



반갑습니다.  
정입니다.

**MetisX** 에서 **Field Application Engineer**  
업무를 맡고 있어요.

그 전에는 그래프 데이터베이스 , 그래프  
솔루션 회사를 다니며  
산전수전공중전까지 겪어본 경험이  
있습니다.

# 오늘 여러분들에게 전달드릴 지식.

---

다음과 같은 콘텐츠를 가져왔어요.

## 1.Graph Understanding

그래프 이해도 상승을 위해 필수적으로 알아두면 좋을 요소들

## 2.Graph Good Use Case

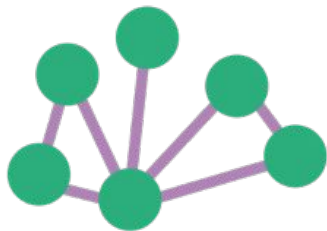
E-commerce 분야에서 검색 엔진성능 향상을 위해 시스템을 구축한 사례

## 3.Graph + LLM(Foundation Model)

KDD 24' 트렌드 Graph AGI 포인트인 Graph Foundation Model 부터 Graph를 LLM에 적용하는 방식까지

# Graph Understanding

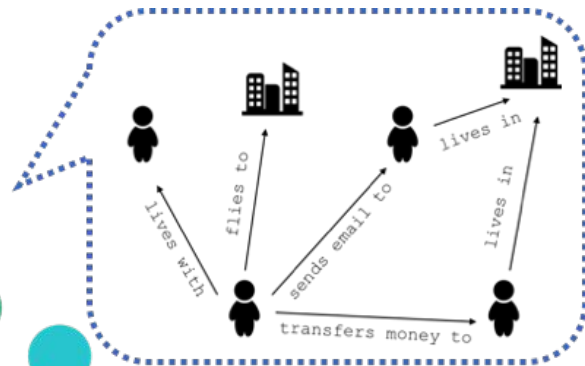
그래프 구성요소인 노드, 엣지 종류가 1개 초과인지 아닌지에 따라 homogeneous , heterogeneous로 나뉘어요.



homogeneous



heterogeneous



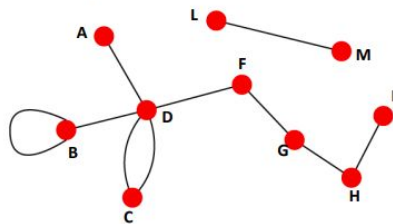
# Graph Understanding

이외에도 그래프 엣지에 방향성, Weights Types Properties Attributes 에 따라 그래프의 성격이 바뀌어요.

## Directed vs. Undirected Graphs

### Undirected

- **Links:** undirected  
(symmetrical, reciprocal)

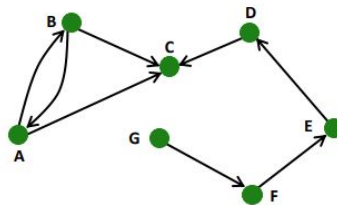


- **Other considerations:**

- Weights
- Properties

### Directed

- **Links:** directed

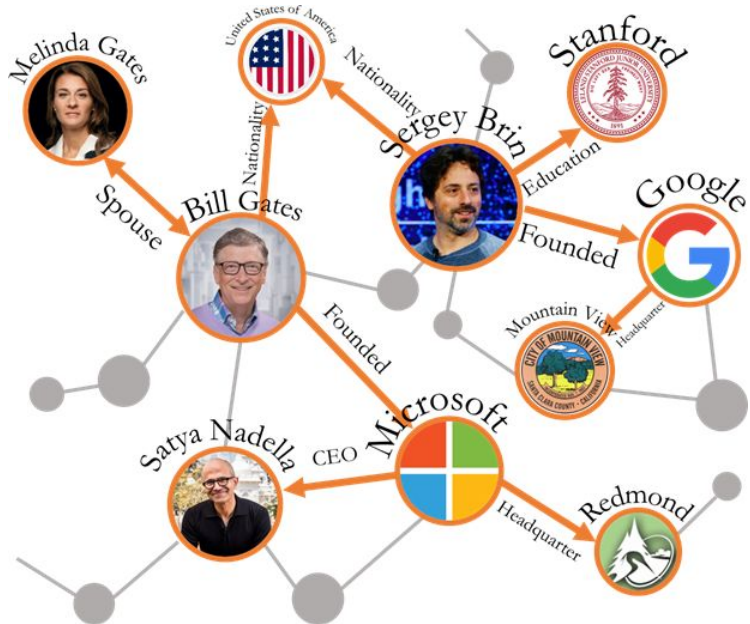


- Types
- Attributes

# Graph Understanding

지식 그래프(Knowledge graph) 또한 heterogeneous graph 에 속해요. Google은 지식 그래프를 검색엔진에 활용했고, 이를 통해 시장에서 급부상할 수 있었죠.

The image shows a Google search result for "Florence Price". Red boxes highlight several knowledge panels: the top image gallery, the "Listen" section with streaming service icons, the "About" section with biographical text, and the "Songs" section with a list of her works. Arrows point from these panels to the right, where a graph visualization is shown.



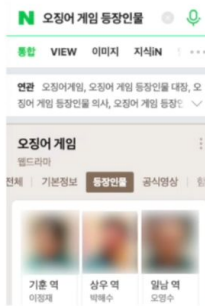
# Graph Understanding

우리나라에서도 지식그래프를 통해 검색 엔진 개선을 하고 있어요.



## 사용자 질의를 이해하는 방식 - Search Topic & Goal

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

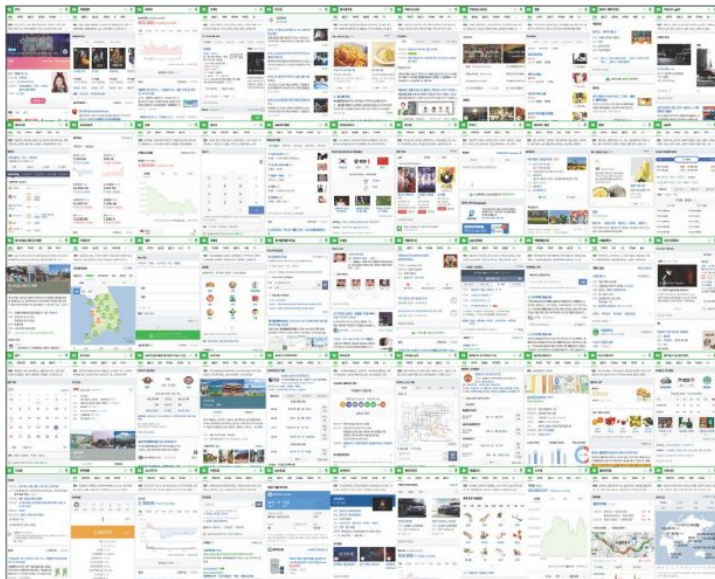


# Graph Understanding

지식그래프로 대응하는 질의어가 상당해요.

<2019 DEVIEW 중>

**N** DEVIEW  
2021



DEVIEW  
2019

40여개 주제 250여개 컬렉션으로  
서비스 중.

통합검색 전체 질의어 중  
약 30% 대응 중.

# Graph Understanding

지식그래프 구축을 위해 필요한 과정들. 그런데 갑자기 지식그래프에서 Ontology 그리고 RDF?

