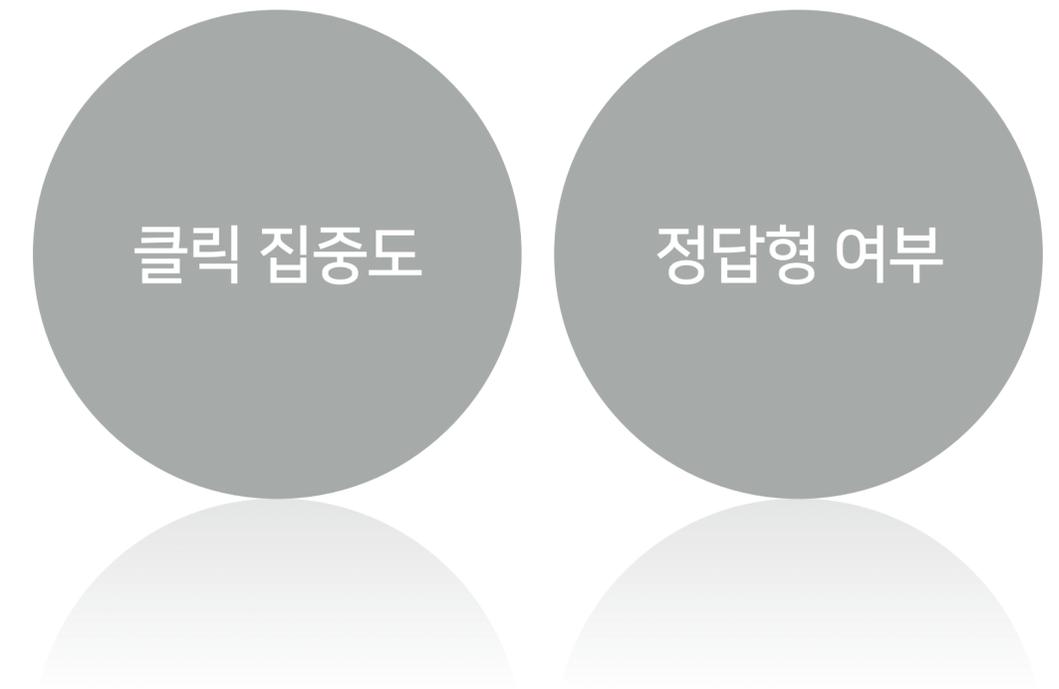
주차는 바로 앞 강릉 불고기 건물 앞 주차장에 주차 가능. 강릉 불고기 초당점 영업시간은 10:00 ~21:00(19:30이 라스트 오더) 수요일은 휴무이고 공휴일은 정상 영업 오늘의 이야기도...

