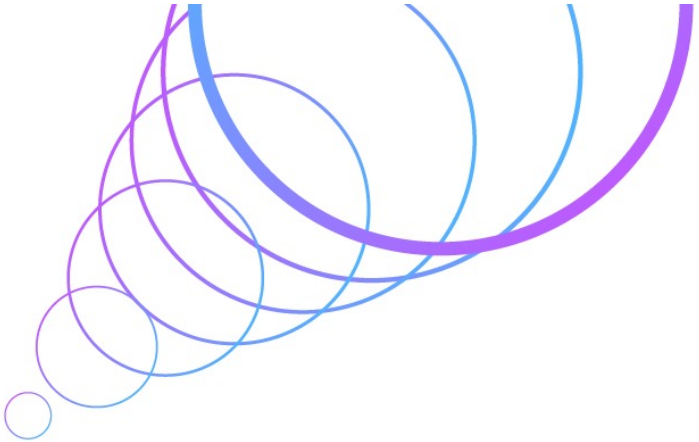


DEVIEW 2021

사용자 의도와 취향을 이해하는 네이버 통합검색

김기동 NAVER / AiRSearch
임희재 NAVER / AiRSearch

CONTENTS

1. 질의 의도에 따라 검색결과 최적화

1-1. 질의 의도 이해

1-2. 질의 의도 자질의 개발 – LIBERO Project

1-3. 의도에 맞는 랭킹 개발 – TIGGER Project

2. 사용자에 따라 검색결과 최적화

2-1. 사용자 이해

2-3. 인플루언서 반응형 검색 개발

DEVIEW 2021

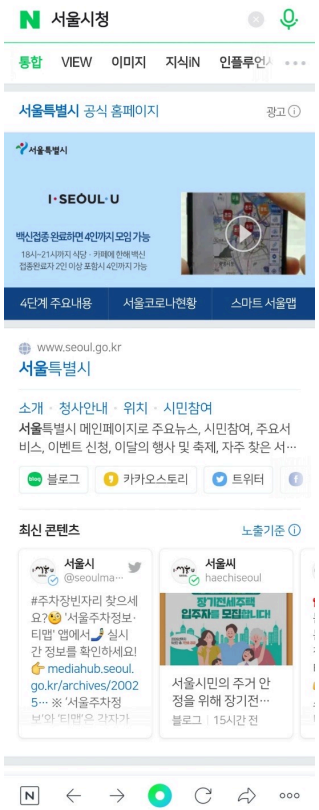
1. 질의 의도에 따라 검색결과 최적화

다양한 출처의 정보들을

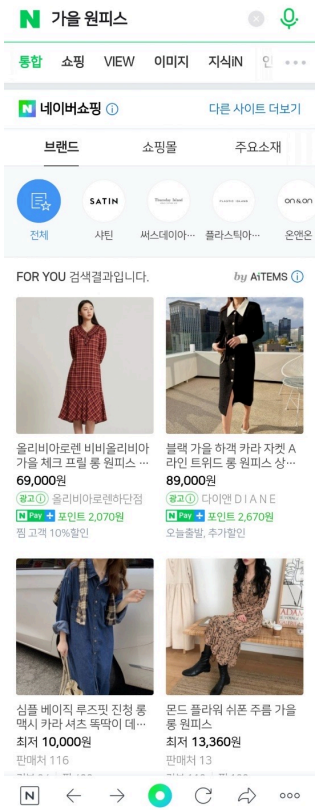
모아서 보여주는 통합검색

N 검색어를 입력해주세요.

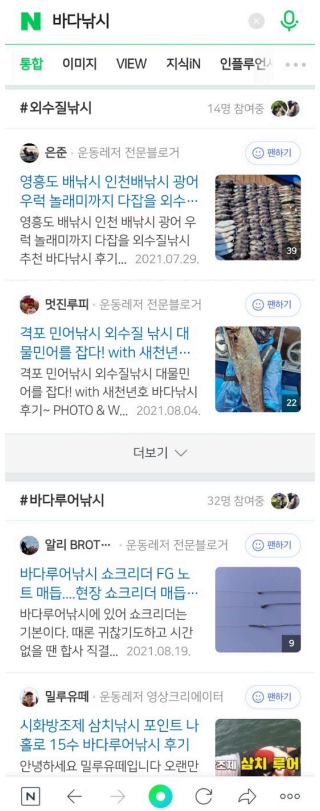
사용자 질의 의도에 부합한 검색결과를 제공하는 역할을 담당합니다.



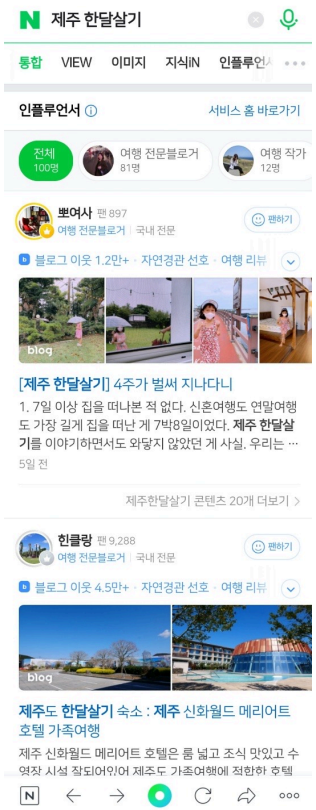
공식 홈페이지



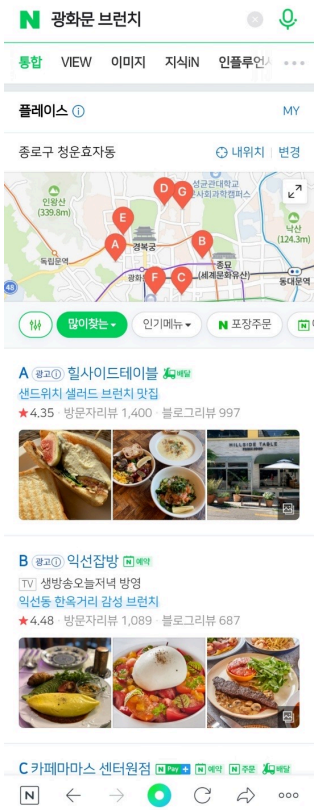
쇼핑



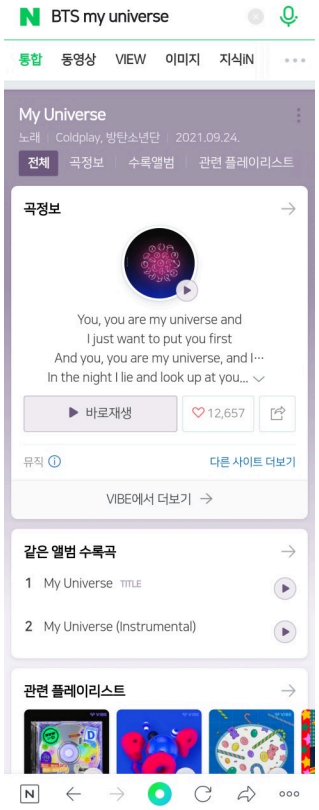
사용자 리뷰 정보



지역 정보



지역 정보

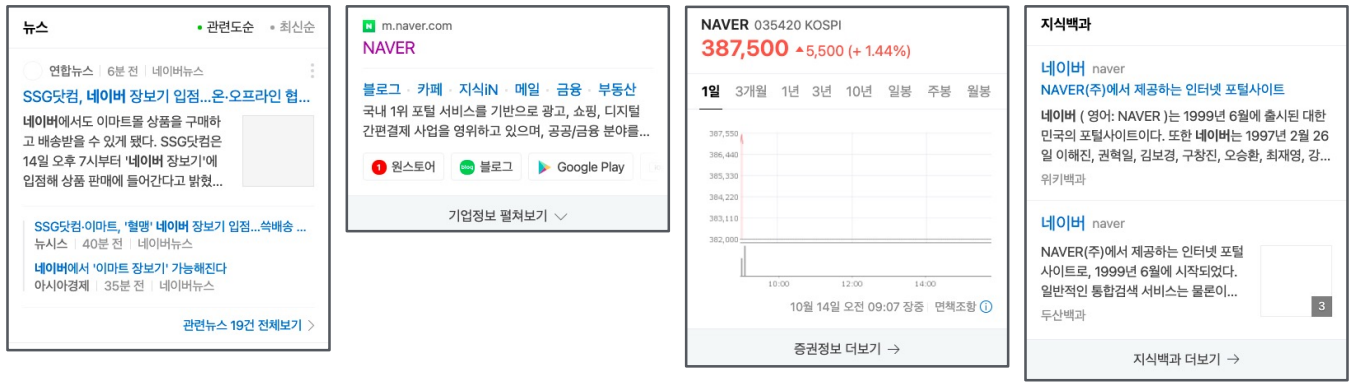


음악 정보

네이버 검색은 2단계의 랭킹 구조를 가지고 있습니다.

버티컬 검색

1차 Ranker – Vertical Search Ranking

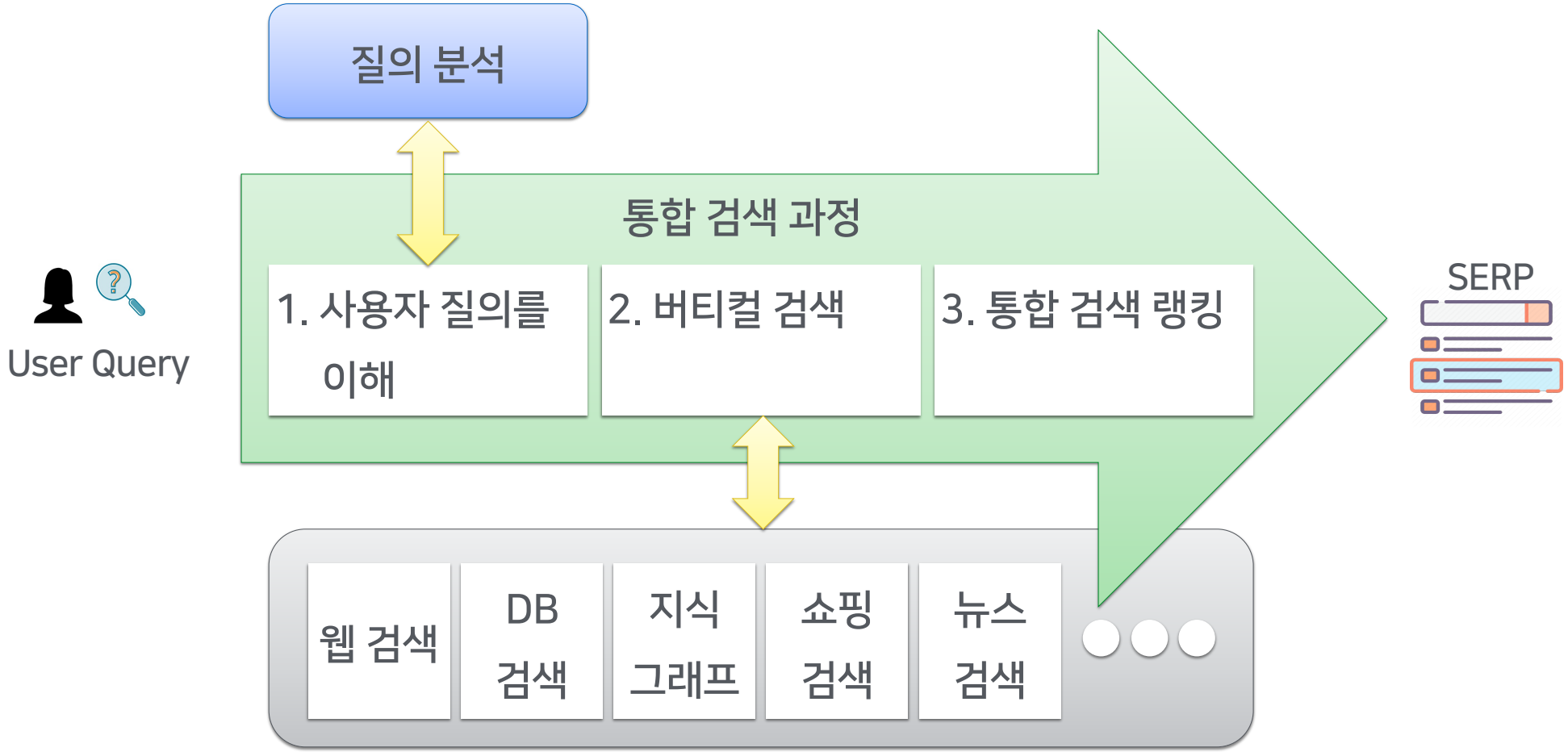


통합 검색

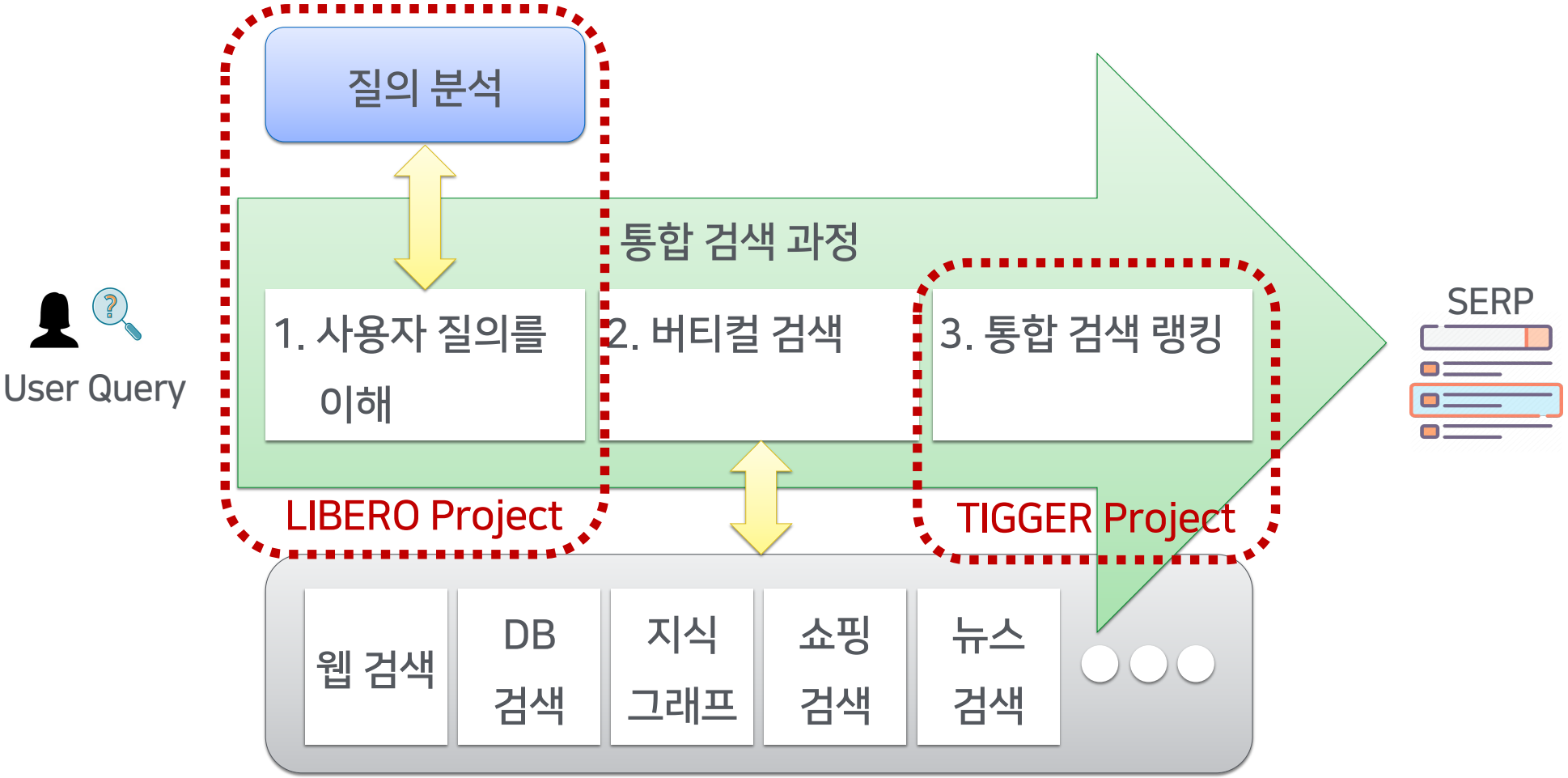
2차 Ranker

– Integrated Search Ranking

통합 검색의 과정은 크게 3단계로 진행합니다.