## 2.1 과정

**N** DEVIEW  
2021

### Ontology

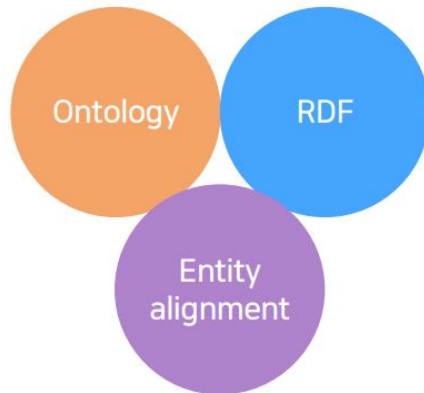
- Taxonomy (class hierarchy)
- Relations, Restrictions

### RDF 형태로 변환

- Resource Description Framework
- URI (Uniform Resource Identifier)

### Entity Alignment

- 같은 Entity를 찾아서 연결
- DB Linkage / Entity Alignment / Entity Resolution





# Graph Understanding

그래프를 표현하고 담는 방식에 따라 LPG(Label Property Graph) 와 RDF(Resource Description Framework) 크게 두 가지로 나뉘어요.

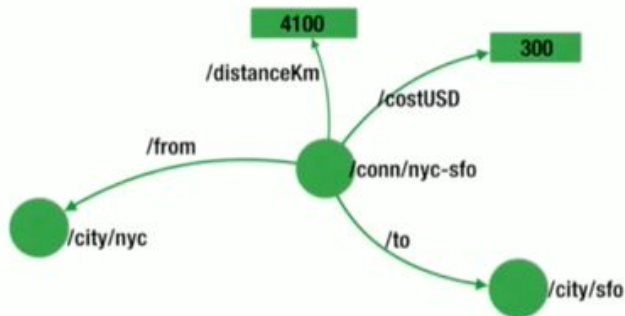
## Modeling workaround

Connection between NYC and SF: 300 USD / 4100 in Km.

LPG (Neo4j)



RDF



# Graph Understanding

---

LPG? RDF? 헷갈리는데, 언제 어떻게 무슨 기준으로 판단 및 선택해서 사용해야해요?

## When to use RDF vs. Property Graphs

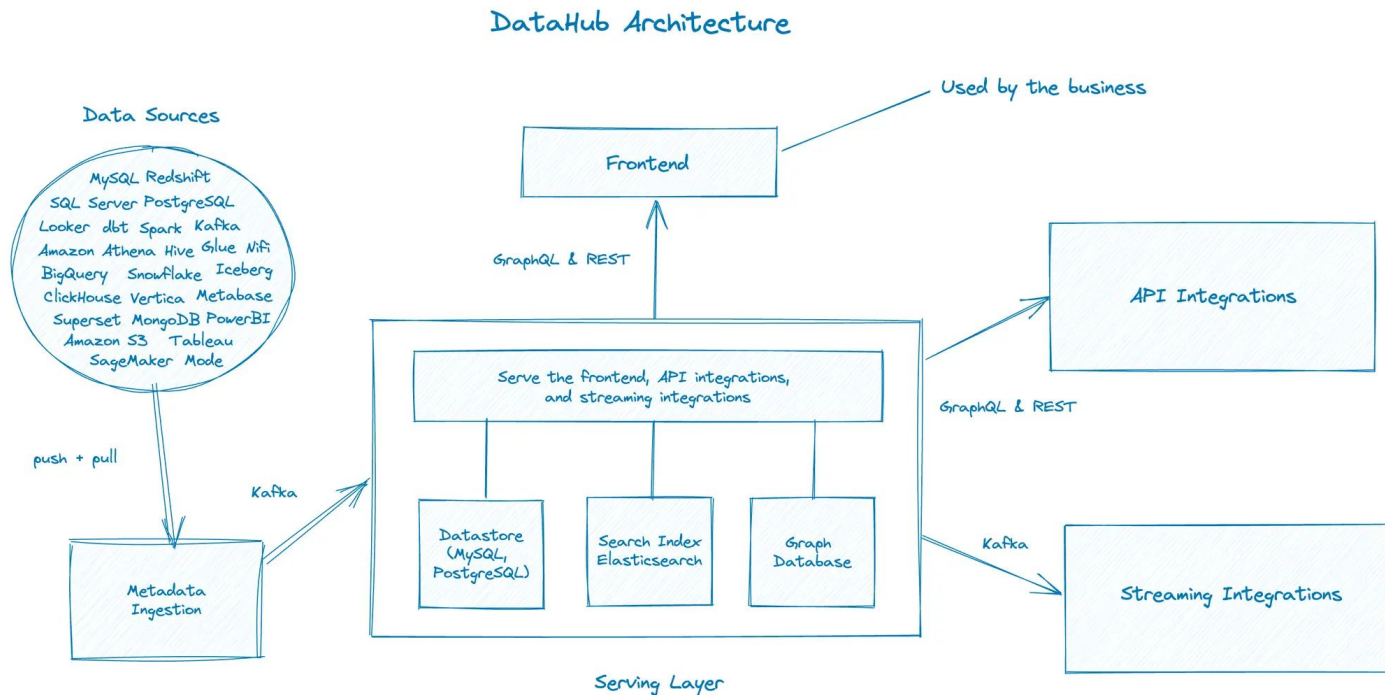
Questions about the choice of data models are at some level, a user decision. A rule of thumb is that if you have sufficient structure over your records and want to model them as a graph (e.g., to find paths, patterns, ask for recursive and/or arbitrary connections between records), you should structure them as property graphs. The general principle is that DBMSs provide fast query performance over large sets of records by exploiting structure. Outside of this, some common scenarios for using **RDF** are the following:

1. When your data is very **heterogeneous and hard to tabulate**.
2. You want to homogeneously represent and query both your data and metadata/schema information in the same format of triples. For the example above, we represented both data and schema information homogeneously as triples.
3. You need some automatic reasoning/inference capabilities.

출처 : [kuzudb](#)

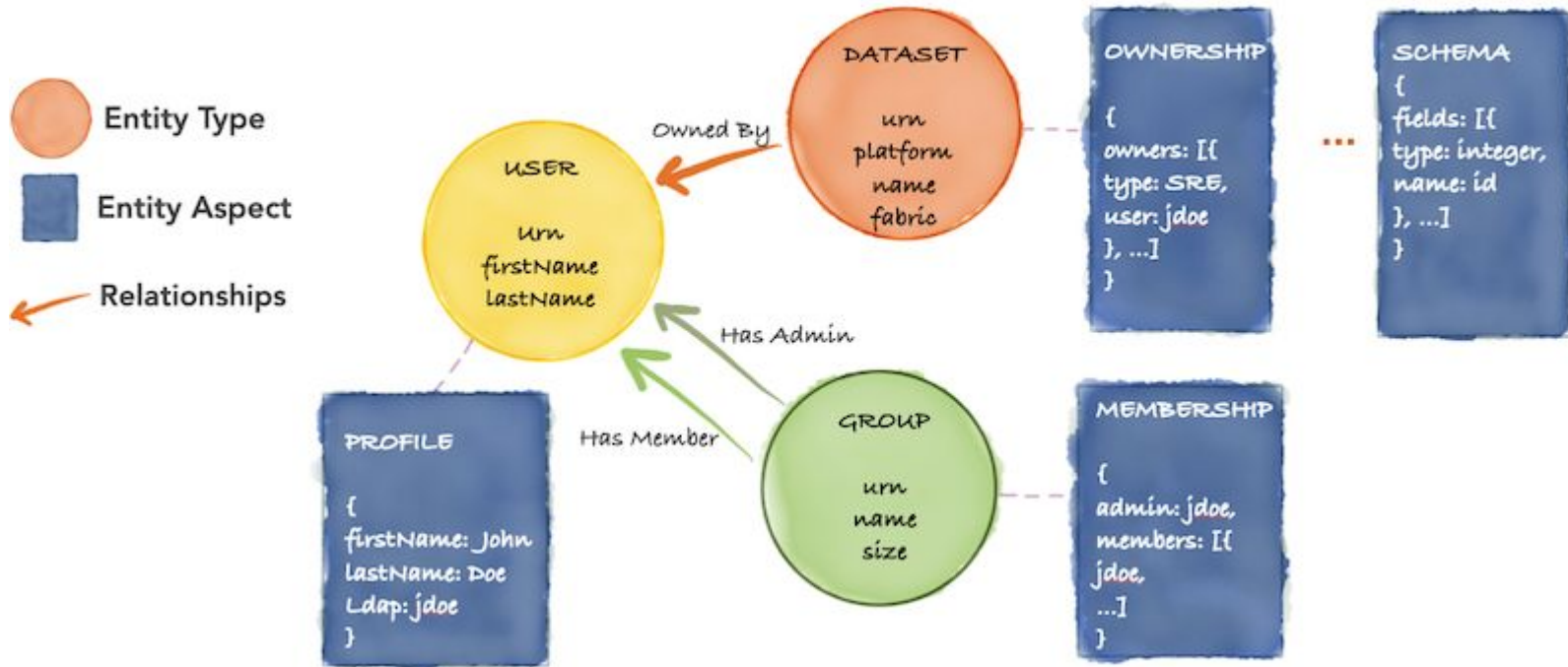
# Graph Good Use case - MDM(Master Data Management)