15

3.7 학습데이터 구성

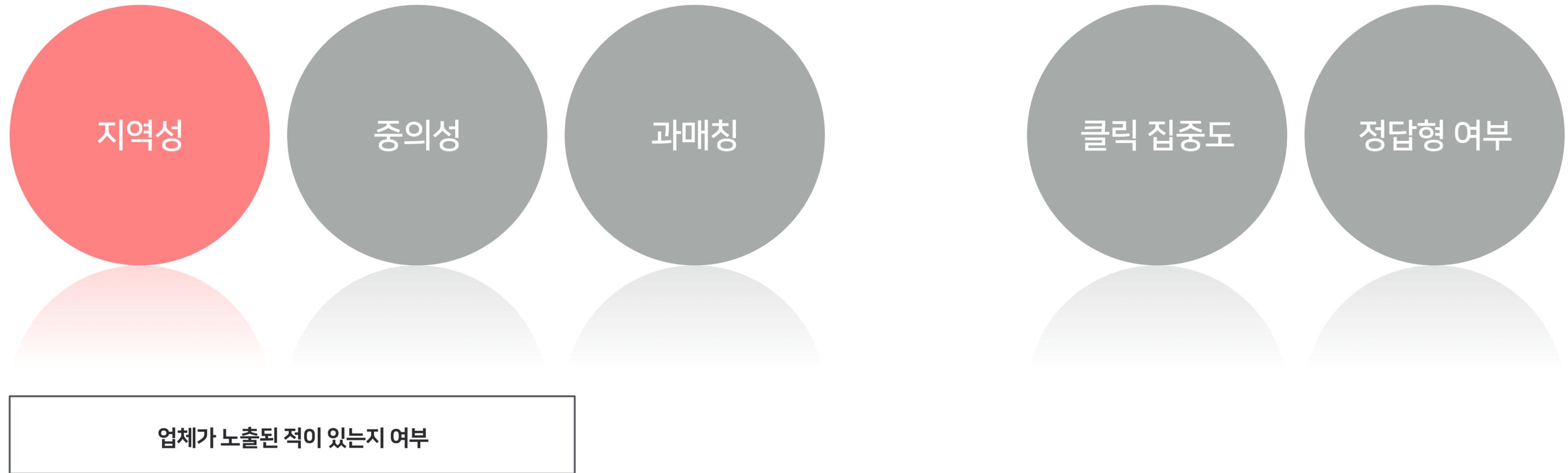


질의 만 있어도 알 수 있는 특성

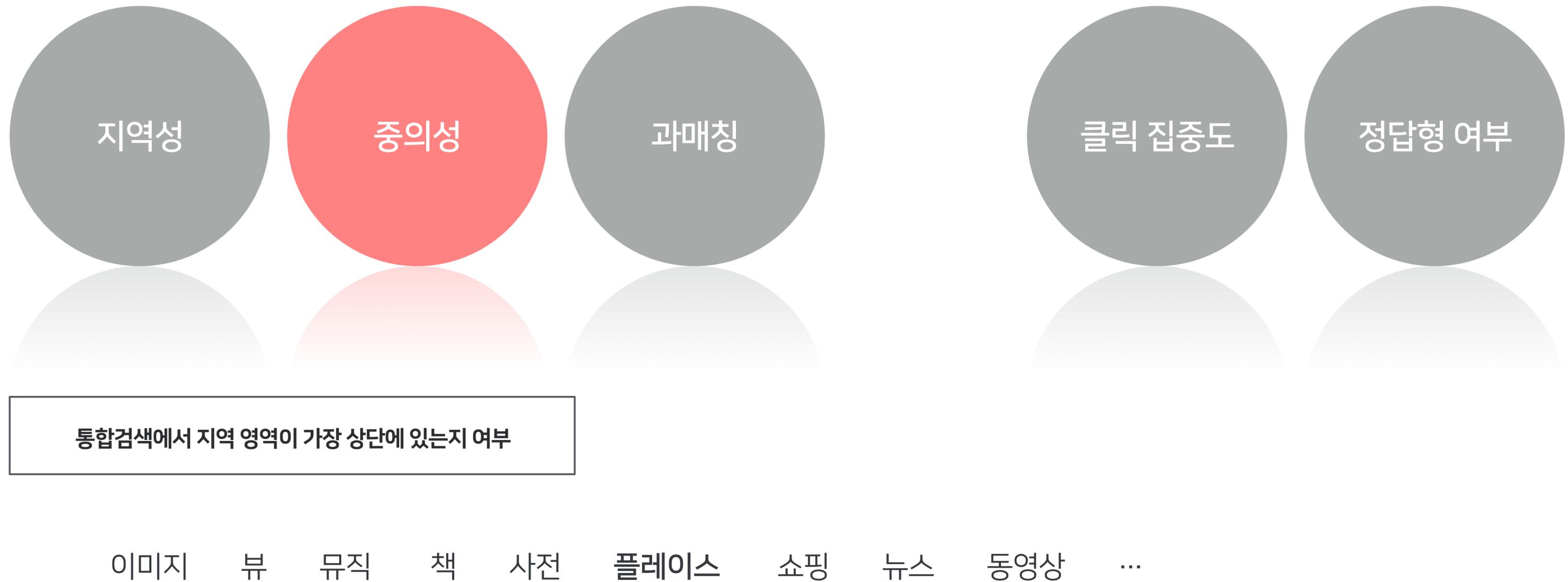


질의-문서 pair 여야만 알 수 있는 특성