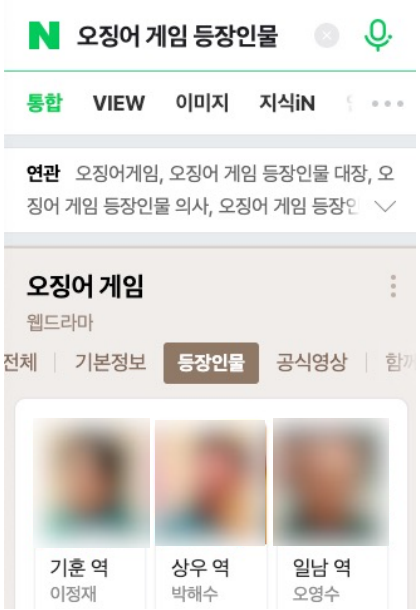


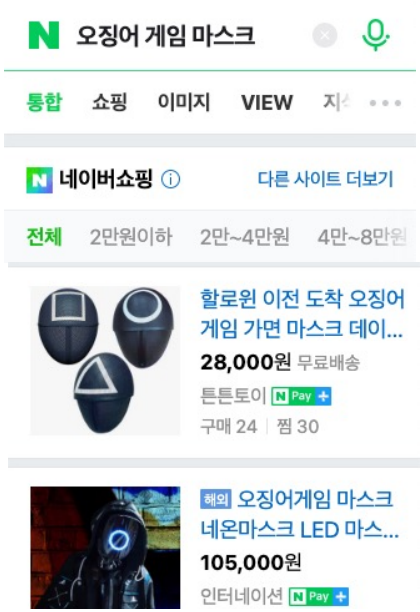
질의를 더 잘 이해하고 의도에 부합한 랭킹을 위한 시도를 소개합니다.



사용자 질의를 이해하는 방식 – Search Topic & Goal

what?	엔터테인먼트	엔터테인먼트	엔터테인먼트	Search Topic
why?	정보를 알려줘	정답을 알려줘	쇼핑하고 싶어	Search Goal





Search Topic

DEVIEW 2021

Table 2.1 The 67 Predefined Categories in KDD Cup 2005 (from [49])

Computers\Hardware	Computers\Internet & Intranet
Computers\Mobile Computing	Computers\Multimedia
Computers\Networks & Telecommunication	Computers\Security
Computers\Software	Computers\Other
Entertainment\Celebrities	Entertainment\Games & Toys
Entertainment\Humor & Fun	Entertainment\Movies
Entertainment\Music	Entertainment\Pictures & Photos
Entertainment\Radio	Entertainment\TV
Entertainment\Other	
Information\Arts & Humanities	Information\Companies & Industries
Information\Science & Technology	Information\Education
Information\Law & Politics	Information\Local & Regional
Information\References & Libraries	Information\Other
Living\Book & Magazine	Living\Car & Garage
Living\Career & Jobs	Living\Dating & Relationships
Living\Family & Kides	Living\Fashion & Apparel
Living\Finance & Investment	Living\Food & Cooking
Living\Furnishing & Houseware	Living\Gifts & Collectables
Living\Health & Fitness	Living\Landscaping & Gardening
Living\Pets & Animals	Living\Real Estate
Living\Religion & Belief	Living\Tools & Hardware
Living\Travel & Vacation	Living\Other
Online Community\Chat & Instant Messaging	Online Community\Forums & Groups
Online Community\Homepages	Online Community\People Search
Online Community\Personal Services	Online Community\Other
Shopping\Auction & Bids	Shopping\Stores & Products
Shopping\Buying Guides & Researching	Shopping\Lease & Rent
Shopping\Bargains & Discounts	Shopping\Other
Sports\American Football	Sports\Auto Racing
Sports\Baseball	Sports\Basketball
Sports\Hockey	Sports\News & Scores
Sports\Schedules & Tickets	Sports\Soccer
Sports\Tennis	Sports\Olympic Games
Sports\Outdoor Recreations	Sports\Other

The taxonomy of topics is the foundation of query topic classification.
It is usually defined as a hierarchical structure,
which may cover many different topics,
like education, economics, travel, sports, and so on.

Ying Li, Zijian Zheng, and Honghua (Kathy) Dai. KDD CUP-2005 report: facing a great challenge. SIGKDD Explorations, 7 (2): 91–99, 2005.

Search Topic @네이버

DEVIEW 2021

- 대표적인 주제 분류 : 대 분류 60여개 / 하위 분류 200여개의 토픽 분류
- 그 외 다양한 분류 체계를 통합하는 과제 진행 (10분류 / 글로벌 검색 기준)

1. 예술/엔터테인먼트

부산 시립미술관, 성경 연대순, 성우 김가령, 가수 이혜리 나이, 중남미문화원, 한자체험관, 쓸쓸하고 찬란한 Hogwarts(책이름)

2. 게임/스포츠/취미/레저

포켓몬스터 일격마스터카드, 필라테스, 자전거, 강서구 복권명당, 애월 사진찍기 좋은 가볼만한 곳, 000움짤

3. 여행/관광

넓은마당 캠핑장, 팜응라오 배낭족거리, 오스트레일리아 1인당국내총생산, 미국국기, 나이지리아 항공권, 목동 깨비시장 주차

4. 건강/의학

다발성근종에좋은음식, 리팅성형외과, 16시간 공복, 의학용어 anemic, 쌀국수 칼로리

5. 라이프스타일

결혼 답례품, 초등학생 어린이날 선물 보드게임, 5학년 추천 책, 금강연화 미용실, 황정음비키니, 복숭아잿빛무늬병, 장수하늘소, 기은세 패딩

6. 경제/사회/정치

인천공항 자가격리숙소, 전기차개별소비세, 아시아나주가, 건강보험 세대주, 월세도배, 공매도 주가흐름, 2021 공기업 채용, 교도소시간표

7. 교육/어학

스크래치 코딩 뮤직배우기, 서울예대 실용음악과 모집요강, 논문명, 택시시험 준비물, 신경쓰지마세요 영어로

8. 자동차/엔지니어링

유아 자동차 햇빛가리개, bmw키홀더, 아시아나, 굴삭기

9. 음식/맛집

닭다리 간장조림, 감자전 레시피, 강화도 카페, 빽다방 가격, 단호박, 장군수산, 계란과 같이먹으면 좋은 음식

10. IT/컴퓨터

에게이밍 조립PC, 복고 폰트, 데스트탑 에어팟, 네이버 휴먼계정, 힐링쉴드, 아이폰 사진 크기 변경, kt vip요금제

no	Search topic	UGC 31	Similar Web
1	예술/엔터테인먼트	문학/책 영화 드라마 공연/전시 미술/디자인 방송 음악 스타/연예인 만화/애니	Arts and Entertainment
2	게임/스포츠/취미/레저	게임 스포츠 취미 사진	Games Sports Gampling Hobbies and Leisure
3	여행/관광	국내여행 세계여행	Travel and Tourism
4	건강/의학	건강/의학	Health
5	라이프스타일	육아/결혼 패션/미용 인테리어/DIY 원예/재배 애완/반려동물	Lifestyle Home and Garden Pets and Animals
6	경제/사회/정치	비즈니스/경제 사회/정치	Buisiness and Consumer Service Finance Community and Society Jobs and Career Law and Covernment News and Media
7	교육/어학	교육/학문 어학/외국어	Science and Education Reference Materials
8	자동차/엔지니어링	자동차	Vehicles Heavy Industry and Engineering
9	음식/맛집	요리/레시피 맛집	Food and Drink
10	IT/컴퓨터	IT/컴퓨터	Computers Electronics and Technology
99	미분류	상품리뷰 좋은글/이미지	E commerce and Shopping

Search Goal

Basically, user goals can be classified based on the type of demanded resources users are seeking for by issuing a query.
Several taxonomies of user goals have been proposed since Broder introduced this concept.
In the first part of this subsection, we will briefly introduce these taxonomies.

web search taxonomy
- Andrei Z. Broder. A taxonomy of web search. SIGIR Forum, 36 (2): 3–10, 2002.

3. A taxonomy of web searches

In the web context the "need behind the query" is often not informational in nature. We classify web queries according to their intent into 3 classes:

1. **Navigational.** The immediate intent is to reach a particular site.
2. **Informational.** The intent is to acquire some information assumed to be present on one or more web pages.
3. **Transactional.** The intent is to perform some web-mediated activity.

Search Goal @네이버

DEVIEW 2021

- 5개 대분류 : Informational / site / location / shopping / transactional
- 20여개 하위 분류 기준으로 분류 모델 개발

1.Informational
- 1-1) Know : Know_simple, Know_broad, Weather, Know_etc
- 1-2) Advice : Review, Comparison, Recommend, Guide/howto/tip
- 1-3) Issue

2.Site
- 2-1) Navigational
- 2-2) Channel
- 2-3) Site_list

3. Location

4. Shopping

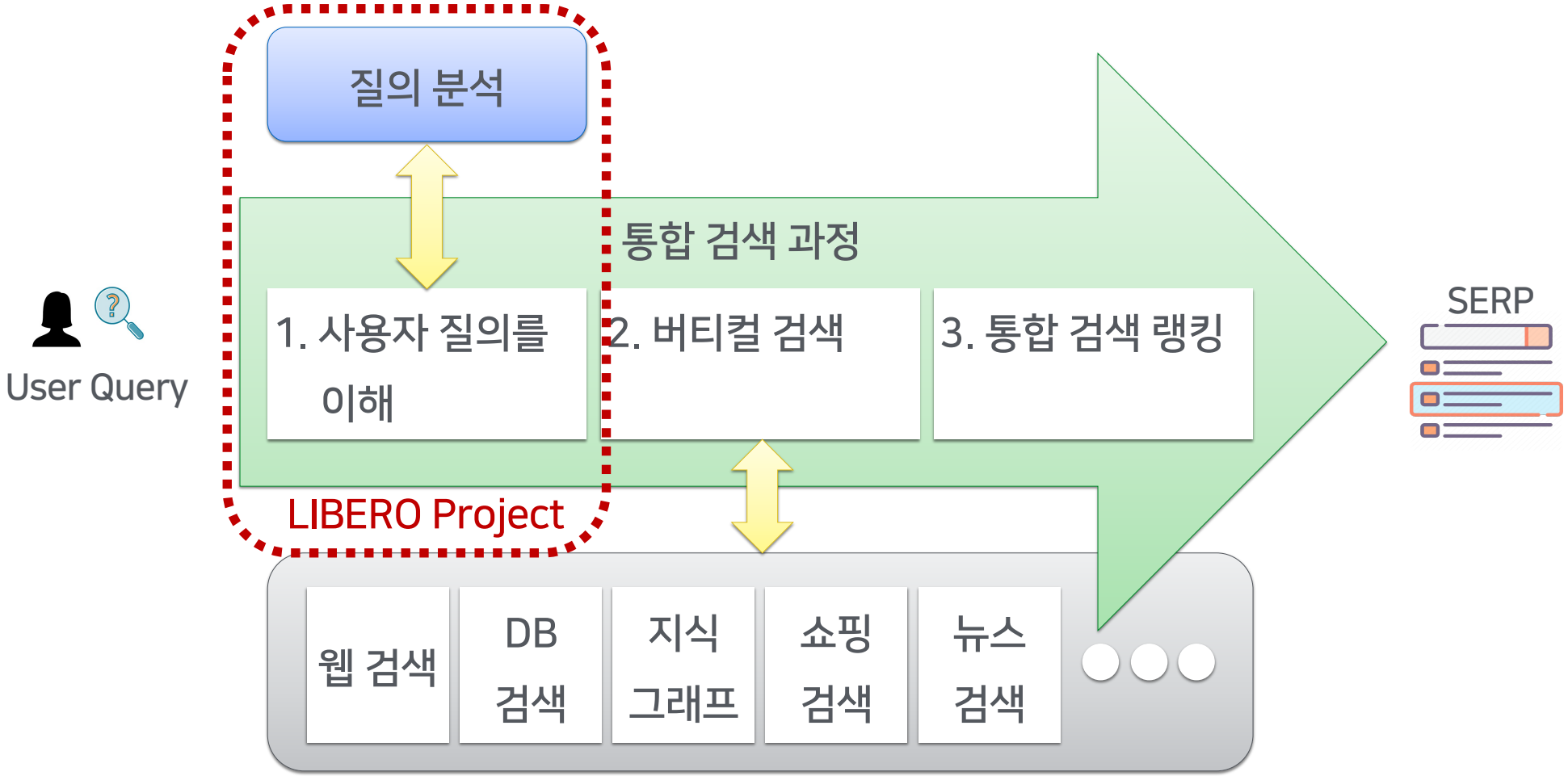
3.Transactional
- 5-1) Tool
- 5-2) Resource : App
- 5-3) Grep
- 5-4) Translate

Table 1: The Search Goal Hierarchy. Queries are only assigned to leaf nodes. All examples are taken from actual AltaVista queries.