서두에 언급한 것처럼 데이터를 통합하고 활용할 때, 그래프 형태로 표현하는걸 선호해요. (Linkedin 사례)



# Graph Good Use case - MDM(Master Data Management)

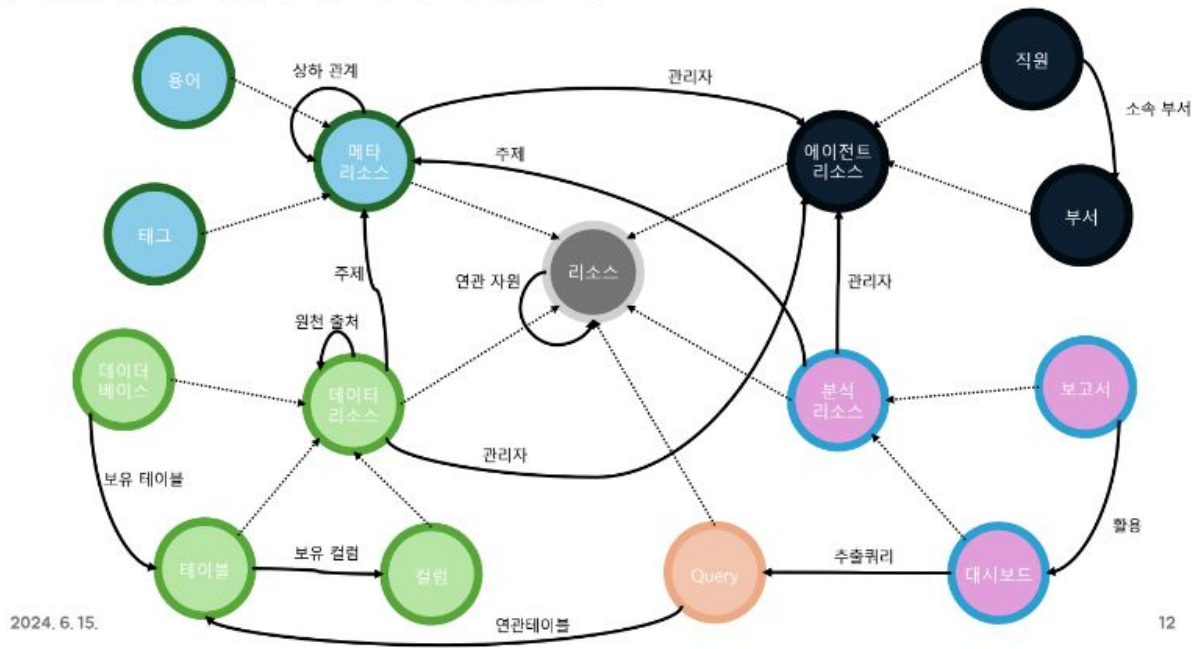
데이터간 관계를 통해 데이터를 이해하고 이를 기반으로 Business 까지



# Graph Good Use case - MDM(Master Data Management)

이를 거시적인 맥락에서 바라본다면,

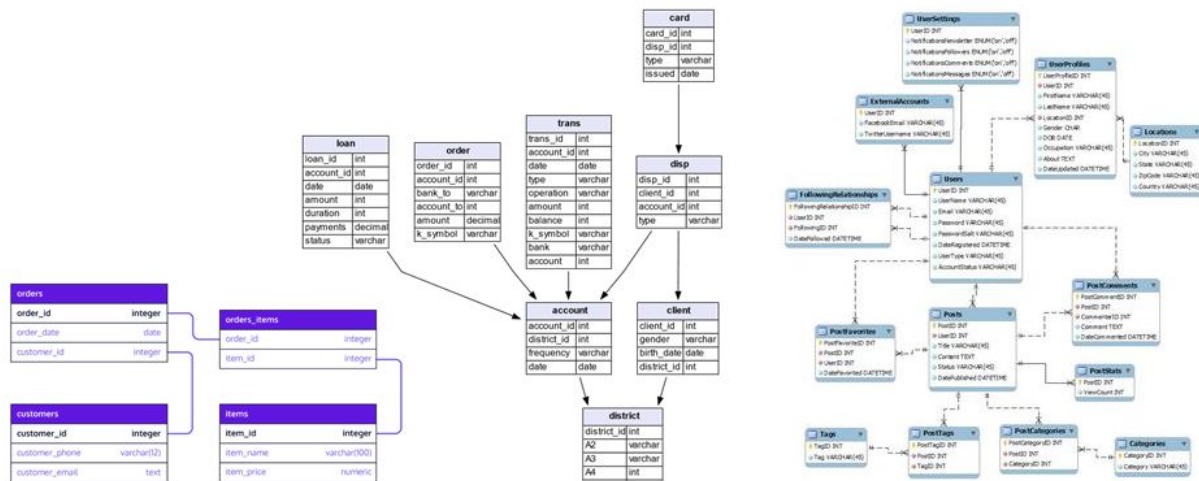
## 데이터 자산과 그래프



# Graph Good Use case - MDM(Master Data Management)

Graph 간단해요! 지금 RDB를 활용하는 것을 비유해서 살펴보자면,

## Databases are Graphs!



Commerce

Finance

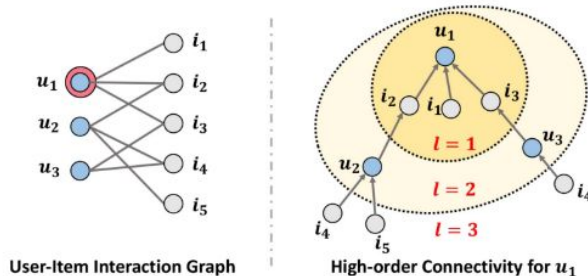
Social Media

# Graph Good Use case - Recommender System

추천 시스템에서도 활발하게 사용하고 있어요. 고객의 구매 행태와 비슷한 사람의 사람의 사람이 구매한 내역을 기반으로 추천할 때 효율적임.(JOIN)

## Motivation: Why GNN are needed for RecSys

- High-order connectivity
  - Recommender systems rely on capturing similarity
    - User-user (User-CF), item-item (Item-CF), user-item (Model-CF)
  - GNN extends similarity to high-orders
    - Connectivity among high-order neighbors
  - Besides, data sparsity issue is well addressed



Figures are from:

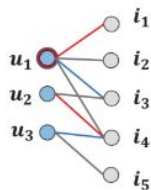
Wang et al. Neural Graph Collaborative Filtering, SIGIR 2019

# Graph Good Use case - Recommender System

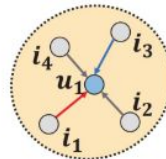
사실, 고객이 남긴 로그가 그렇게 많지않기에 고객의 로그를 기반으로 연결하고 그 연결에 기대어 추천하는 형태라고 보시는거예요.(Cold-start)