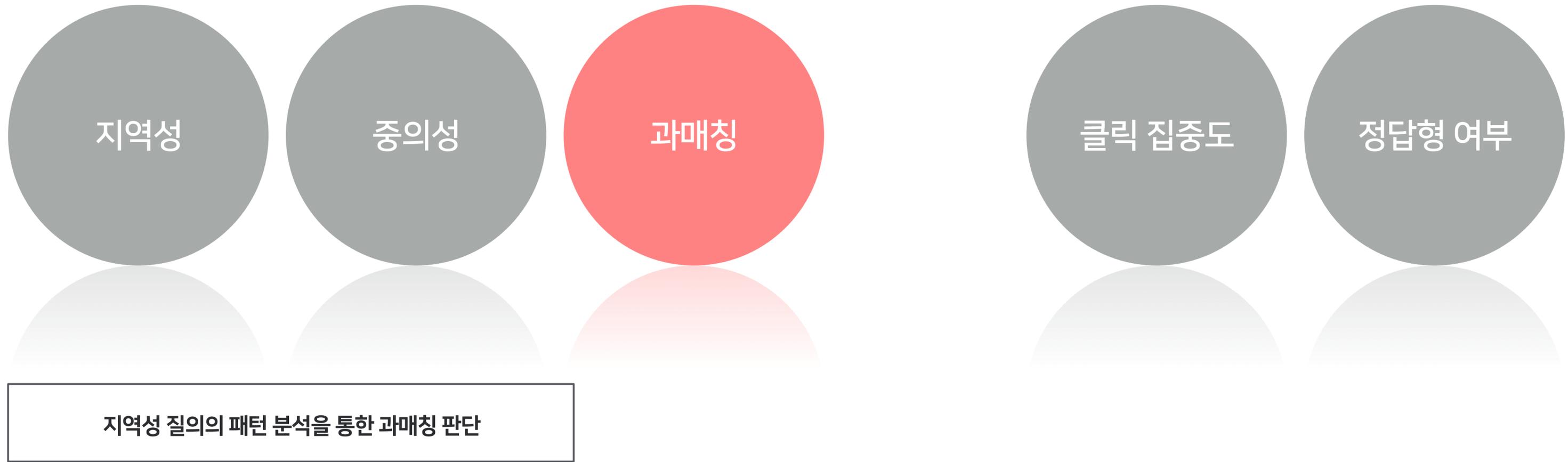
3.7 학습데이터 구성



3.7 학습데이터 구성



3.7 학습데이터 구성



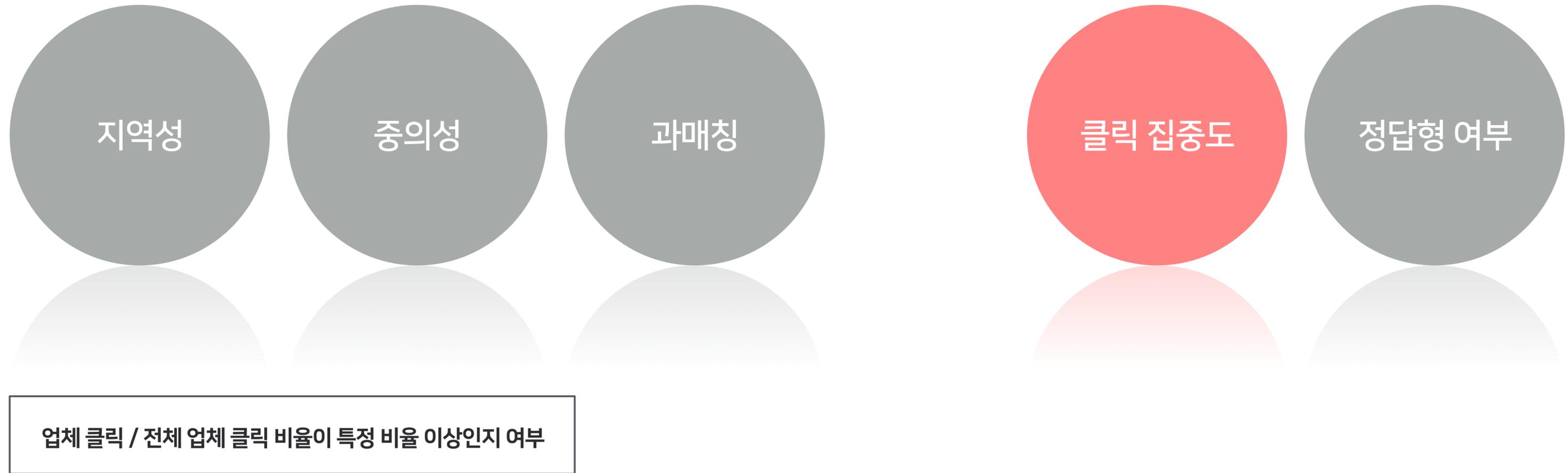
지역성 질의의 패턴 분석을 통한 과매칭 판단

가락동 막국수 맛집
울산진장동 마사지 후기