SEARCH GOAL	DESCRIPTION	EXAMPLES
1. Navigational	My goal is to go to specific known website that I already have in mind. The only reason I'm searching is that it's more convenient than typing the URL, or perhaps I don't know the URL.	aloha airlines duke university hospital kelly blue book
2. Informational	My goal is to learn something by reading or viewing web pages	
2.1 Directed	I want to learn something in particular about my topic	
2.1.1 Closed	I want to get an answer to a question that has a single, unambiguous answer.	what is a supercharger 2004 election dates
2.1.2 Open	I want to get an answer to an open-ended question, or one with unconstrained depth.	baseball death and injury why are metals shiny
2.2 Undirected	I want to learn anything/everything about my topic. A query for topic X might be interpreted as "tell me about X."	color blindness jfk jr
2.3 Advice	I want to get advice, ideas, suggestions, or instructions.	help quitting smoking walking with weights
2.4 Locate	My goal is to find out whether/where some real world service or product can be obtained	pella windows phone card
2.5 List	My goal is to get a list of plausible suggested web sites (I.e. the search result list itself), each of which might be candidates for helping me achieve some underlying, unspecified goal	travel amsterdam universities florida newspapers
3. Resource	My goal is to obtain a resource (not information) available on web pages	
3.1 Download	My goal is to download a resource that must be on my computer or other device to be useful	kazaa lite mame roms
3.2 Entertainment	My goal is to be entertained simply by viewing items available on the result page	xxx porno movie free live camera in l.a.
3.3 Interact	My goal is to interact with a resource using another program/service available on the web site I find	weather measure converter
3.4 Obtain	My goal is to obtain a resource that does not require a computer to use. I may print it out, but I can also just look at it on the screen. I'm not obtaining it to learn some information, but because I want to use the resource itself.	free jack o lantern patterns ellis island lesson plans house document no. 587

- Daniel E. Rose / Yahoo! inc. Understanding User Goals in Web Search. WWW2013

LIBERO Project



LIBERO – Search goal 분류기 개발

DEVIEW 2021

- 질의만 가지고 의도를 학습하여,
- 컬렉션에서 올라오는 다른 정보 없이 질의 의도를 6가지로 분류
- 비슷한 패턴의 질의의 의도 유추 (unseen 질의 이해도 고취)
- 의도에 따른 랭킹개선

	의도	부가설명
의도1	Buy	구매/쇼핑
의도2	Visit-in-person	위치 기반 컨텐츠 소비
의도3	Issue	이슈/뉴스
의도4	Website	특정 웹사이트 이동
의도5	Know	리뷰/의견 이나 컨텐츠 내 다양한 정보 소비
의도6	Answer	전문 지식이나 컨텐츠 내 특정 정답을 찾는 의도

LIBERO – Search goal 분류기 개발

DEVIEW 2021

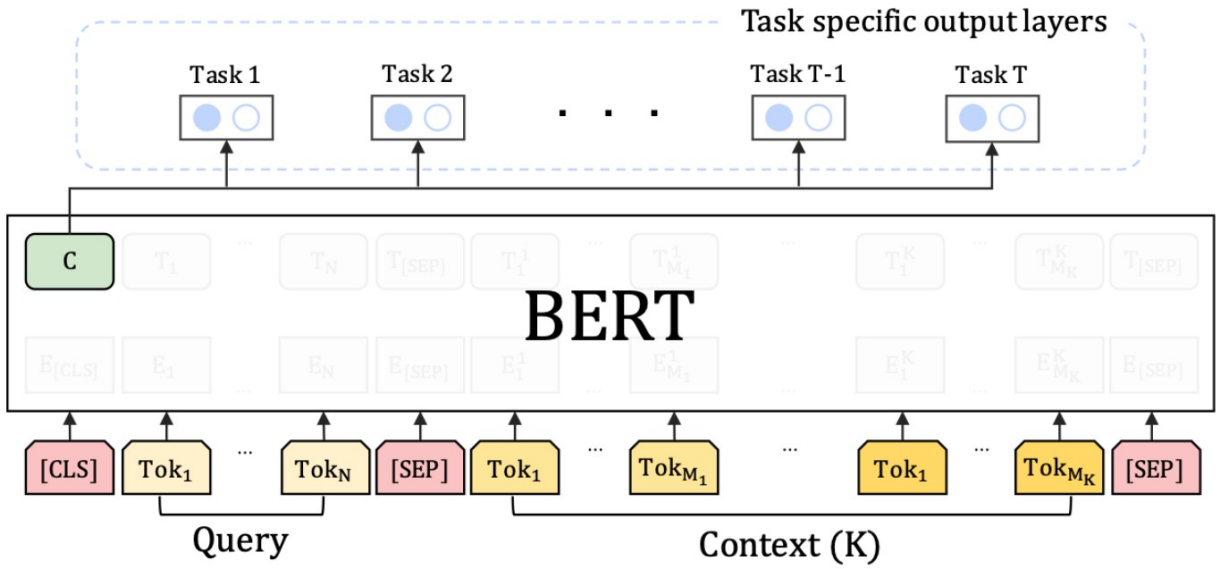
Intent	Description
Answer	전문 지식 / 특정 답변을 찾으려는 의도
Buy	구매 / 쇼핑
Issue	이슈 / 뉴스
Know	리뷰 & 코멘트 / 다양한 정보 탐색
Navigational	특정 웹사이트로 이동하려는 의도
Visit	위치 기반

base 의도

조합 의도

LIBERO – 모델

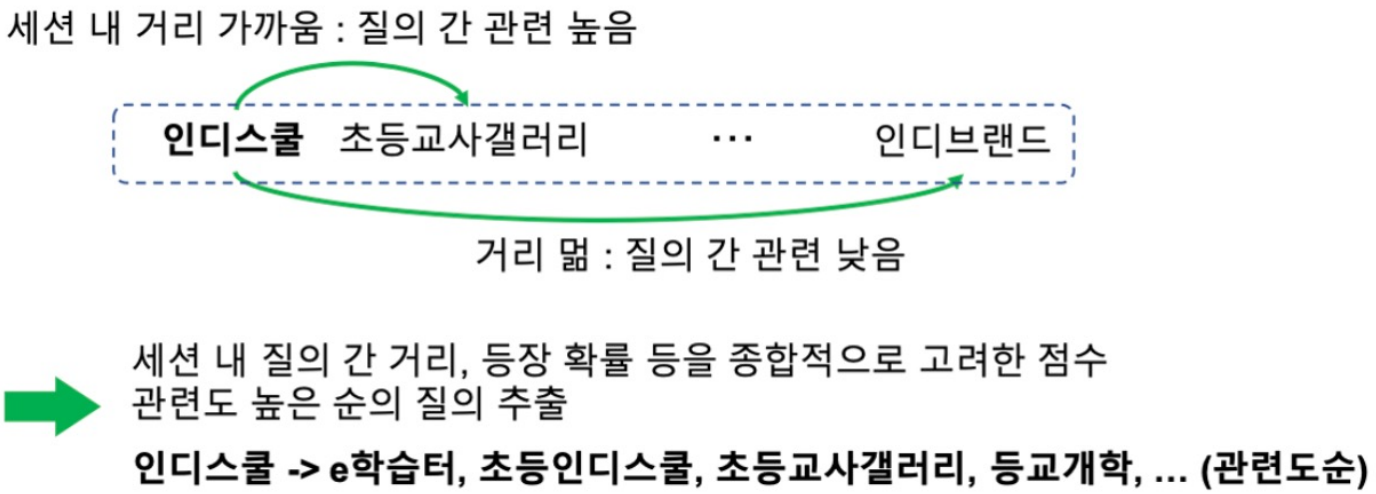
- Pretrained model: RoBERTa 기반 모델
- 6개의 의도별 독립적인 score를 얻기 위해 **Multi-label Classification** 모델로 fine-tuning
- loss 계산 시 의도 class 들의 softmax 대신 각 의도에 대한 sigmoid 계산
- 학습 시 **context 정보** 추가 (단어 내림차순)



•Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. RoBERTa: A robustly optimized BERT pretraining approach. arXiv, 2019

LIBERO - 모델

- 검색 질의와 함께 session co-occurrence 질의들도 모델의 input으로 사용
- Input으로 넣기 전 session co-occurrence 질의들을 모두 aggregate 하여 context 정보가 풍부하고 길이가 긴 sequence로 만들어 사용
- 추출 방법



LIBERO – Label Smoothing

DEVIEW 2021

- Hard label만 사용 시 다른 클래스에 대한 정보 모두 손실
- auto 정답셋이기 때문에 label에 대한 정확도가 100%라고 보긴 어려움
- 각 의도에 해당하는 사용자들의 클릭 피드백 값의 합산을 1로 scaling 하여 **soft label**로 사용.

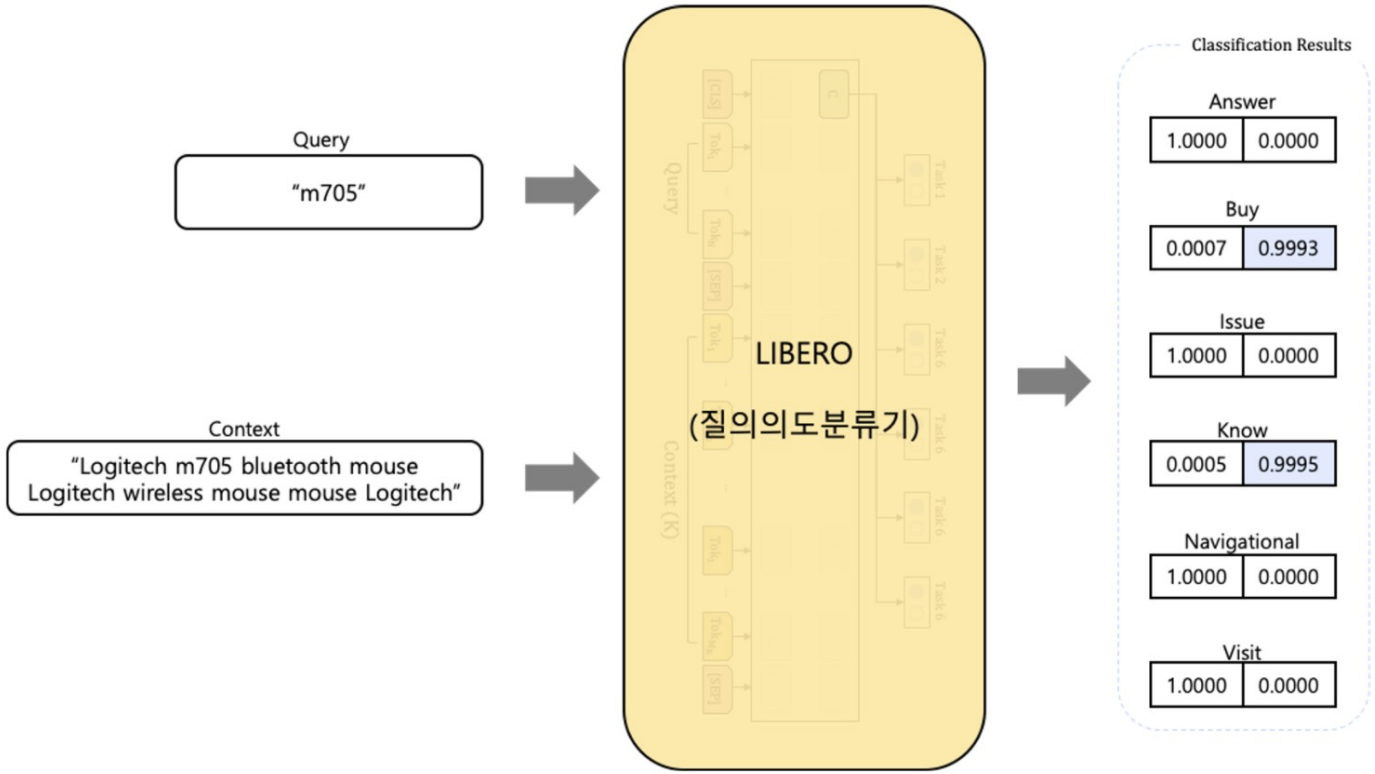
query	intent	answer	buy	issue	know	navi	visit	answer_soft	buy_soft	issue_soft	know_soft	navi_soft	visit_soft
여주세종병원	navi|visit	0	0	0	0	1	1	0.030390	0.000000	0.025959	0.087745	0.414649	0.441257
c202msf933	buy	0	1	0	0	0	0	0.000000	0.872174	0.000000	0.127826	0.000000	0.000000
인스타그램 링크복사	know	0	0	0	1	0	0	0.000000	0.000000	0.036146	0.963854	0.000000	0.000000
반면교사	answer	1	0	0	0	0	0	0.753559	0.000000	0.015956	0.230485	0.000000	0.000000
메로나 칼로리	answer|know	1	0	0	1	0	0	0.365027	0.000000	0.050859	0.584115	0.000000	0.000000
도안 어울림하트	visit	0	0	0	0	0	1	0.000000	0.000000	0.027554	0.144422	0.000000	0.828024

hard labels

soft labels

LIBERO – Results (sample)

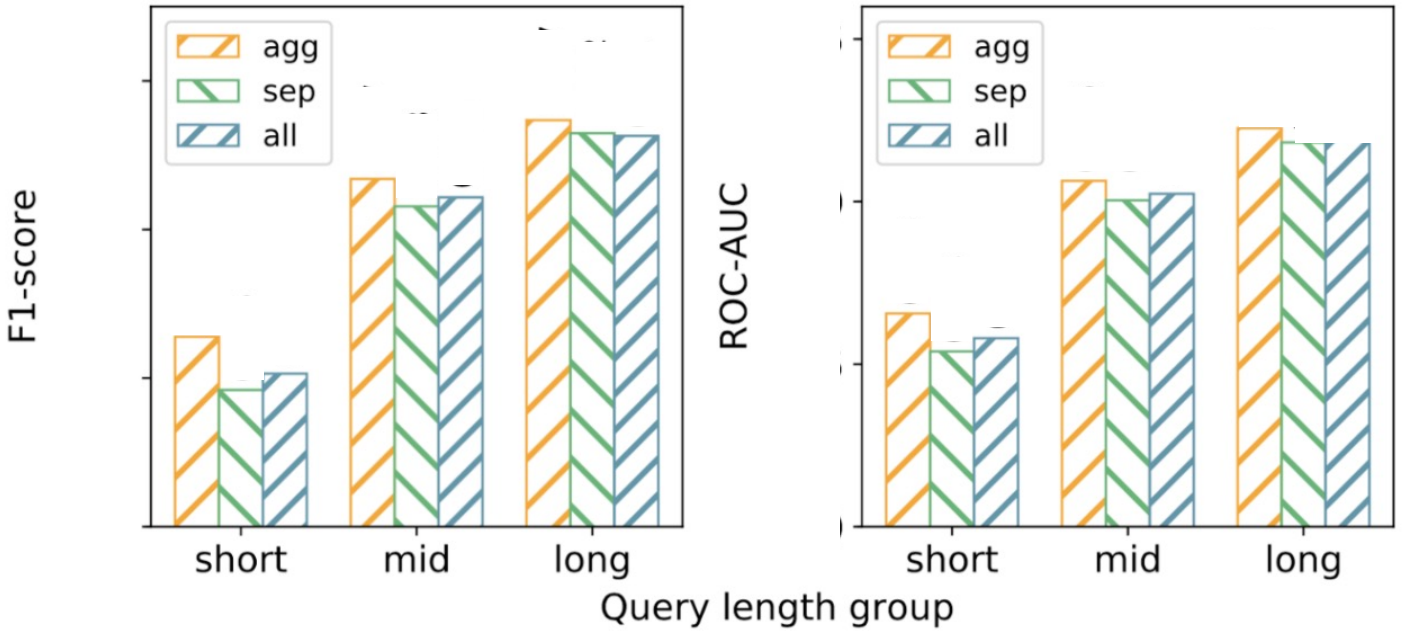
- 질의: “m705”
- context: [“Logitech m705”, “bluetooth mouse”, “Logitech wireless mouse”, “mouse”, “logitech”]
- results: [answer: 0.0, buy: 0.9993, issue: 0.0, know: 0.9995, navigational: 0.0, visit: 0.0]