## Motivation: Why GNN are needed for RecSys

- Supervision signal
  - Users' feedback can be sparse
    - Semi-supervised signal in GNN learning
  - Users' feedback can be various
    - Well handled by various-form graph (nodes and edges)



(a) U-I Interaction Graph



(b) Local Graph of  $u_1$

Figures are from:

Jin et al. Multi-behavior Recommendation with Graph Convolutional Networks, SIGIR 2020



# Graph Good Use case - Recommender System

로그를 좀 더 다양화해서 엣지에 표현해본다면,

## Motivation: Why GNN are needed for RecSys

### ➤ Structured data

- The input of today's recommender system is always structured
  - Can be utilized to construct graph
- Learning from not only features but also structural information
  - Structural reveals implicit signals that cannot be learned by traditional works
- GNN's strong power to learn from graph-structured data

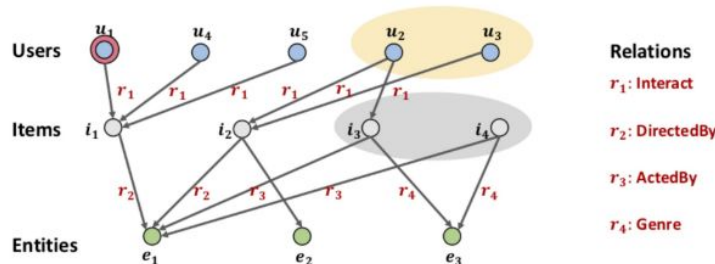
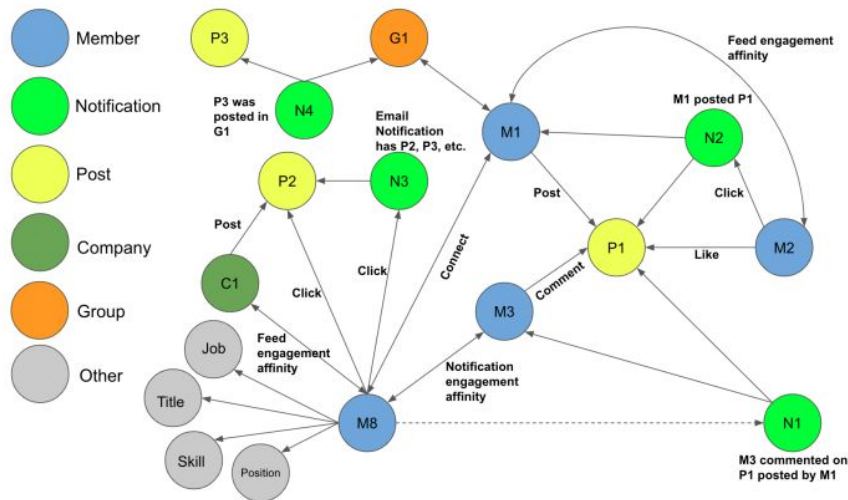


Figure from:

Wang et al. KGAT: Knowledge Graph Attention Network for Recommendation, KDD 2019

# Graph Good Use case

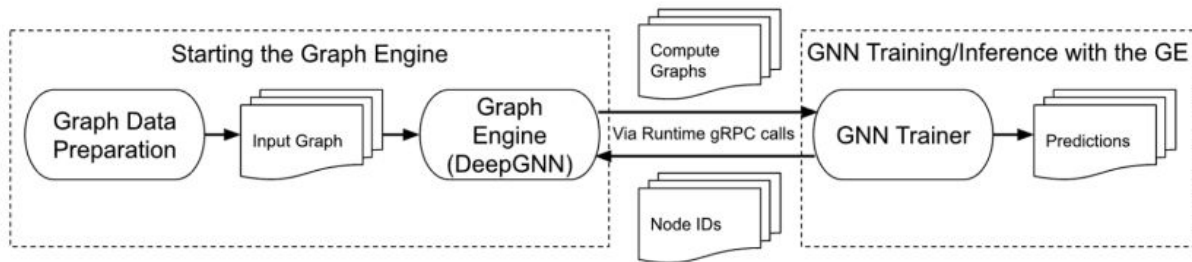
GNN을 프로젝트에 활용하려면, 다음을 참고하시는걸 강력 추천드려요! (Linkedin)



**Figure 1: Schematic representation of LinkedIn Graph. Members engaging with Posts, Jobs, Groups, Companies and other members.**

# Graph Good Use case

GNN을 프로젝트에 활용하려면, 다음을 참고하시는걸 강력 추천드려요! (Linkedin)



**Figure 2: High level view of GNN pipelines.**

# Graph Good Use case

---

GNN을 프로젝트에 활용하려면, 다음을 참고하시는걸 강력 추천드려요! (Linkedin)

## DeepGNN Overview

DeepGNN is a framework for training machine learning models on large scale graph data. DeepGNN contains all the necessary features including:

- Distributed GNN training and inferencing on both CPU and GPU.
- Custom graph neural network design.
- Online Sampling: Graph Engine (GE) will load all graph data, each training worker will call GE to get node/edge/neighbor features and labels.
- Automatic graph partitioning.
- Highly performant and scalable.

# Graph + LLM

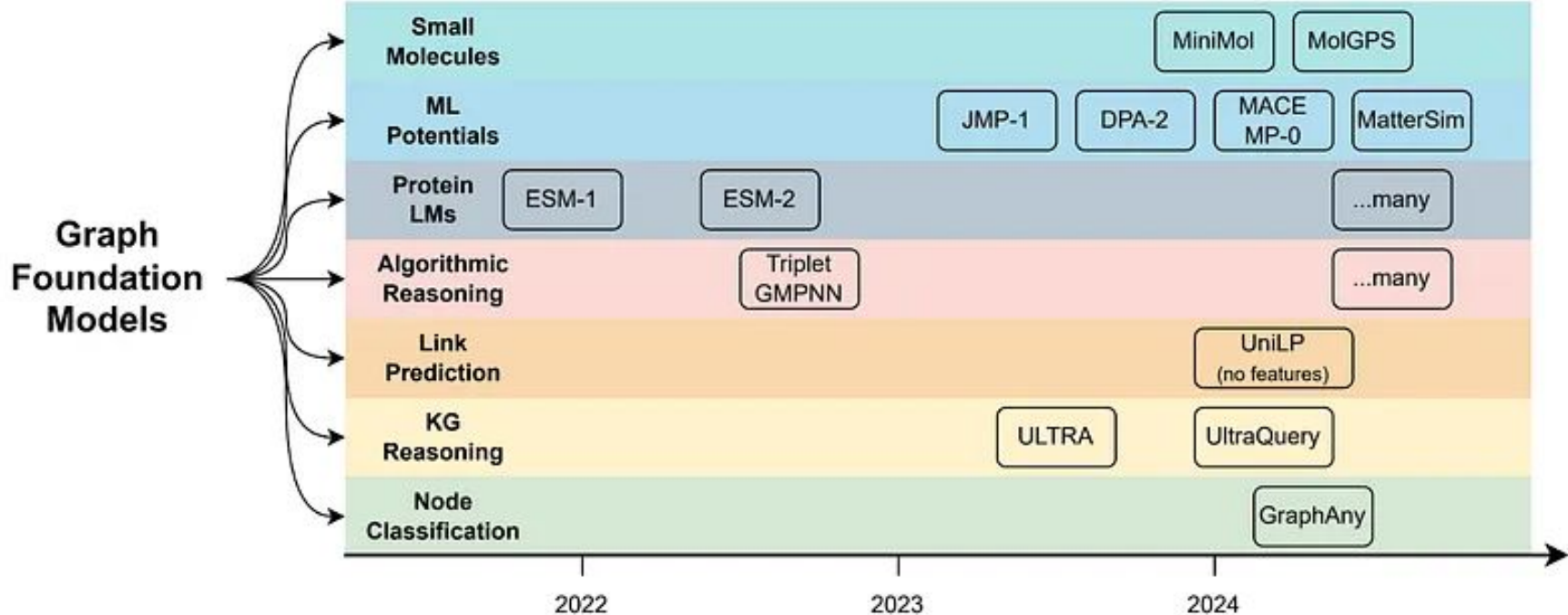
---

Graph AGI , 핵심인 Graph Foundation 이란?