119 뒷고기 양덕

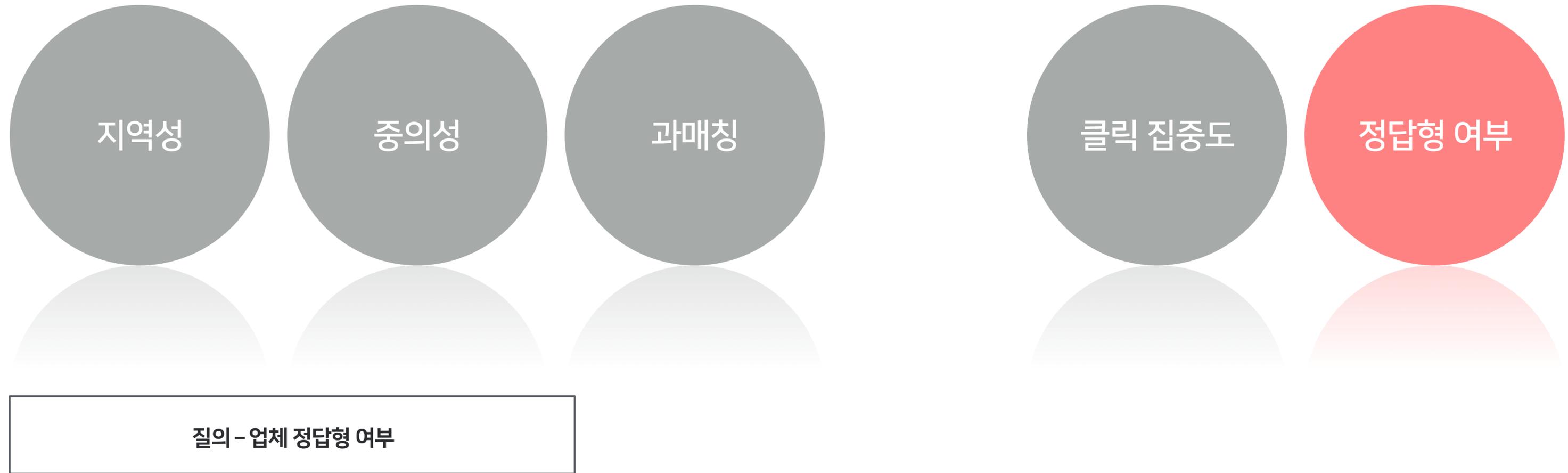
카테고리성 질의
리뷰성 질의
업체명 질의

→ 과매칭
→ 과매칭
→ 정답형 후보

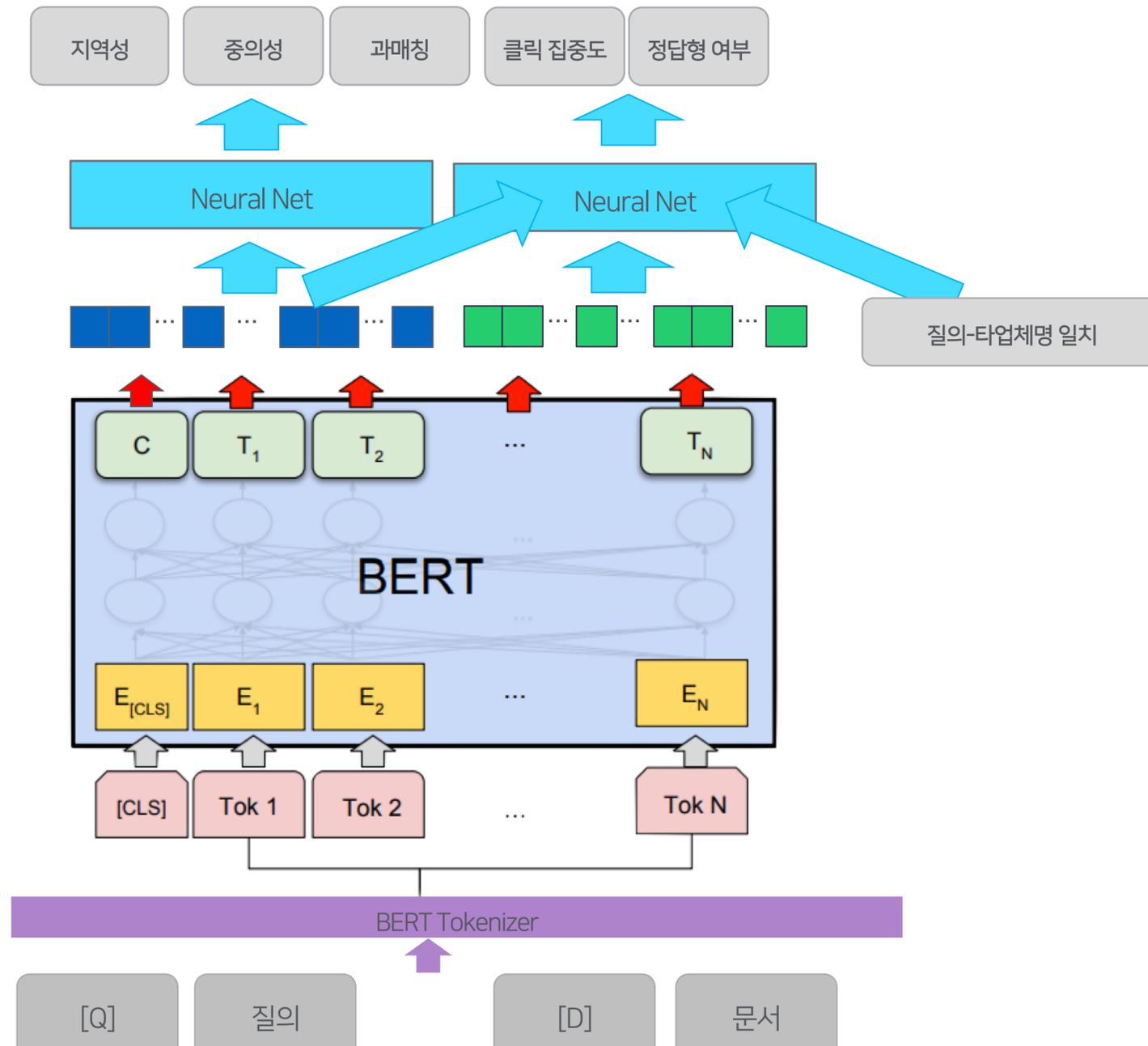
3.7 학습데이터 구성