LIBERO – Context

Context 정보 추가 방법에 따른 성능 비교

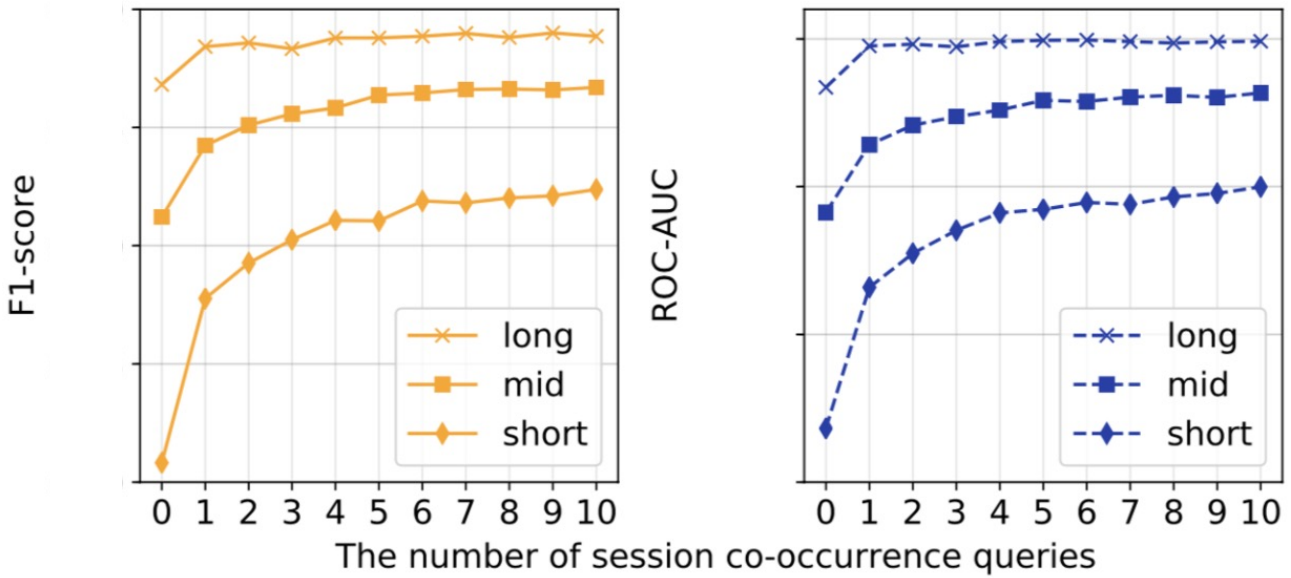
- 세션 질의를 활용하기 위해 여러 context 구조를 실험하였습니다.
- 그 중 agg 방식이 가장 좋은 성능을 보여 최종적으로 이 방식을 선택하여 사용하였습니다.



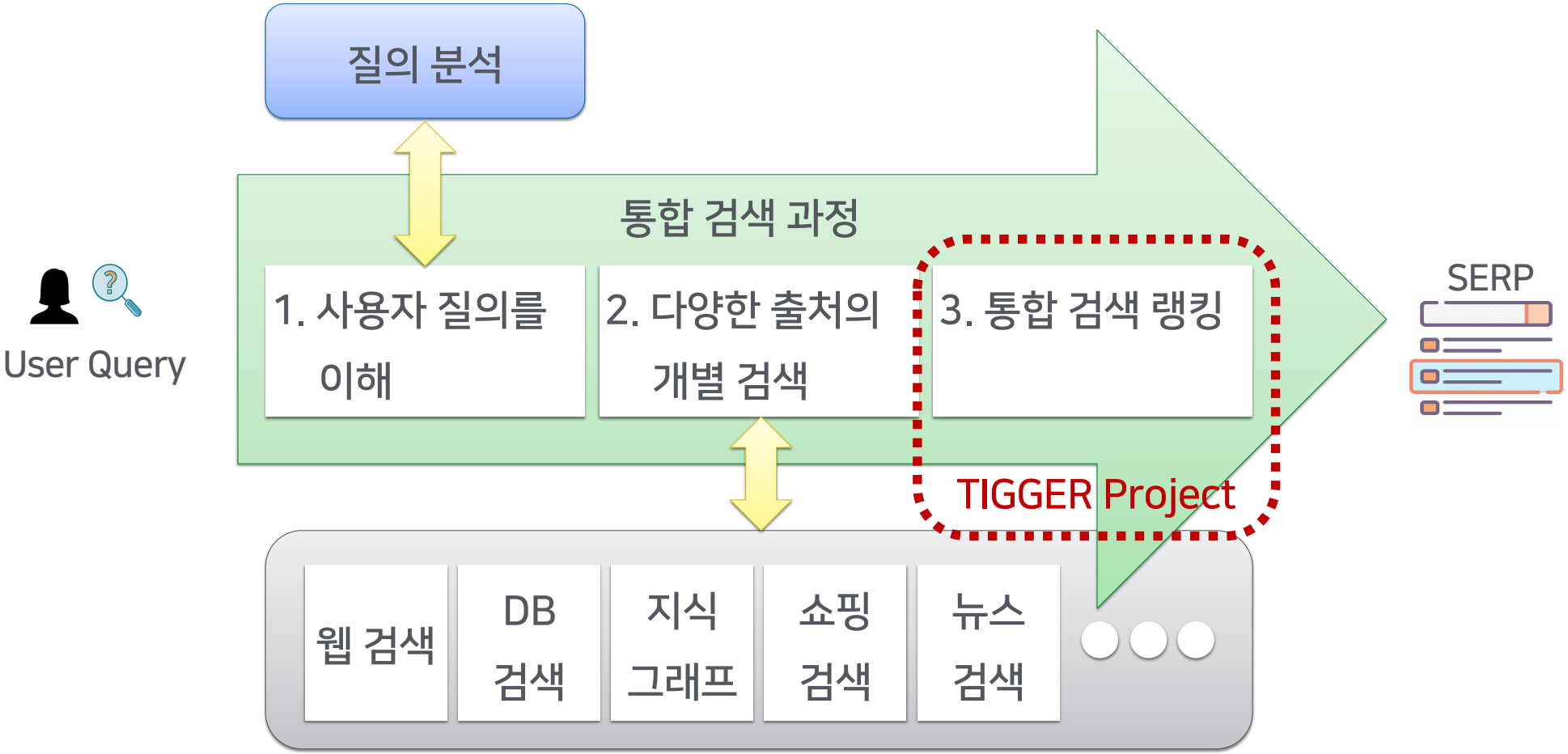
LIBERO – Context

세션질의 수에 따른 성능 비교

- 모델에서 사용될 적정 세션질의 수를 파악하기 위해 실험을 진행하였습니다.
- 실험 결과 short, mid 질의에 대해서는 세션질의 수가 증가함에 따라 성능도 증가하고, long 질의의 경우 1개 이후로 수렴하는 걸 볼 수 있습니다.
- 이 실험 결과에서도 세션질의 추가가 짧은 질의에서 특히 효과를 보이는 걸 알 수 있습니다.
- 이러한 경향과 계산 비용을 고려하여 5개의 세션 질의를 사용하였습니다.