“A Graph Foundation Model is a single (neural) model that learns transferable graph representations that can generalize to any new, previously unseen graph”

# Graph + LLM

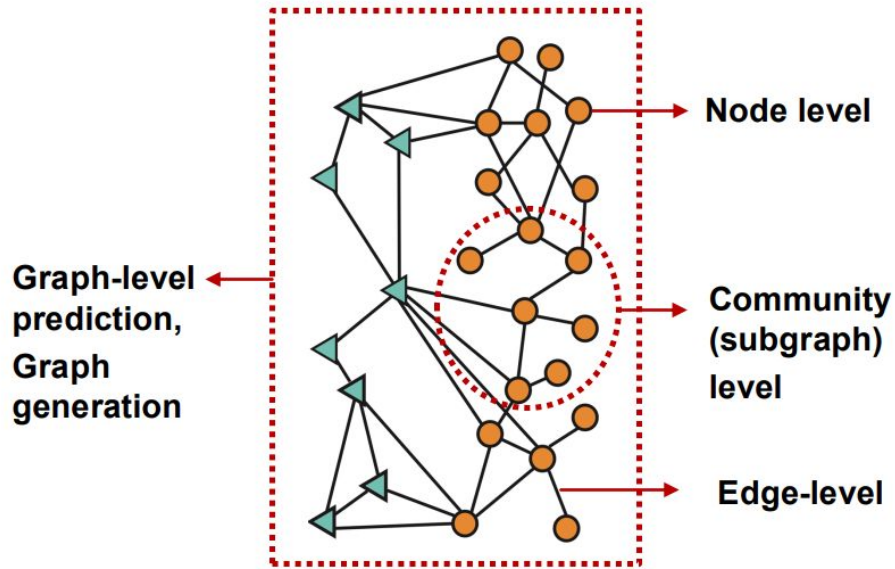
활발하게 연구되고 있는 분야 GFM



# Graph + LLM

GFM을 구축하기 위해, 수행해야 할 Task

## Different Types of Tasks



# Graph + LLM

GFM 구축을 위해 요즘 많이 쓰는 방식 Graph MoE(Mixture of Experts)

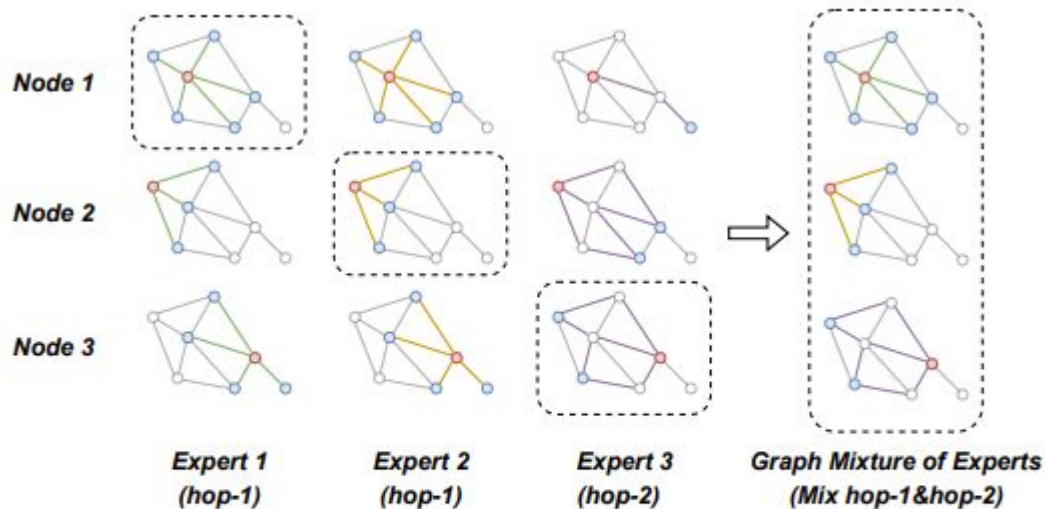


Figure 1: Each row represents the aggregation of a single node, and each column corresponds to a different network or sub-module. Blue dots (•) denote the input features passed to the red dots (•) via the colorful edges. On the left, we demonstrate two hop-1 experts with distinct weights, along with one hop-2 expert. On the right, GMoE is depicted. In this instance, the proposed GMoE selectively chooses one expert for each node while masking the others. Best viewed in color.



# Graph + LLM

잠시, 고품질의 Graph + LLM 리소스로 넘어가서 설명해볼게요.

LS-17	Multimodal Pretraining, Adaptation, and Generation for Recommendation: A Tutorial	Lecture-Style	Sunday, August 25	10:00 AM – 1:00 PM
LS-18	RAG Meets LLMs: Towards Retrieval-Augmented Large Language Models	Lecture-Style	Sunday, August 25	10:00 AM – 1:00 PM
LS-20	Recent and Upcoming Developments in Randomized Numerical Linear Algebra for Machine Learning	Lecture-Style	Sunday, August 25	10:00 AM – 1:00 PM
LS-23	Symbolic Regression: A Pathway to Interpretability Towards Automated Scientific Discovery	Lecture-Style	Sunday, August 25	10:00 AM – 1:00 PM
HO-8	Domain-Driven LLM Development: Insights into RAG and Fine-Tuning Practices	Hands-on	Sunday, August 25	2:00 PM – 5:00 PM
HO-9	Tutorial on Graph Reasoning with LLMs (GReaL)	Hands-on	Sunday, August 25	2:00 PM – 5:00 PM
LS-2	A Survey on Hypergraph Neural Networks: An In-Depth and Step-By-Step Guide	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-6	Causal Inference with Latent Variables: Recent Advances and Future Perspectives	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-11	Graph Machine Learning Meets Multi-Table Relational Data	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-14	Inference Optimization of Foundation Models on AI Accelerators	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-15	Large Language Models for Graphs: Progresses and Directions	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-19	Reasoning and Planning with Large Language Models in Code Development	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-21	Safe Multi-Modal Machine Learning	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
LS-25	Towards Urban General Intelligence Through Urban Foundation Models	Lecture-Style	Sunday, August 25	2:00 PM – 5:00 PM
HO-1	Practical Machine Learning for Streaming Data	Hands-on	Monday, August 26	10:00 AM – 1:00 PM
HO-4	Breaking Barriers: A Hands-On Tutorial on AI-Enabled Accessibility to Social Media Content	Hands-on	Monday, August 26	10:00 AM – 1:00 PM
HO-2	A Hands-on Introduction to Time Series Classification and Regression	Hands-on	Monday, August 26	2:00 PM – 5:00 PM
HO-3	DARE to Diversify: Data Driven and Diverse LLM Red Teaming	Hands-on	Monday, August 26	2:00 PM – 5:00 PM
LS-12	Grounding and Evaluation for Large Language Models: Practical Challenges and Lessons Learned	Lecture-Style	Monday, August 26	2:00 PM – 5:00 PM
HO-5	Privacy-Preserving Federated Learning using Flower Framework	Hands-on	Thursday, August 29	10:00 AM – 1:00 PM
LS-9	Foundation Models for Time Series Analysis: A Tutorial and Survey	Lecture-Style	Thursday, August 29	11:00 AM – 3:00 PM



00...