3.7 학습데이터 구성



3.8 정답형 예측 모델



지역성	94.80%
중의성	93.38%
과매칭	94.59%
클릭 집중도	92.91%
정답형 여부	98.81%

	True	False
Positive	1059	17
Negative	1468	7456

recall : 41.91%

accuracy : 85.15%

precision : 98.42%

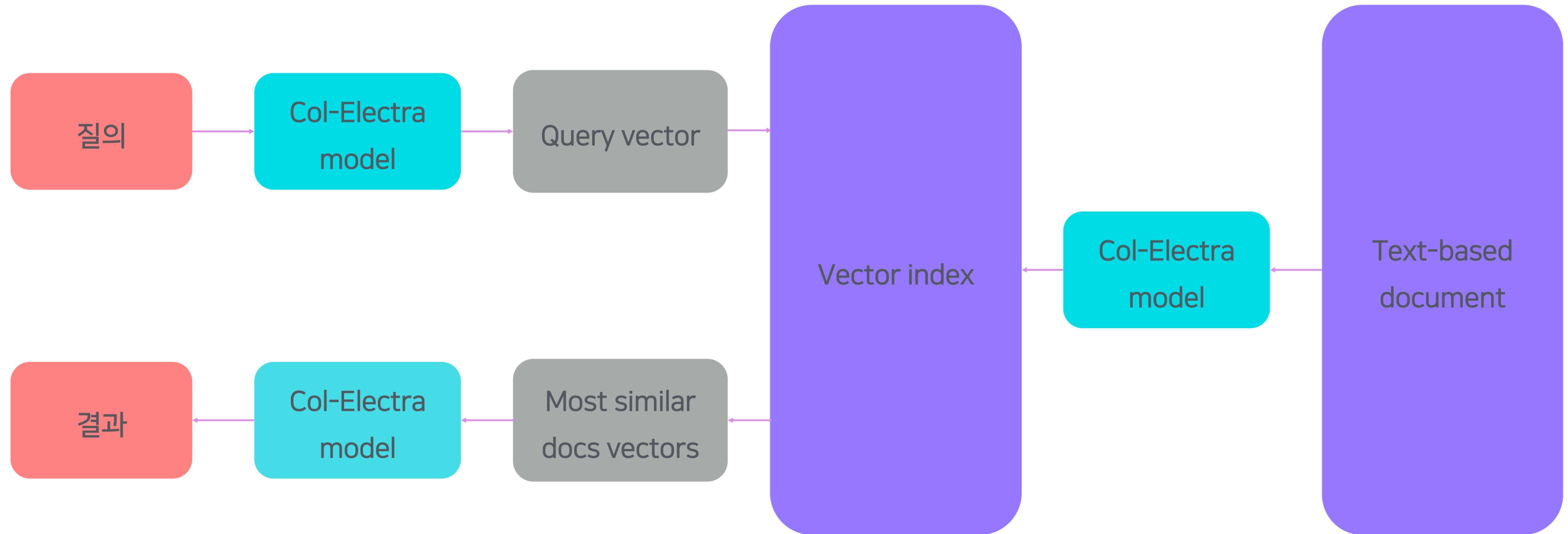
4. What's Next?

4.1 What's Next?

매칭 개선

- Embedding 기반의 matching 개선 (user-feedback 개선)