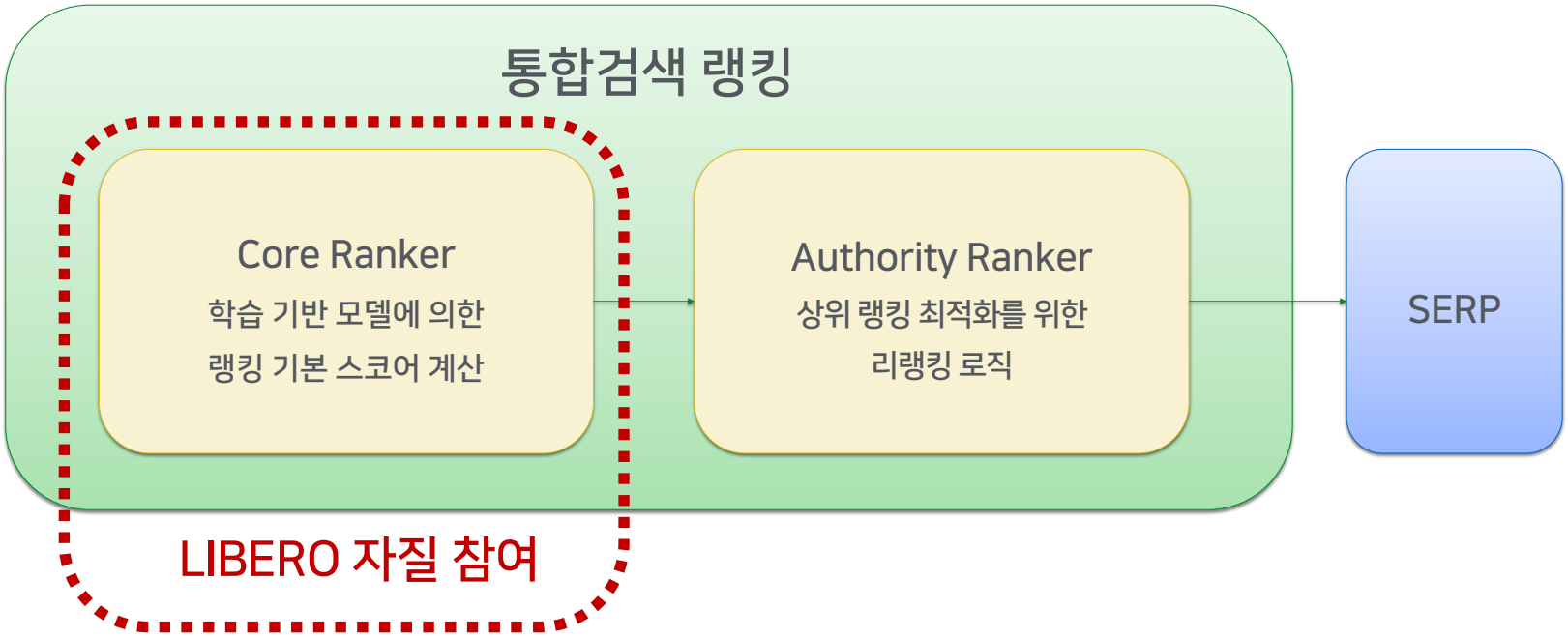


TIGGER Project

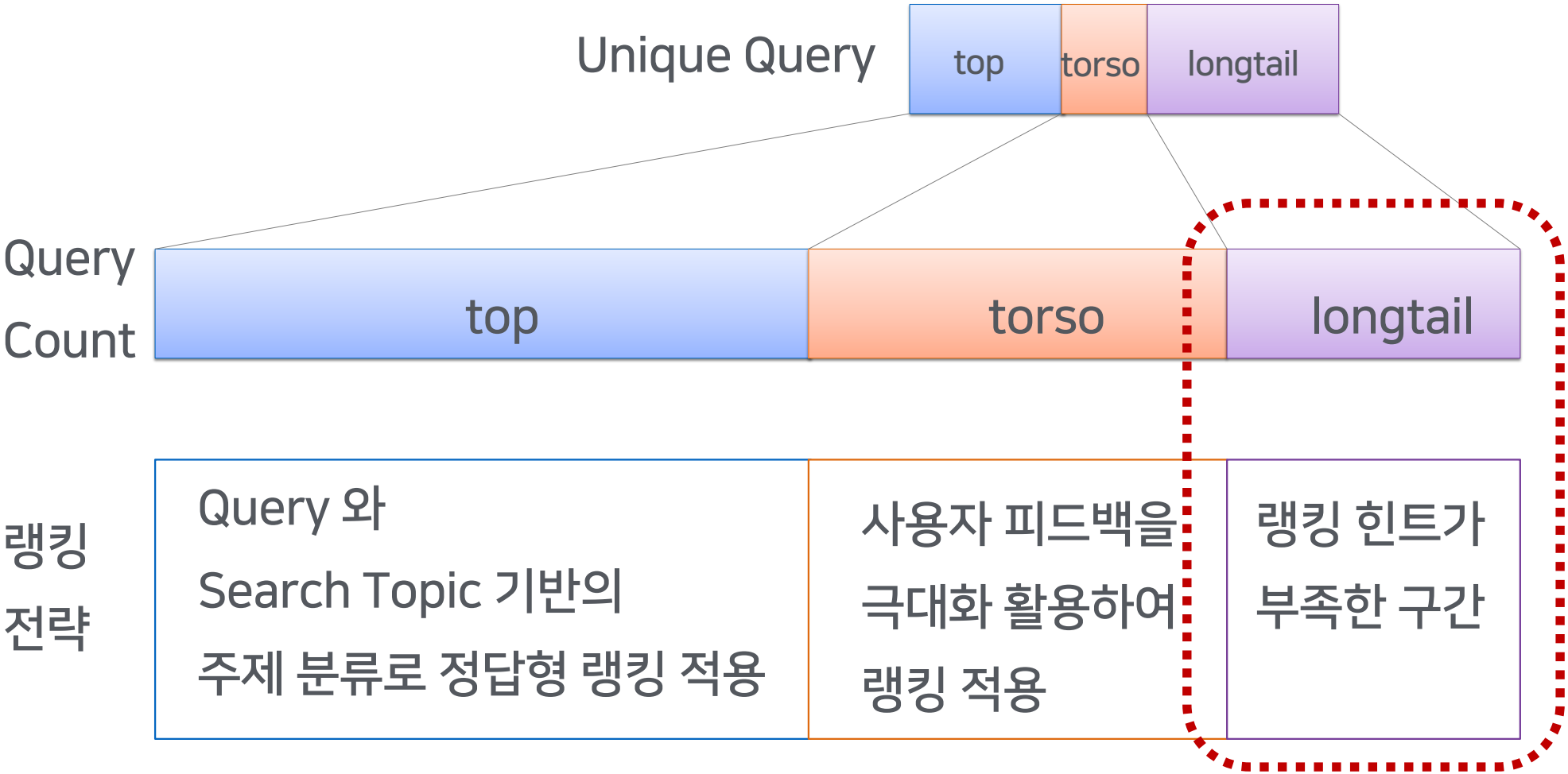


통합검색 랭킹 구조

1. Core Ranker – 랭킹 기본 스코어 계산
2. Authority Ranker – 상위 랭킹 최적화를 위한 리랭킹

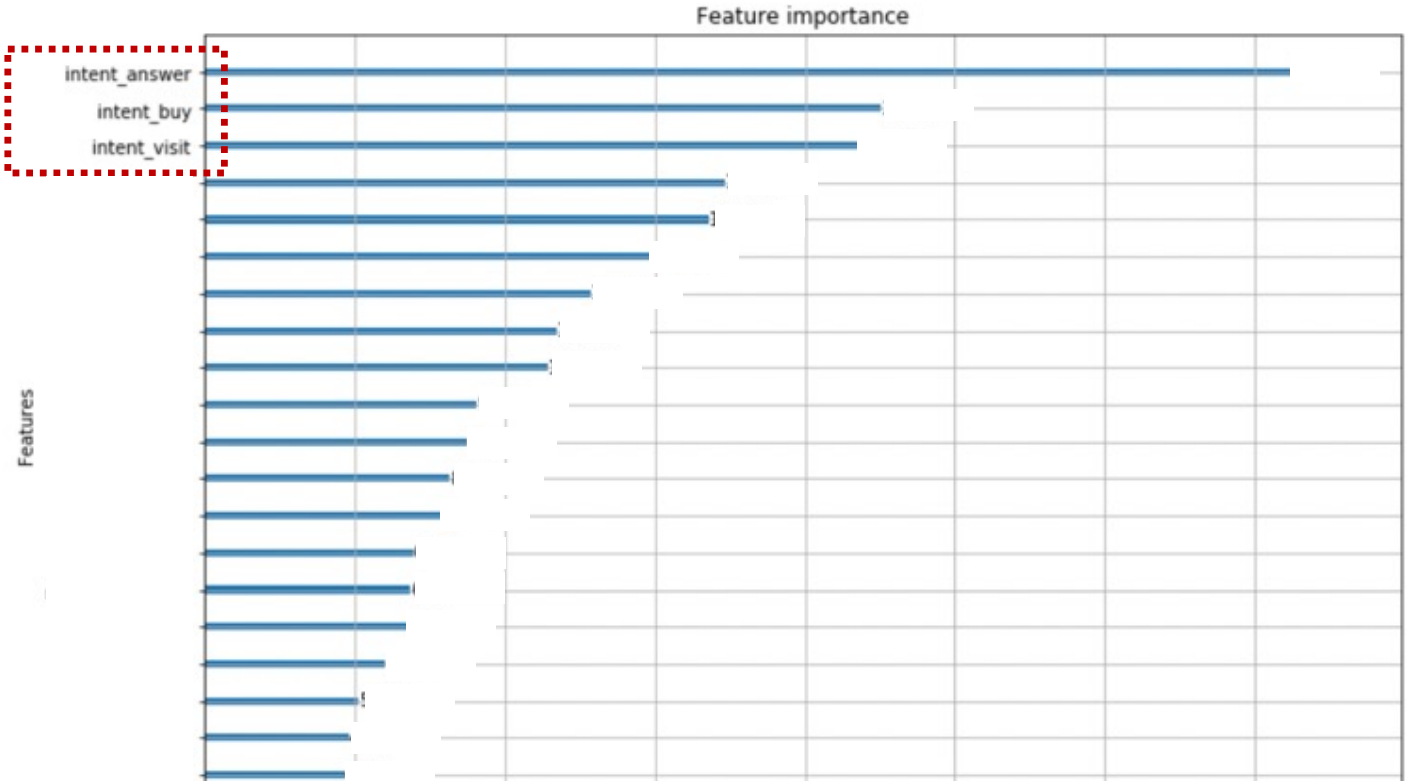


질의 구간 별 랭킹 특징

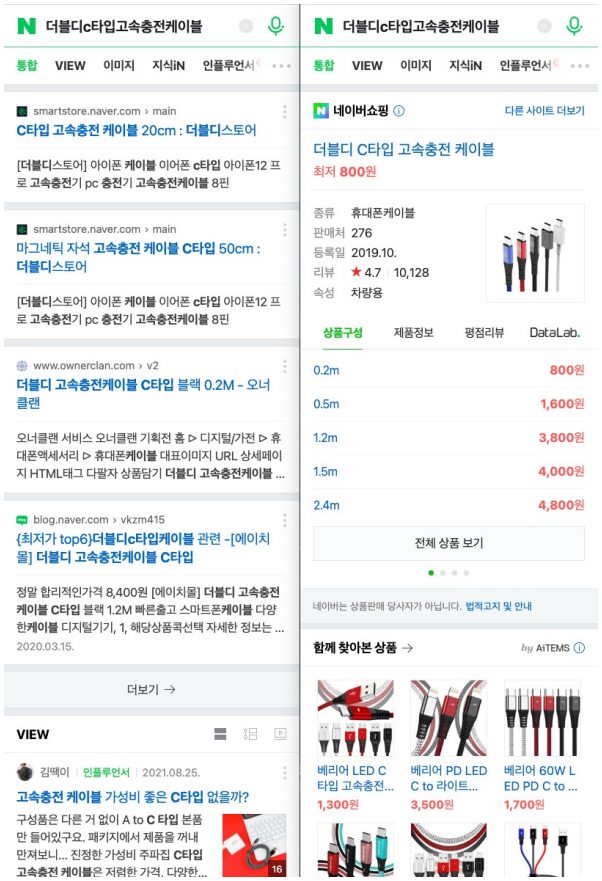

Unique Query
top
torso
longtail
Query Count
top
torso
longtail
랭킹 전략
Query 와 Search Topic 기반의 주제 분류로 정답형 랭킹 적용
사용자 피드백을 극대화 활용하여 랭킹 적용
랭킹 힌트가 부족한 구간

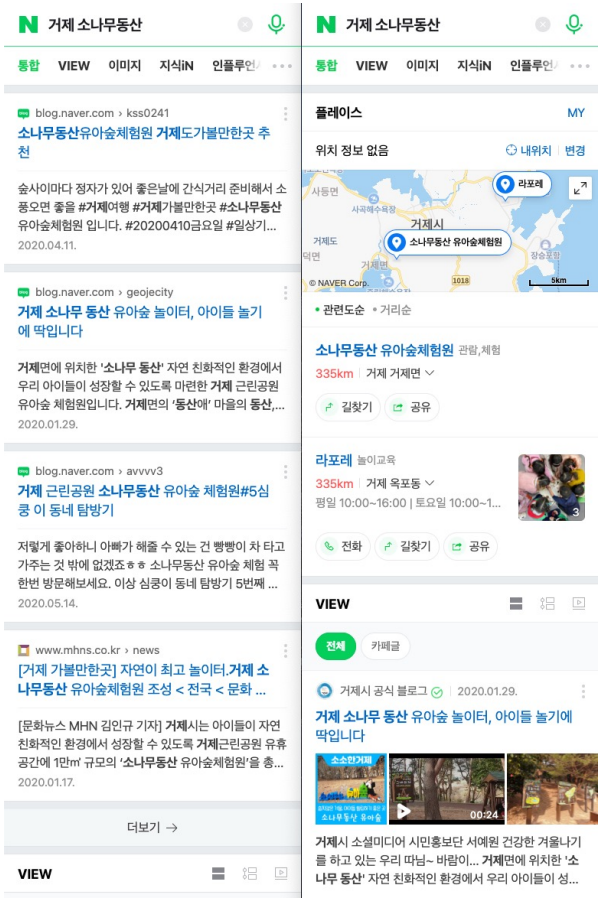
모델의 특징 및 실험 성능

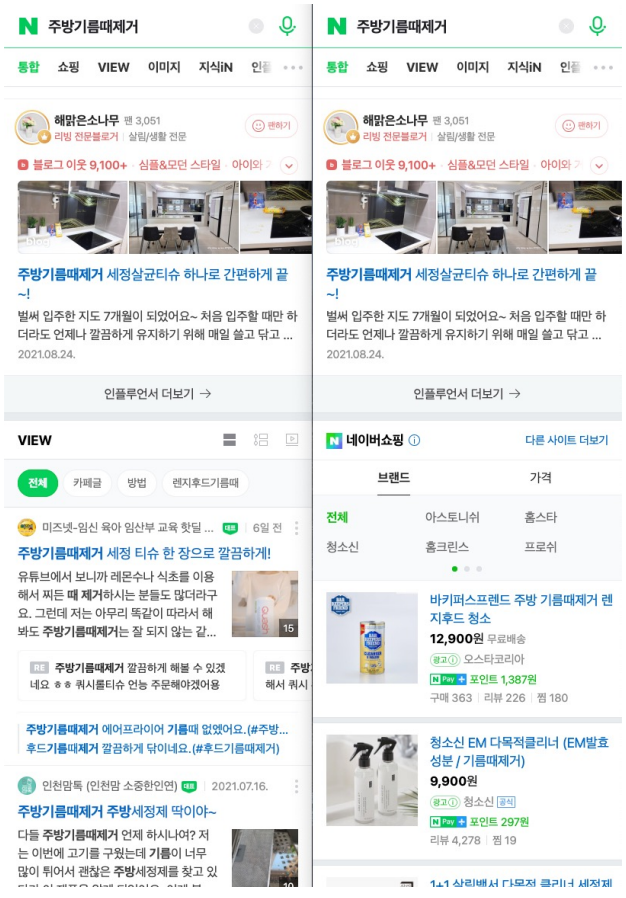
- 질의 의도 자질 적극 참여
- 특히 **즉답형, 쇼핑/플레이스** 의도에서의 품질 개선 기대



개선 사례



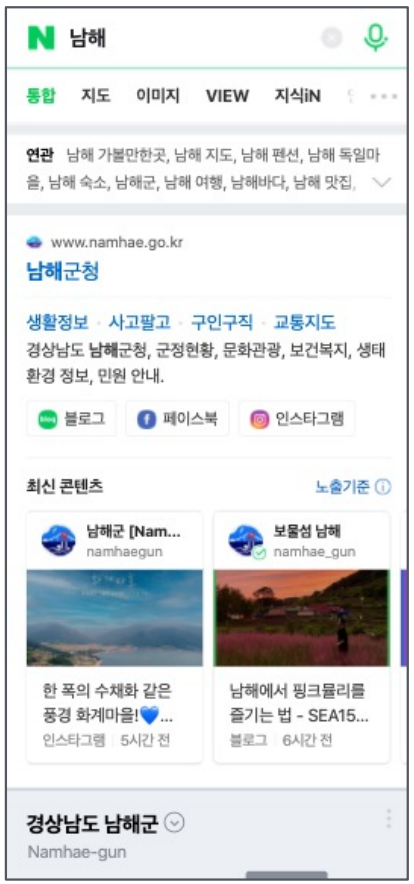


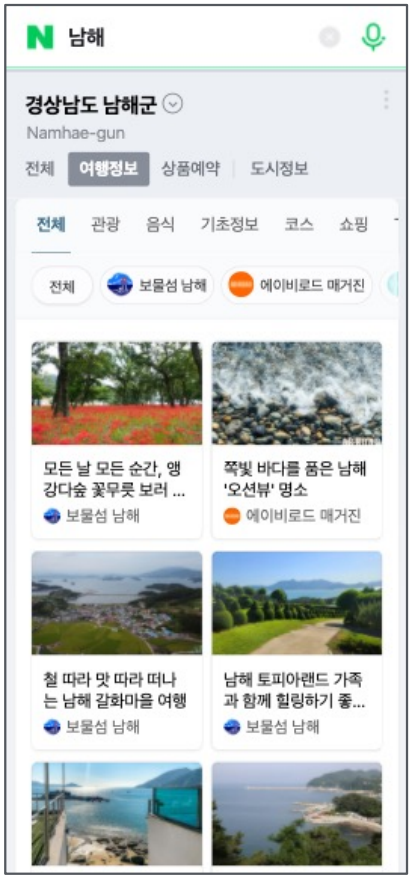


첫 클릭 랭킹 지표 8% 개선

2. 사용자에 따라 검색결과 최적화

Q=남해

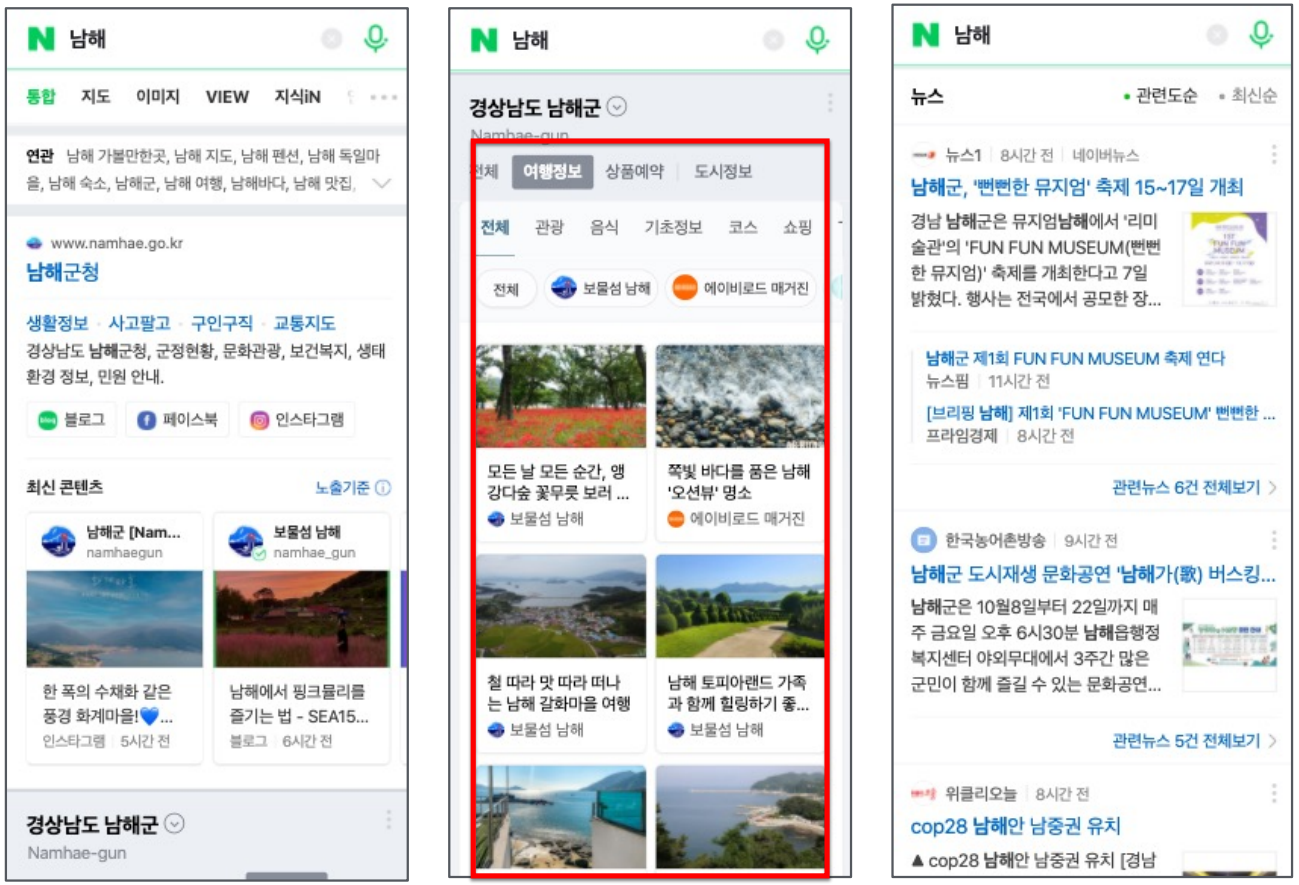




일반적인 의도 분류

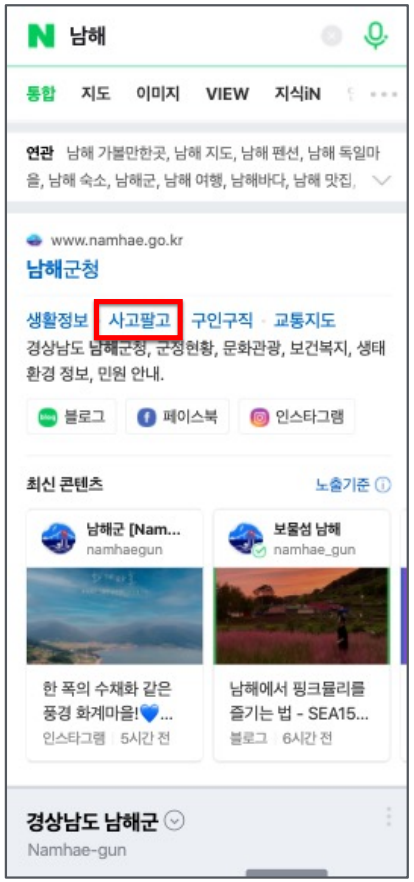
- location : 0.93
- know : 0.82
- shopping : 0.01
- ...