<https://github.com/HKUDS/Awesome-LLM4Graph-Papers>

<https://www2024.thewebconf.org/docs/tutorial-slides/large-language-models-for-graphs.pdf>



A16

디지털테크

# “핫플 어디지?”… 플랫폼 ‘맞춤검색’ 각축전

인스타로 트렌드 찾기 늘자  
‘ 선호 콘텐츠’ 검색기능 강화  
네이버 ‘20대 인기글’ 노출  
구글 ‘인기상승’ 검색어 도입  
통신사, 대화형 ‘AI검색’으로  
원하는 결과 원스톱 서비스

ICT 업계 신규 검색 서비스		
NAVER	네이버	패션·맛집 등 트렌드 정보 최우선 노출 테스트, 연내 AI 검색 서비스 ‘큐’ 모바일 출시 예정
G	구글	검색창에 ‘인기 급상승 검색어’ 실시간 서비스 시작, 생성형 AI 검색 기능 ‘AI 오버뷰’ 공개
SK 텔레콤	SK Telecom	미국 생성형 AI 검색엔진 ‘파플렉시티’에 100만달러 투자 및 관련 서비스 개발 본격화
Apple	애플	음성 비서 ‘시리’ 통해 오픈AI의 챗GPT 검색 기능 제공
Instagram	인스타그램	콘텐츠 창작자 지원 활성화로 검색에 활용되는 정보 확대

국내의 플랫폼 기업들이 잇달아 트 렌드에 맞춰 개인화된 추천 검색 기능 을 활성화하기 위해 공을 들이고 있 다. 정보를 찾는 목적에 따라 구글과 네이버 등 기존 검색엔진 외에도 인스 타그램 같은 소셜미디어부터 생성형 인공지능(AI)이 가미된 잿보 등 플랫 폼 간 경계를 넘나드는 이용자 성향이 두드러지면서 자사 플랫폼 안에 다 오 며 머물도록 하기 위한 소위 ‘록인 (Lock-in) 전략’의 일환이다.

25일 정보기술(IT) 업계에 따르면 네이버는 최근 일부 사용자를 대상으 로 검색 결과 상단에 ‘20대가 작성한 인기글’을 볼록 형태로 노출해 선호도 를 조사하는 테스트를 진행했는데 현 지 그 결과를 토대로 정식 서비스 출시 여부를 검토하고 있다. 가령 이용자가 ‘서울 가볼 만한 곳’ ‘강남 맛집’ 등 트 렌드에 민감한 패션·미용·식음료 매 장 등을 검색했을 때만 20~29세 블로 그 창작자들이 작성한 최신 인기 문서 (블로그 등)가 통합 검색 결과로 가장 먼저 노출되는 식이다.

이는 젊은 층을 중심으로 핫플레이 스(명소), 맛집 등 장소 관련 정보나 소문 관련 정보를 찾을 때 인스타그램 이나 유튜브, 틱톡 등을 검색 수단으 로 활용하고 있다는 점을 의식한 대응 으로 분석된다. 실제로 지난 2월 소비 자 데이터 플랫폼 오픈서베이가 전국 에 거주하는 만 15~59세 남녀 1000명 을 대상으로 설문해 진행한 결과, ‘국 급한 것을 검색하기 위해 이용하는 서 비스’(중복 응답 기준)로 네이버 (97%)에 이어 유튜브(79.9%)와 인 스타그램(38.6%)이 각각 2위와 4위 를 차지했다. 오픈AI의 챗GPT를 검 색 용도로 활용한다는 응답도 17.8% 였다. 특히 이용자 체류시간 측면에 서도 네이버(이하 5월 와이즈업·리 테일·아웃 집게 기준 3억4352만시 간)는 인스타그램(3억8994만시간) 에 두 달 연속 밀리고 있는 상태다. 네 이버에 맞서 구글은 최근 포털 검색한 을 누르며 실시간으로 ‘인기 급상승

검색어’를 보여주는 서비스를 시작하 기도 했다.

이처럼 기존 검색시장 강자로 군림 하던 포털 사업자들이 선점한 자리를 빼앗기지 않고자 이용자 구미에 맞는 다양한 검색 실험을 이어나가고 있 다면, 상대적으로 신흥 검색 엔진으로 주목받고 있는 유튜브, 인스타그램, 틱톡 등 플랫폼은 콘텐츠 창작자(크 리에이터)를 적극 공략하면서 검색에 활용되는 양질의 정보를 확보해 나가 고 있다. 사회관계망서비스(SNS) 특 성상 실시간으로 유입되는 사용자 반 응과 선호도, 유행 등이 반영된 생활 밀접형 정보가 이들 플랫폼이 지향하 는 검색의 차별화 전략이다. 한편 최 근 들어 부상하고 있는 생성형 AI 기 술이 접목된 ‘대화형’ 검색 서비스의 경우 정보의 간결성과 신속성 및 정화 성에 보다 역점을 두는 정보통신기술 (ICT) 기업들이 많은 편이다. 종전 포털 중심의 검색 구조는 사용자가

‘키워드’를 입력해 원하는 결과값을 얻을 때까지 페이지마다 여러 커뮤니 티 ‘링크’를 하나하나 들어가 보고 확 인하는 과정을 반복해야 했다. 그러나 AI 검색은 ‘질문’과 ‘답변’이라는 대 화의 과정을 통해 최적화된 검색 결과 를 일목요연하게 정리한 내용으로 제 공해 주기 때문에 사용자 편의성이 높 다는 특징이 있다.

일단 네이버는 지난해 9월 내놓은 AI 검색 서비스 ‘큐’ ‘자기 바전을 모 바일 환경에서 이미지 검색까지 가능 한 멀티모달 형태로 준비하고 있다. 상대적으로 후발 주자인 SK텔레콤은 미국의 유니콘 기업 ‘파플렉시티’에 100만달러(약 137억원)를 투자하 며 생성형 AI 검색엔진 서비스를 위 한 속도전에 나섰다. 그동안 폐쇄적인 생태계를 고수해온 애플이 전략을 바 뀌며 GPT와 결합된 음성 비서 ‘시리’ 를 시장에 공개한 것 역시 같은 맥락으 로 풀이된다.

한 업계 관계자는 “검색 엔진의 경 계가 모호해짐에 따라 이제는 포털 외 에도 SNS, 동영상 플랫폼 등을 통해 정보를 얻는 과정에서 이용자는 콘텐 츠를 소비하고 소파킹을 하며 커뮤니티 에 참여하는 복합적인 활동을 이어나 간다”면서 “그만큼 검색을 통해 수반 되는 사용자 유입 효과가 크기 때문에 플랫폼 기업들이 앞다퉀 자사 서비스 에서 차별화할 수 있는 검색 특화 기능 에 힘을 주는 것”이라고 분석했다.

고민서 기자

## AWS 기술 블로그

# GS SHOP 패션 검색의 진화, Amazon Bedrock 멀티모달 기반 패션 검색 시스템 구현 사례

by DongHoon Han | on 22 8월 2024 | in [Amazon Aurora](#), [Amazon Bedrock](#), [Generative AI](#) | [Permalink](#) | [Share](#)

목표 : 고객들에게 편의성 제공하기 !

=> 이미지와 텍스트 정보를 바탕으로 원하는 상품을 보다 쉽고 직관적으로 찾을 수 있게 되었습니다.

# Graph + LLM

아키텍처는 비교적 간단!? 요새 데이터 이관 privacy 이슈 해결을 위해 Cloud VPC가 잘 되어있나봐요.

GS SHOP | Amazon

## 시스템 구현 아키텍처



# Graph + LLM

상품 속성 추출을 위해, 테스트 데이터셋 1,000개를 무작위로 추출했고 이를 기반으로 Haiku 모델에게 정제된 데이터 전달했어요. 그리고 결과물을 기획자와 개발자가 함께 검증하고 프롬프트를 튜닝하는 방식으로 개선했다고 하네요.