4.2 Embedding 기반의 매칭/랭킹 프로세스





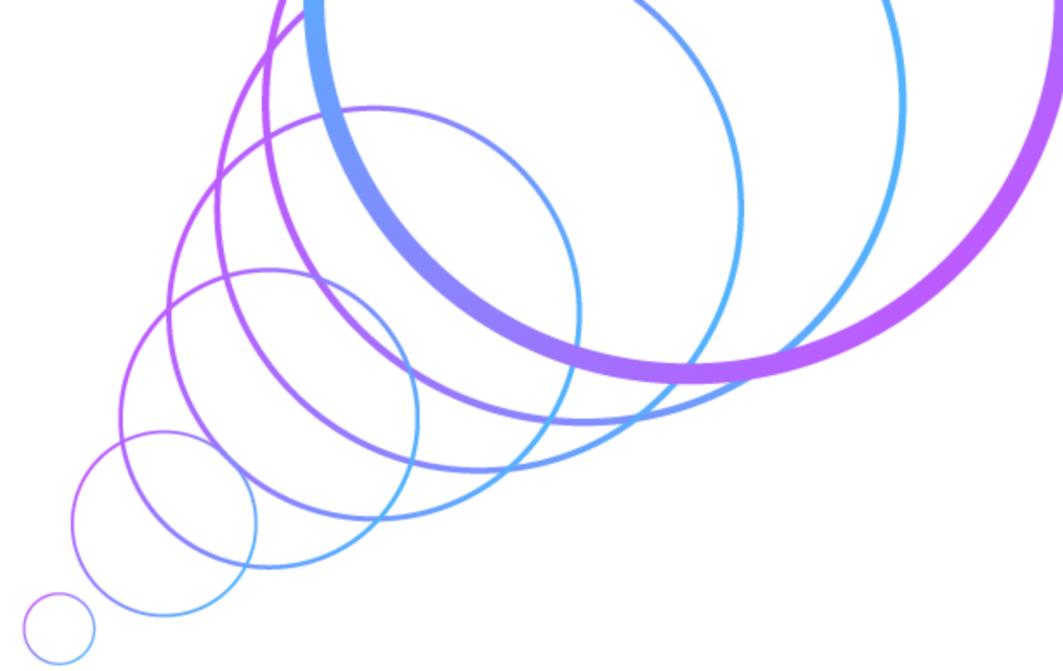
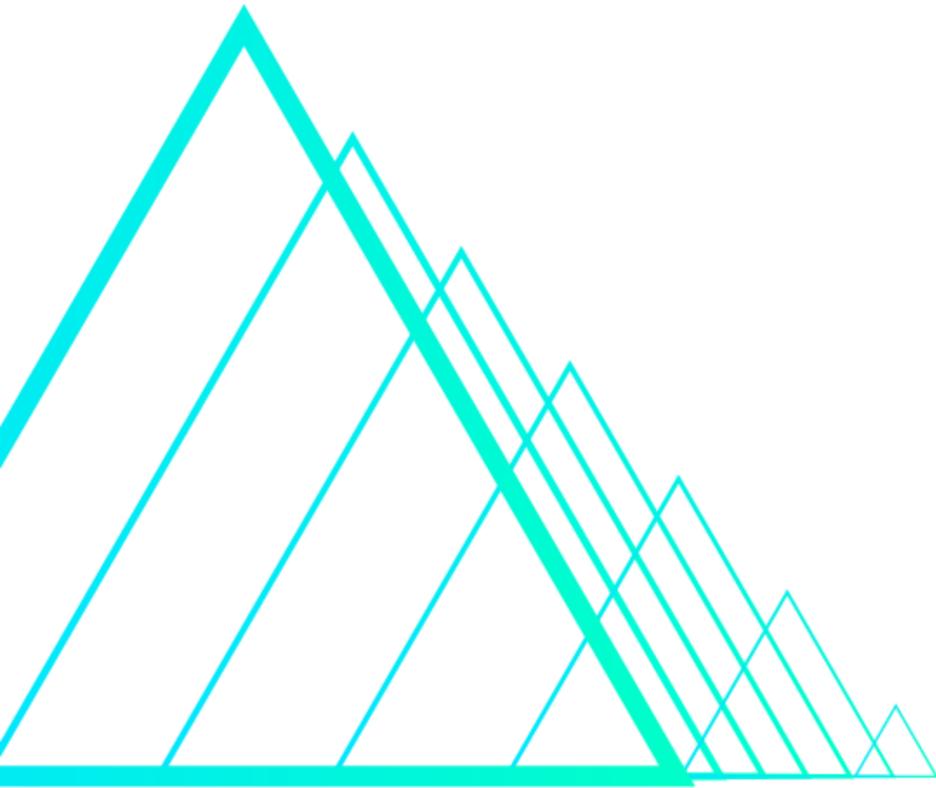
We are Hiring!

네이버/LINE 지역검색 및 추천서비스 개발 (신입/경력)

<https://naver-career.gitbook.io/kr/service/search/local/local-search>



Q & A



Thank You