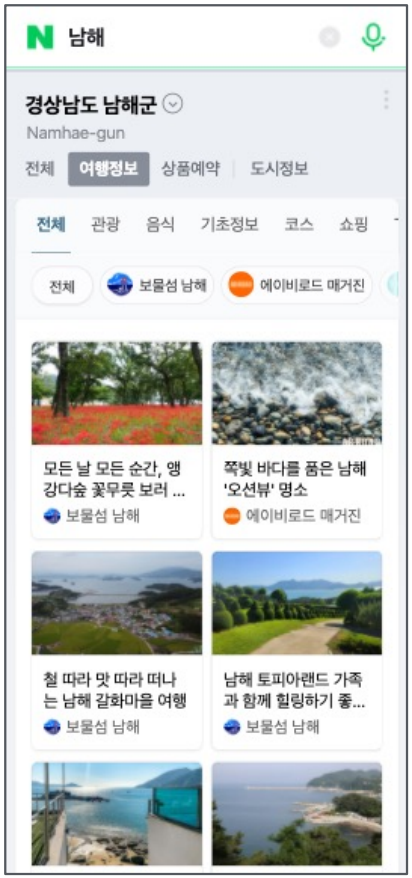
Q1=국내여행 -> Q2=남해



직전 검색어가
"국내여행"인 사용자에게는
여행 컨텐츠

매주 Q=남해



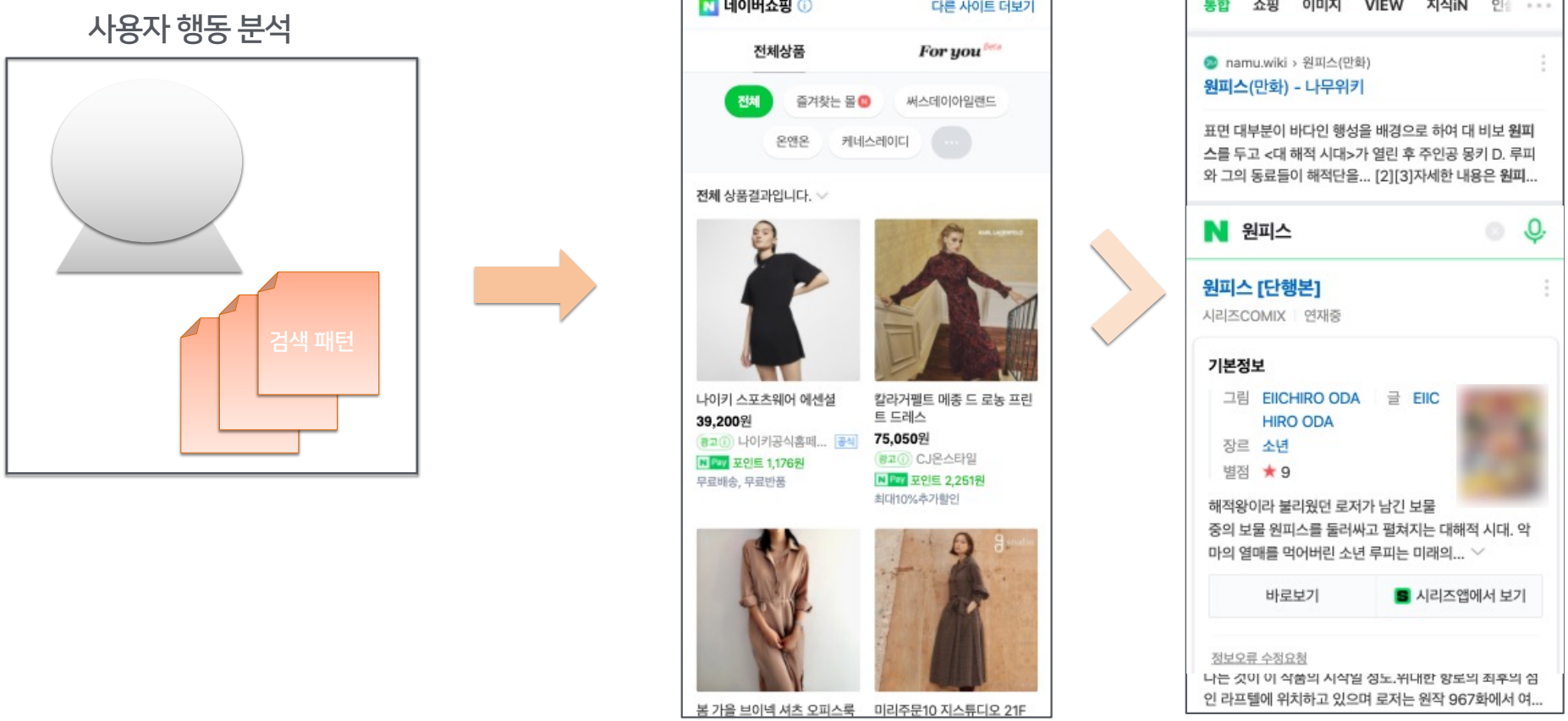


매주 "남해" 검색하고
"사고팔고" 클릭하는 사용자에게는
남해군청 사이트

2.1 사용자 이해

- 사용자에 따라 검색 의도는 다를 수 있음
- 취향/행동을 분석 -> 검색결과 최적화

서비스 적용 예제1) 쇼핑을 즐겨하는 사용자의 검색 결과 최적화



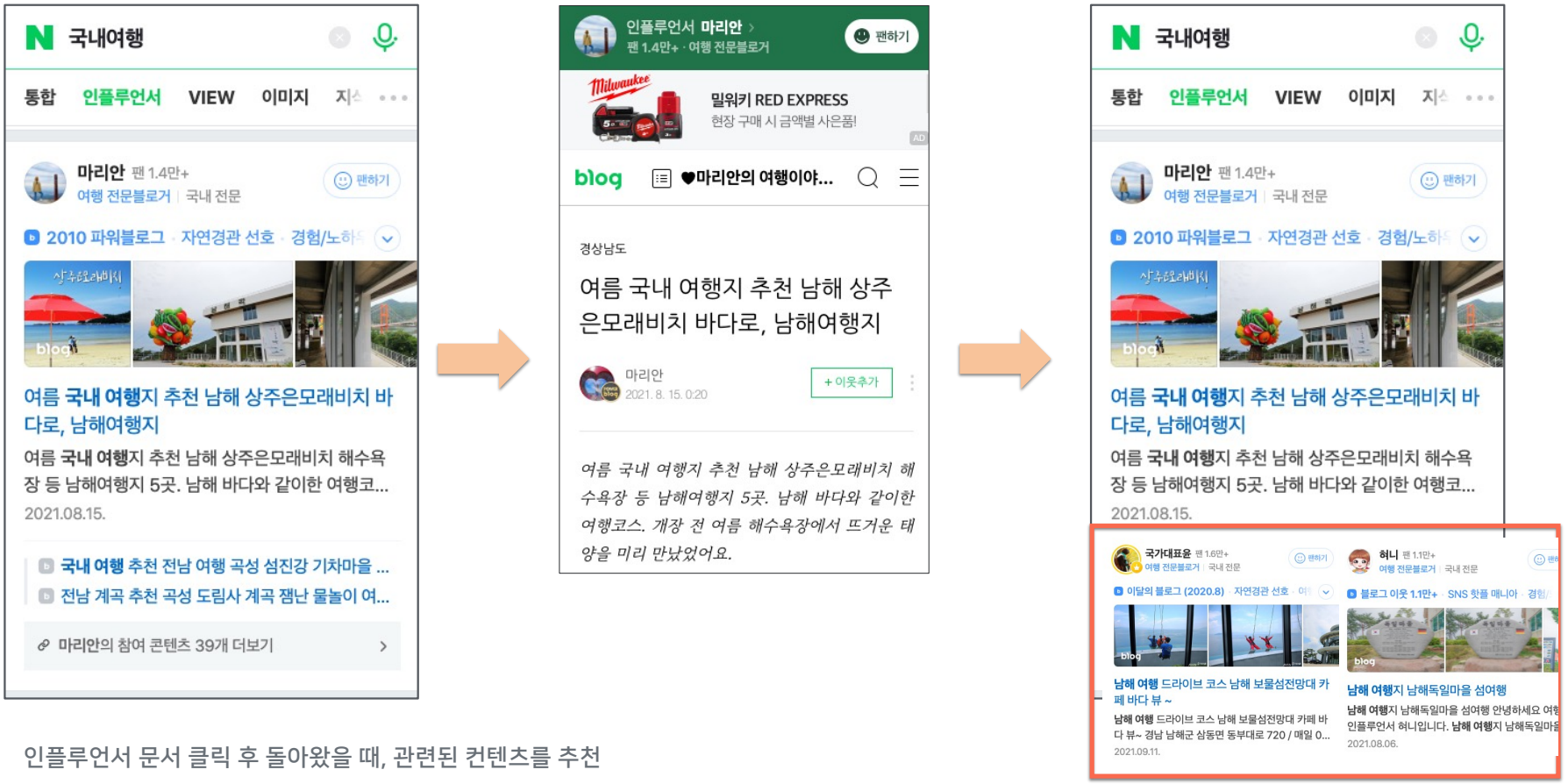
서비스 적용 예제2) 문서 소비 후 관련 상품 재검색 추천

문서 클릭 후 돌아왔을 때 검색어를 추천

https://blog.naver.com/naver_search/2223293431

2.2 반응형 검색 개발 소개

2.2 반응형 검색 개발 소개 : 인플루언서



인플루언서 문서 클릭 후 돌아왔을 때, 관련된 컨텐츠를 추천

반응형 검색의 문제 정의

Input

- 검색어 Q
- 클릭한 문서 D
- ~~사용자 취향 정보~~

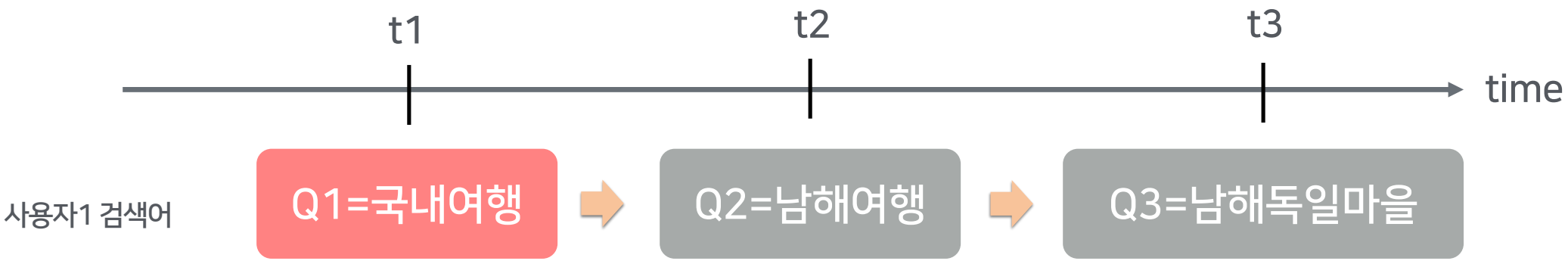
Output

- 사용자가 클릭할 법한 문서들

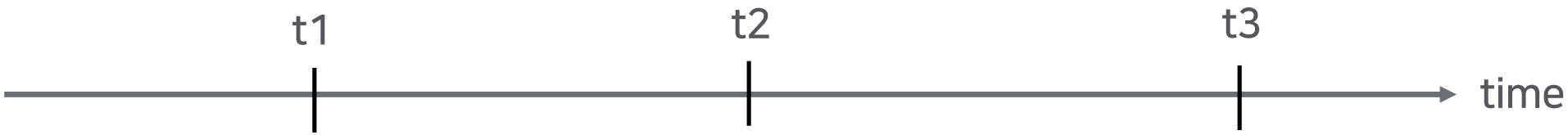
Goal

- 사용자의 검색 탐색 시간 감소
- 다양한 정보 제공