GS SHOP | Amazon

## 상품 속성 파악 및 정제 과정

상품코드	구분	상품명	상품핏	상품색상	Print	Neckline	Image
1055653032	티셔츠	[블루종] 데일리 브이넥 반팔 티셔츠_B2405TS082B_H	기본핏	그레이	무지	브이넥	
1055653008	티셔츠	[블루종] 가오리 루즈핏 라운드 티셔츠_B2405TS080B_D	루즈핏	그레이	무지	라운드넥	
1055652977	티셔츠	[블루종] 쿨링 카울렉 민소매 티셔츠_B2405TS079B_H	기본핏	블루	무지	브이넥	
1055652938	티셔츠	[블루종] 도형 핫픽스 반팔 티셔츠_B2405TS078B_H	기본핏	핑크	그래픽	라운드넥	

# Graph + LLM

LLM가 상품 속성을 잘 이해할 수 있게 프롬프트 엔지니어링해주는 작업이 대다수였어요.

GS SHOP | Amazon

## [Prompt Example]

Prompt	Description
You are the top fashion designer of a clothing company. You should look at the picture ...	프롬프트 역할 지정. 상품명과 카테고리 대분류/중분류를 Context로 입력, 색상/프린트를 다중 선택 가능하다는 내용 추가
<pre>Rules. {   "Color": {     "black": "01",     "white": "02",     "beige": "03",     ...   },   "Print": {     "check": "01",     "stripe": "02",     "zigzag": "03",     ...   }, }</pre>	현업 기획자와 협의한 색상 및 프린트 목록을 JSON으로 구성하여 Context로 전달
If you are uncertain, it is fine to classify it as unknown...	색상/프린트가 명확하지 않은 경우 예외 처리 프롬프트 구성

# Graph + LLM

결과가 나오면 해당 결과를 DB에 Insert! (여기에서 DB는 Postgres를 사용했어요.)

GS SHOP | Amazon

## [Output Example]

Prompt	Description
<pre>{   "Color": "02:06",   "Print": "02" } {   "Color": "08",   "Print": "10" }</pre>	JSON 형태로 응답을 받아 Output Token 을 최소화하고 베치프로그램에서 JSON 파싱하여 DB 에 Insert



## 패션 상품 데이터 아키텍처 구현 프로세스

GS SHOP은 약 1억 개에 달하는 방대한 상품 데이터베이스를 보유하고 있습니다. 이 중에서 패션 상품 데이터를 효과적으로 분류하고 검색 시스템에 활용하기 위해 Amazon Bedrock Claude 3 Haiku 모델을 도입했고, Amazon Bedrock API 호출 최소화를 통한 비용절감을 위해 추출 속성결과를 [Amazon RDS Aurora PostgreSQL](#)에 저장하였습니다. 이 과정에서 데이터 관련 이슈를 다음과 같이 해결하였습니다.



### 1. 데이터 스키마 설계

패션 상품의 특성을 정확하게 반영할 수 있는 데이터 구조를 설계하는 것이 첫 번째 과제였습니다. 접근 방식은 다음과 같습니다.

- 상품 코드 테이블 : 다양한 패션 필터(색상, 소재, 디자인 등)를 활용하여 협업과 데이터를 직접 확인하며 Amazon Bedrock 결과물의 경우를 진행했습니다. 이를 바탕으로 GS SHOP의 패션 상품에 최적화된 결정을 정의했습니다. 이 과정을 통해 상품의 특성을 정확하게 표현하면서도 검색 효율성을 높일 수 있는 구조를 마련했습니다.
- 상품 로그 테이블 : 정의된 상품 코드를 기반으로 패션 상품 검색 데이터를 저장했습니다. 특히 실패에 따른 여러 코드를 별도로 분리하여 기록했다는 것입니다. 신규 상품에 대해서 잘못 된 상품 이미지로 인해서 Bedrock 이 정상 동작 되지 않는 경우 잘못된 상품코드를 Bedrock API를 계속 호출 할 수 있어서 호출 실패에 대한 여러 메시지를 남겼고, 여러 메시지가 남겨진 상품 코드에 대해서는 필터링을 통해 Bedrock 배치 수행 때 제외하도록 처리하였습니다. 이를 통해 불필요한 Bedrock 호출을 방지하여, 시스템 비용/성능을 최적화할 수 있었습니다.

### 2. RDB 테이블 인덱스 최적화

RDB 기반 검색 시스템의 성능은 테이블 인덱스 설계에 크게 좌우됩니다. 잘 구축된 인덱스는 데이터 검색 속도를 대폭 향상시켜 줍니다. 우리는 다음과 같은 전략으로 인덱스를 최적화했습니다.

- Aurora PostgreSQL 활용 : 데이터베이스로 Aurora PostgreSQL을 선택했습니다. 이는 온프레미스 PostgreSQL보다 가용성, 내결함성, 성능 면에서 뛰어납니다. 패션 상품 정보를 저장하는 전용 테이블을 구축하고, 상품명, 카테고리, 가격 등 주요 검색 조건 칼럼에 적절한 인덱스를 생성하여 검색 성능을 최적화했습니다.
- 성능 향상 전략 : 상품 코드 칼럼에 Primary Key 인덱스를 설정하여 데이터 중복을 근본적으로 방지했습니다. 이를 통해 데이터 무결성을 보장하고 검색 성능도 향상되었습니다. 또한 페이징 처리를 구현하여 대량의 검색 결과에도 빠르게 응답할 수 있도록 했습니다.

### 3. 데이터 이관 프로세스

대규모 데이터를 안전하고 효율적으로 이관하는 것은 중요한 과제였습니다. 데이터 이관 프로세스는 다음과 같습니다.

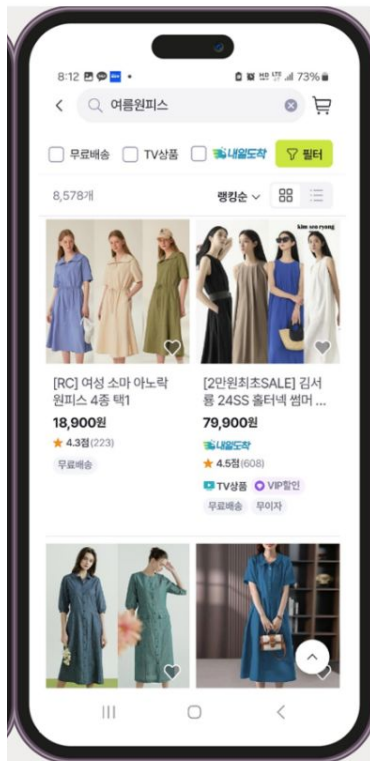
- 초기 데이터 적재 : AWS 환경에 운영 데이터베이스와 분리된 임시 데이터베이스를 구축했습니다. 이 임시 데이터베이스에 Amazon Bedrock을 활용한 병렬 처리로 초기 데이터를 적재함으로써 적재 시간을 대폭 단축했습니다.
- 임시 DB에서 운영 DB로 데이터 이관 : 초기 데이터 적재가 완료된 후, 테스트를 거쳐 검증된 임시 데이터베이스의 데이터를 운영 데이터베이스로 안전하게 이관했습니다. 이 과정에서도 운영 시스템에 대한 영향을 최소화하기 위해 비즈니스 운영에 영향이 적은 시간대를 선택하여 이관작업을 수행했습니다.
- 증분 데이터 적재 : 초기 적재 이후 추가되는 상품 데이터는 일일 배치 작업을 통해 제한적으로 적재했습니다. 이러한 방식으로 운영 중인 시스템에 미치는 영향을 최소화하면서 데이터의 최신성을 유지할 수 있었습니다.