문제 정의 : 검색 시나리오



문제 정의 : 검색 시나리오



사용자1 검색어

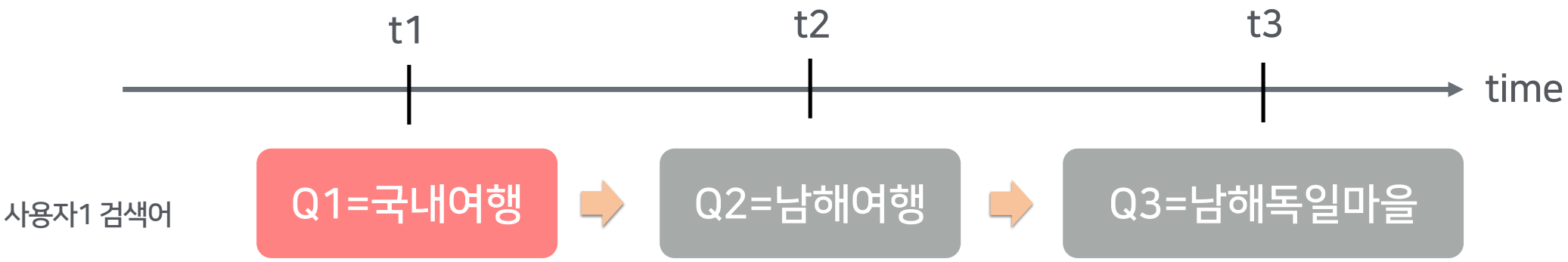
Q1=국내여행 → Q2=남해여행 → Q3=남해독일마을

검색 결과

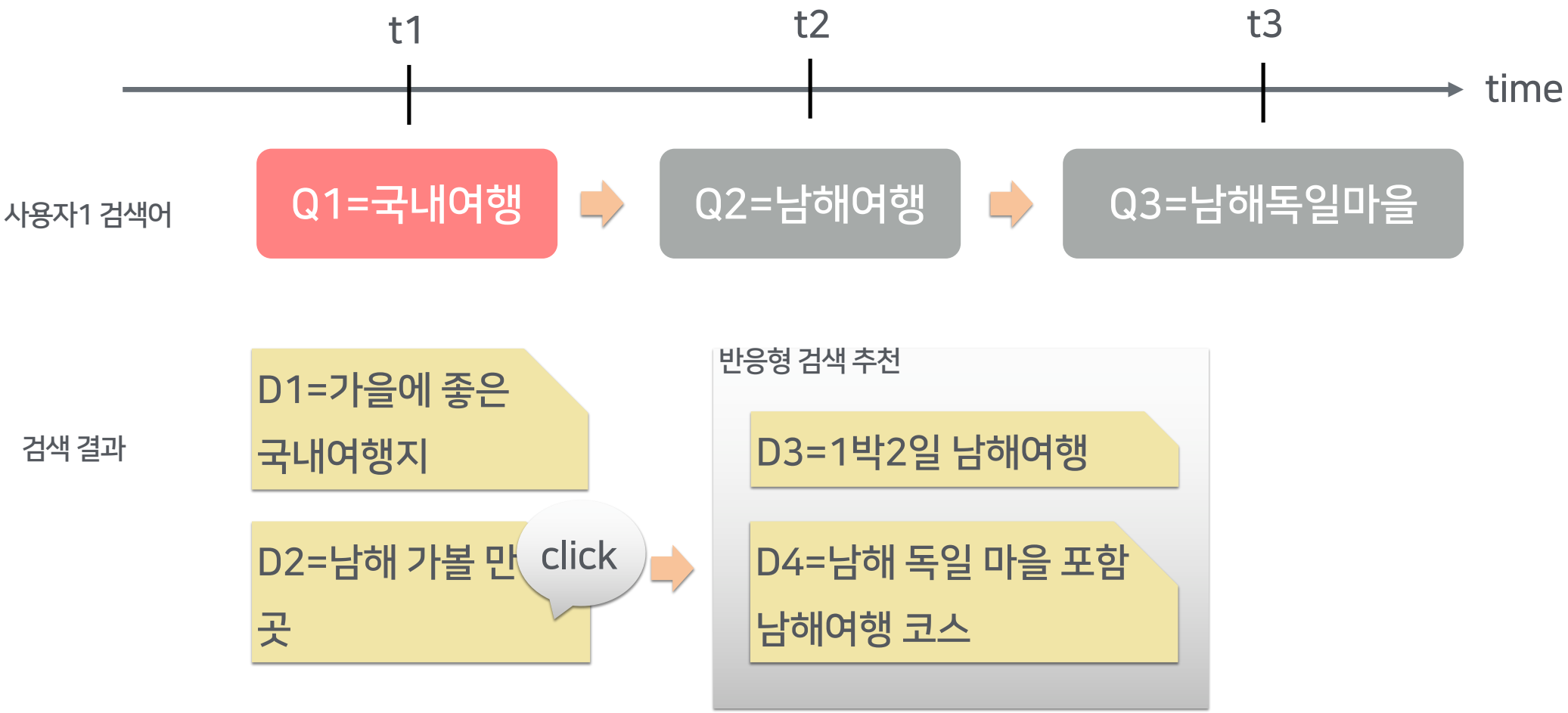
D1=가을에 좋은 국내여행지

D2=남해 가볼 만한 곳

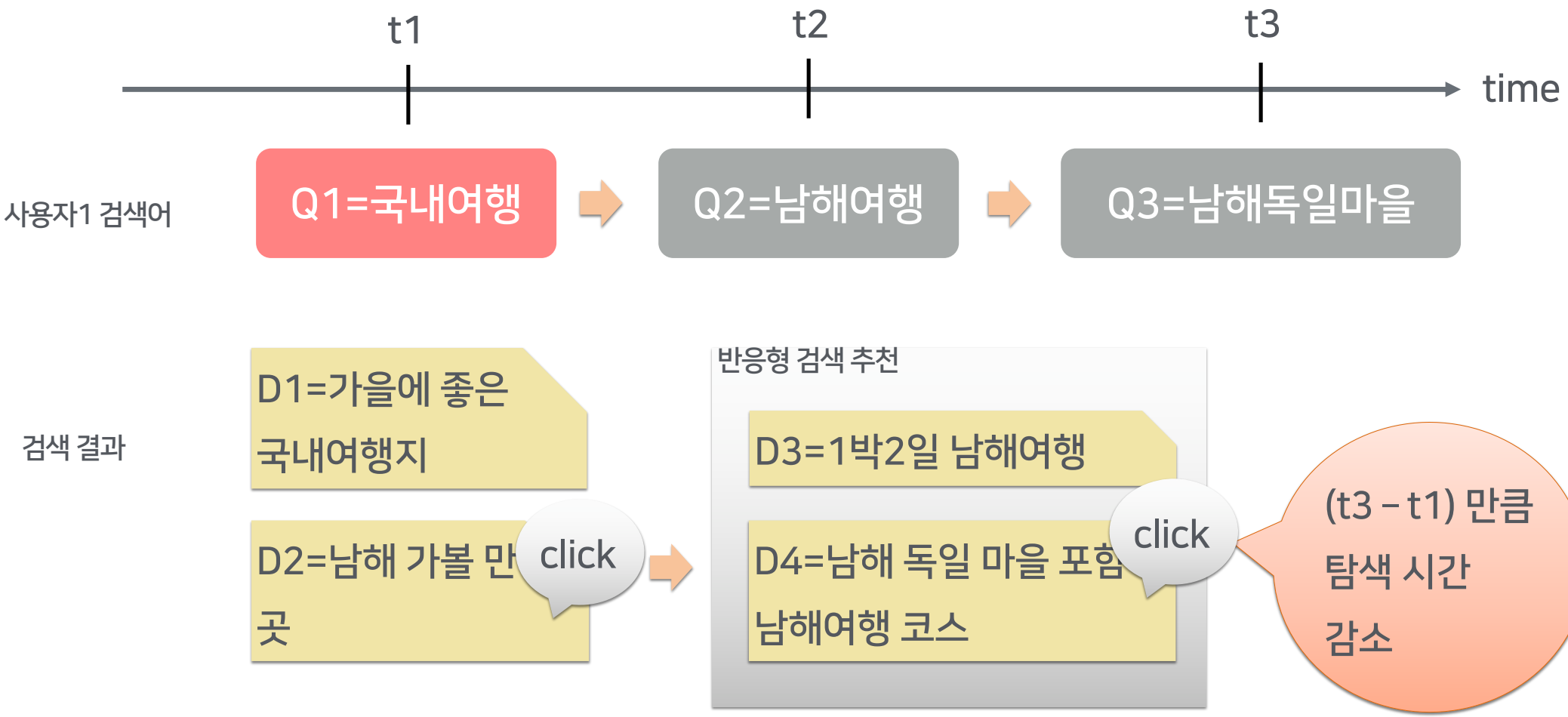
문제 정의 : 검색 시나리오


t1
t2
t3
time
사용자1 검색어
Q1=국내여행
Q2=남해여행
Q3=남해독일마을
검색 결과
D1=가을에 좋은 국내여행지
D2=남해 가볼 만한 곳
click

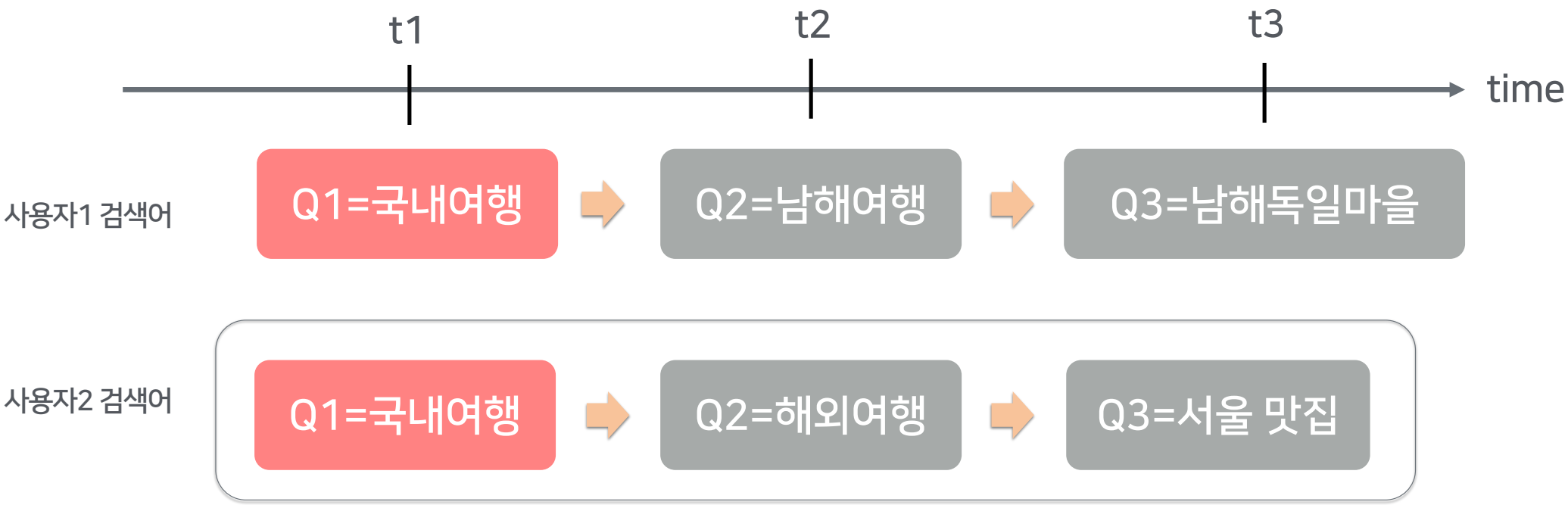
문제 정의 : 검색 시나리오


t1
t2
t3
time
사용자1 검색어
Q1=국내여행
Q2=남해여행
Q3=남해독일마을
검색 결과
D1=가을에 좋은 국내여행지
D2=남해 가볼 만 곳
click
반응형 검색 추천
D3=1박2일 남해여행
D4=남해 독일 마을 포함 남해여행 코스

문제 정의 : 검색 시나리오


t1
t2
t3
time
사용자1 검색어
Q1=국내여행
Q2=남해여행
Q3=남해독일마을
검색 결과
D1=가을에 좋은 국내여행지
D2=남해 가볼 만 곳
click
반응형 검색 추천
D3=1박2일 남해여행
D4=남해 독일 마을 포함 남해여행 코스
click
(t3 – t1) 만큼 탐색 시간 감소

문제 정의 : 가정

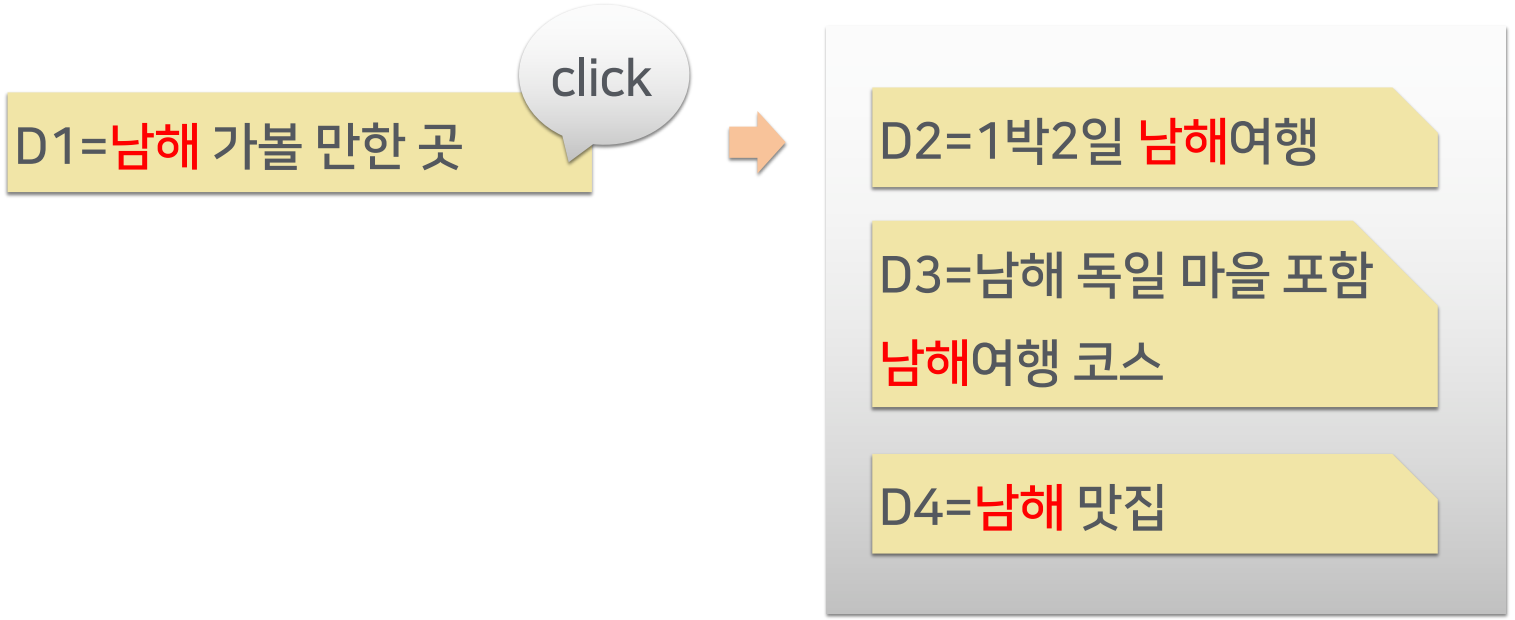


주제 확장 -> 관심있는 문서를 예측하기 어려움

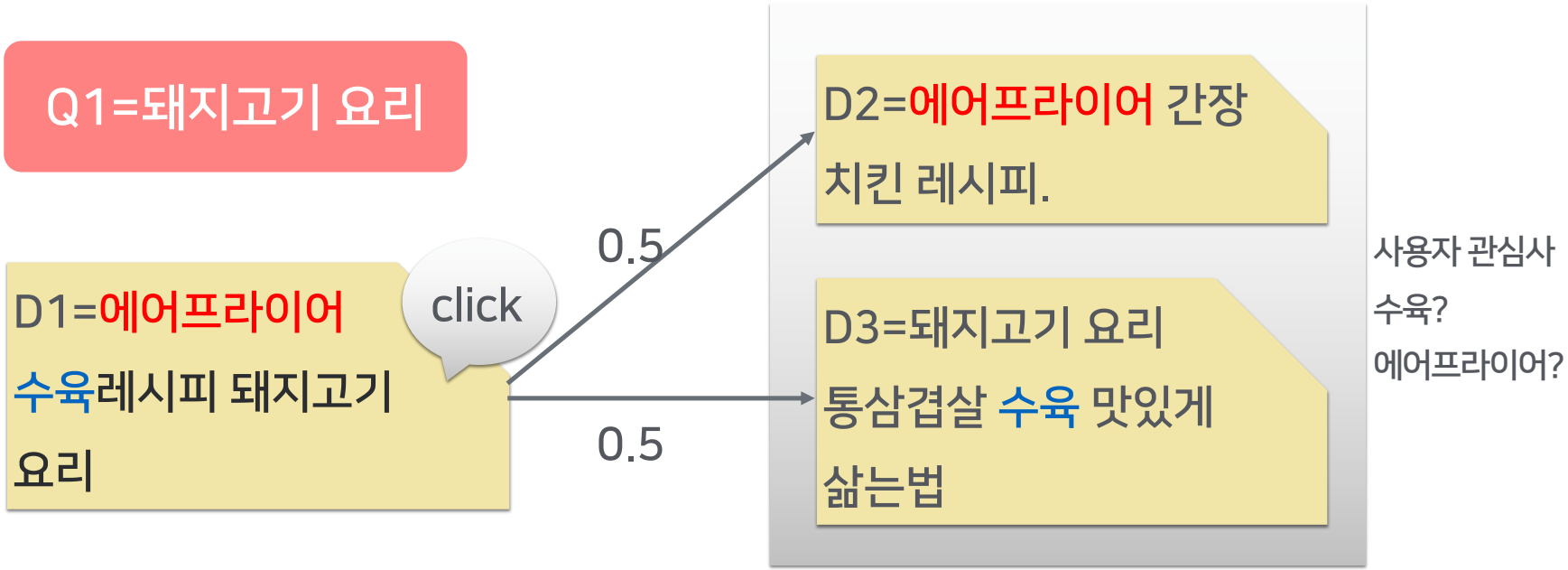
사용자가 의도를 좁혀가면서 검색을 수행한다고 가정

Baseline 모델 : 클릭 문서 기반 추천

- 클릭한 문서와 제목이 유사하거나, 문서 클릭한 사용자들이 공통적으로 선호 문서

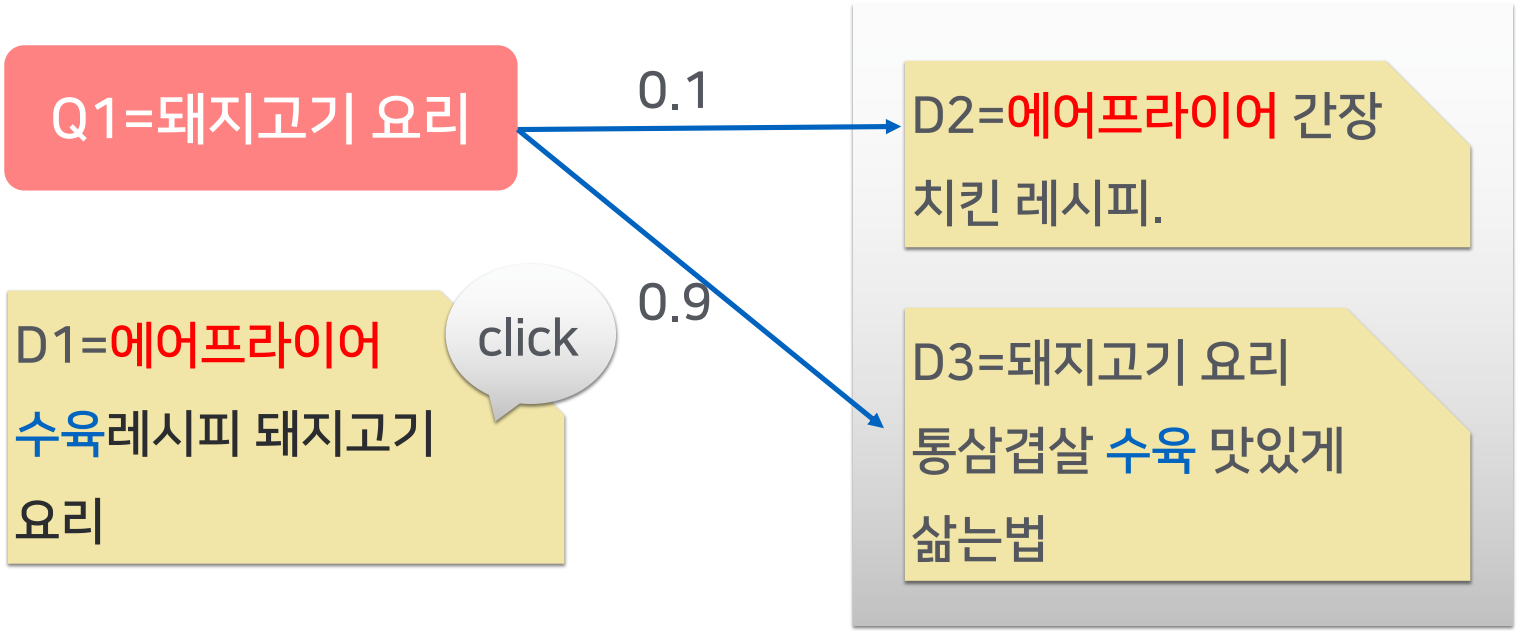


Baseline 모델의 문제점



문서 기반 추천만으로는 사용자 관심사에서 벗어날 가능성이 있음

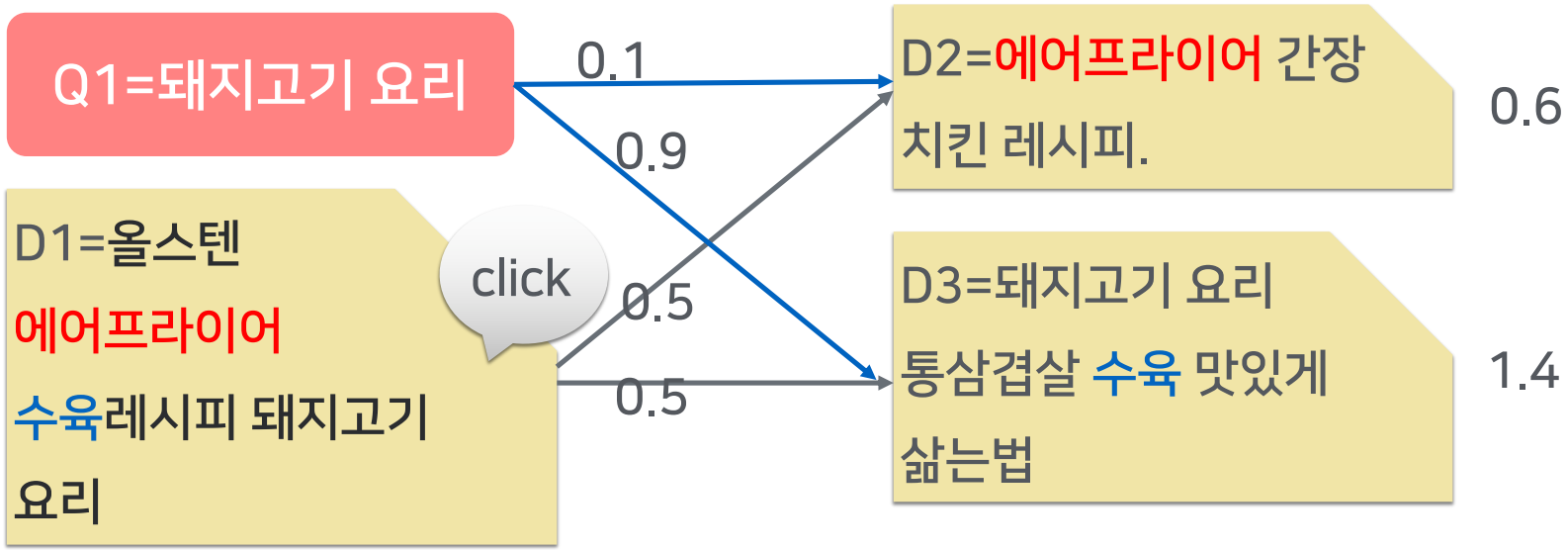
Baseline 모델의 문제 해결 : Query 정보 이용



Query 정보를 이용하여 사용자의 관심 있는 문서를 더 높은 정확도로 예측 가능

해결 방안 : 검색과 추천 기술 융합

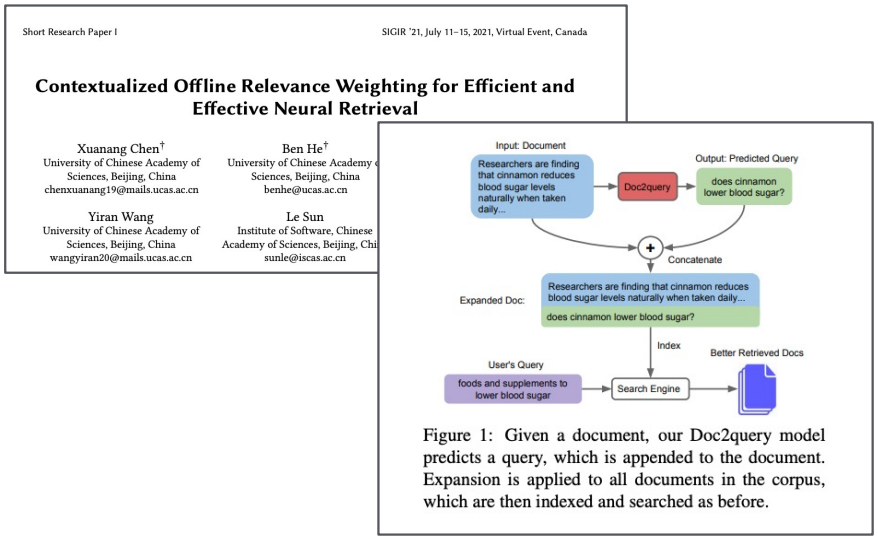
- 검색 : Query에 가장 적합한 문서를 랭킹 (bm25, semantic matching, doc2query 등)
- 추천 : 소비한 컨텐츠를 기반으로 클릭할 문서를 예측



Feature Example : Doc2Query

DEVIEW 2021

- Doc2Query : 문서마다 적합할 질의를 추출하는 기술
- 사용자가 문서를 클릭했을 때, 관심사를 예측할 때 사용
  - 로그 기반에서 추출 -> Language Model 기반으로 확장 진행 중



서비스 적용

DEVIEW 2021

- 11월 말 네이버 통합검색 및 인플루언서 탭검 적용 목표로 개발 중

Q1=돼지고기 요리

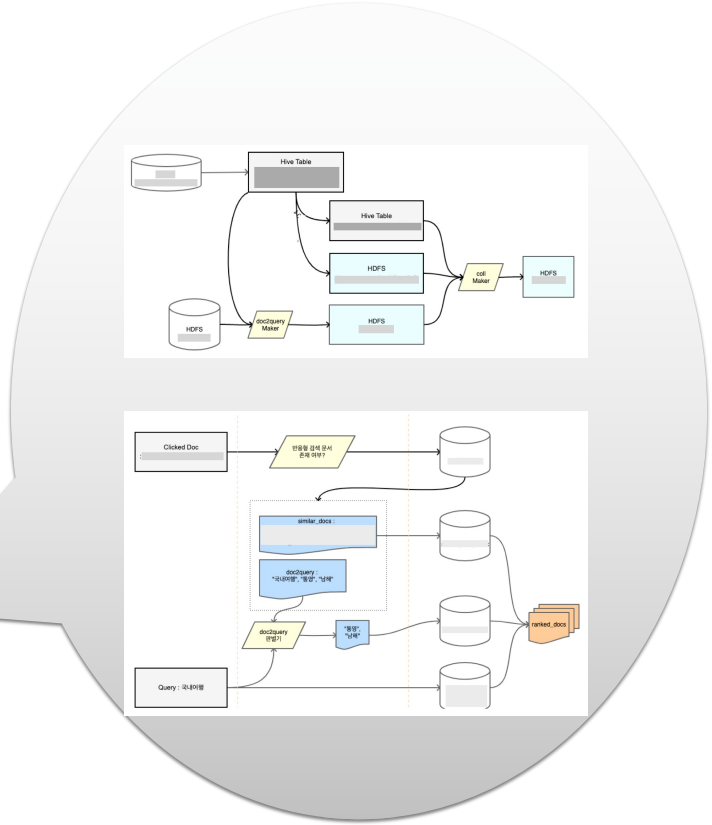
D1=올스텐 에어프라이어 수육레시피 돼지고기 요리

click

반응형 검색 추천

D3=돼지고기 요리 통삼겹살 수육 맛있게 삶는법 1.4

D2=에어프라이어 간장 치킨 레시피. 0.7



Future Work : 문제 확장 - 유저 정보 사용

DEVIEW 2021

Input

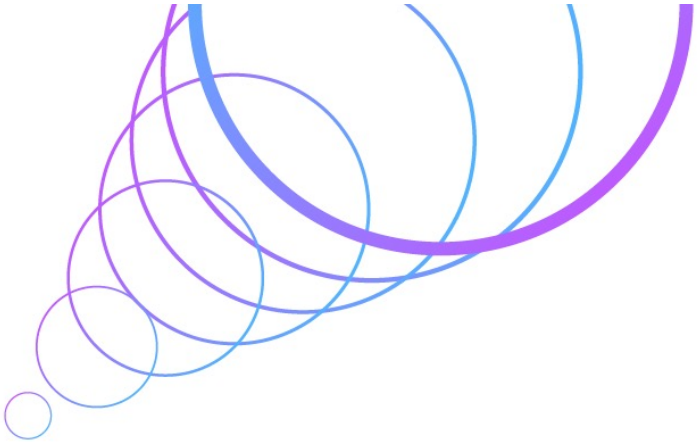
- 검색어 Q
- 클릭한 문서 D
- 사용자 정보

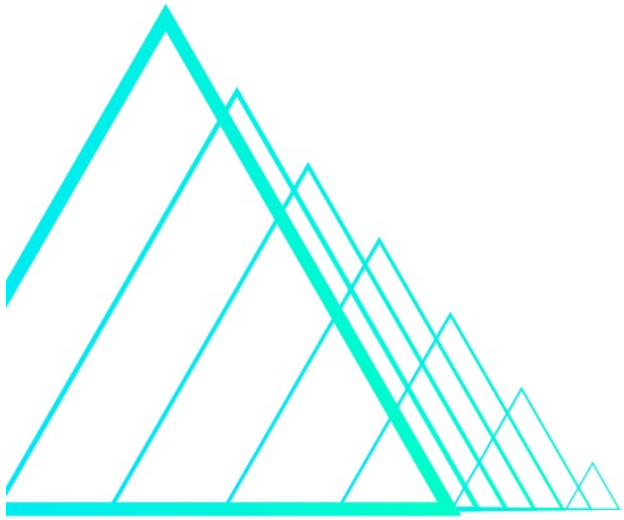
Output

- 사용자가 클릭할 법한 문서들

사용자의 과거 이력과 취향을 분석해서(*1)
사용자 별 만족도 개선을 목표

(*1)발표에서 사용한 모든 로그는 직접적인 개인 식별이 불가능하도록 비식별 상태로 분석에 사용되었습니다.





Thank You