# Graph + LLM

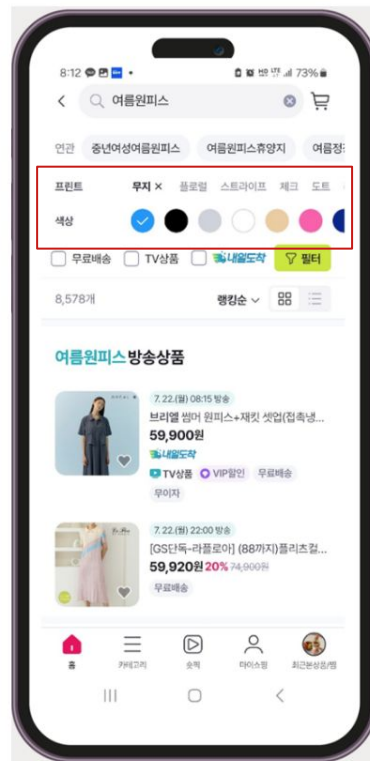
짜잔! AS-IS TO-BE

GS SHOP | Amazon

## AS-IS



## TO-BE



# Graph + LLM

잠깐 멈춰보고, 어떻게하면 적용해 볼수 있을까 고민해보아요.

Kurly마켓컬리 | 뷰티컬리

다이어트 밀키트

📍❤️🛒

카테고리

신상품

베스트

알뜰쇼핑

특가/혜택

셋별 하루 배송안내

## '다이어트 밀키트'에 대한 검색결과

필터

초기화

총 15건

추천순 | 신상품순 | 판매량순 | 혜택순 | 낮은 가격순 | 높은 가격순

카테고리

밀키트 7

메인요리 4

냉동·이식·간편채소 2

파장·볶음·파스타·면류 2

샐러드·샌드위치 1

도시락·밥류 1

배송

셋별배송 15

포장타입

상온 1

냉장 10

냉동 4

셋별배송

[도리개집] 월남쌈 밀키트

풍성한 재료에 담긴 다채로운 맛

21,200원

9,999+

Kurly Only

한정수량

셋별배송

[채선당] 월남쌈 밀키트 (2인)

재소의 신선함을 그대로

16,980원

9,999+

Kurly Only

셋별배송

[채선당] 샤브샤브 밀키트 (2인)

해장의 노하우로 완성

16,980원

9,999+

Kurly Only

Kurly마켓컬리 | 뷰티컬리

선크림

📍❤️🛒

카테고리

신상품

베스트

특가/혜택

브랜드관

셋별 하루 배송안내

## '선크림'에 대한 검색결과

필터

초기화

총 297건

추천순 | 신상품순 | 판매량순 | 혜택순 | 낮은 가격순 | 높은 가격순

카테고리

선크림 214

스킨케어 40

선크림 30

베이스 메이크업 25

메이크업 22

선크림 20

선크림 14

크림 10

선크림 7

에센스/세럼/앰플 4

카테고리 더보기 >

배송

셋별배송 297

라이브특가

셋별배송

[에스트라] 더마UV 365 장벽수분 무기자차 선크림 40ml (+무기자차 ...

피부 장벽 케어 무기자차 선크림

31,000원

28% 22,200원

34

라이브특가

셋별배송

[에스트라] 더마UV 365 레드진정 톤업 선크림 40ml (+톤업 선크림 1...

붉은기 진정 톤업 선크림

31,000원

28% 22,200원

22

15%쿠폰+응징

셋별배송

[피지오겔] DMT UV 선크림 30ML

온가족 산케어 오늘날 최대 할인특가!

36,200원

32% 10,900원

999+

# Graph + LLM

이번에는 LLM에 Graph를 추가 활용한 E-commerce 사례예요.

GS SHOP | Amazon



**Figure 1: An example of mining implicit commonsense knowledge from e-commerce user behavior.**

# Graph + LLM

유저의 특성과 의도가 최대한 반영된 검색 결과가 핵심임.

GS SHOP | Amazon

유저는 자신이  
임산부임을 '굳이'  
플랫폼에 기입하지  
않는다.

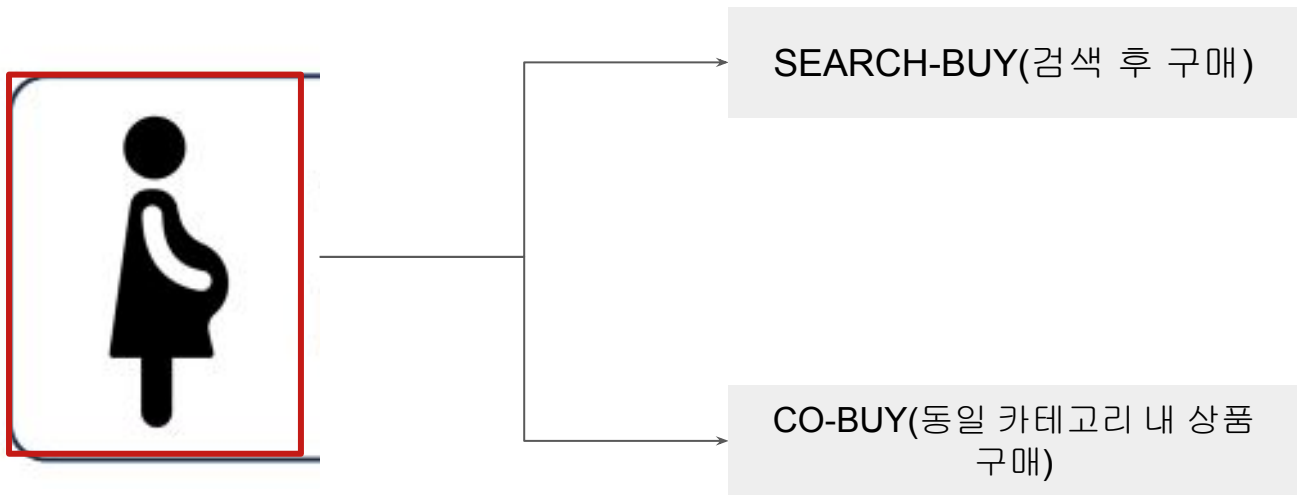


**Figure 1: An example of mining implicit commonsense knowledge from e-commerce user behavior.**

# Graph + LLM

유저 의도를 파악하기 위해 크게 두 가지로 분류했어요.

GS SHOP | Amazon



# Graph + LLM

---

지식 그래프 구축의 목적은 결국 alignment!

GS SHOP | Amazon

User behavior KG와 Item KG로 Intent를 탐지하고, 유저 선호를 생성하여 활용하기.

=> ***'We adopt instruction tuning for effective e-commerce commonsense knowledge generation to better align with human preferences.'***

Table 2: Mined e-commerce commonsense relations for the COSMO KG.

Relation Type	Tail Type	Example
USED_FOR_FUNC	Function / Usage	dry face
USED_FOR_EVE	Event / Activity	walk the dog
USED_FOR_AUD	Audience	daycare worker
CAPABLE_OF	Function / Usage	hold snacks
USED_TO	Function / Usage	build a fence
USED_AS	Concept / Product Type	smart watch
IS_A	Concept / Product Type	normal suit
USED_ON	Time / Season / Event	late winter
USED_IN_LOC	Location / Facility	bedroom
USED_IN_BODY	Body Part	sensitive skin
USED_WITH	Complementary	surface cover
USED_BY	Audience	cat owner
xINTERSTED_IN	Interest	herbal medicine
xIs_A	Audience	pregnant women
xWANT	Activity	play tennis

예시)

‘customers bought camera case and screen protector glass together because they are capable of providing protection for camera’

\*\* Yu, Changlong, et al. "COSMO: A large-scale e-commerce common sense knowledge generation and serving system at Amazon." Companion of the 2024 International Conference on Management of Data. 2024.



# Graph + LLM

검색 결과를 위해 만든 프롬프트는 대략적으로 다음과 같아요.

GS SHOP | Amazon

Task: Please provide typical explanation for the following search-purchase behavior and complete the answer.

Search Query: {Query}

Product: {Product Title}

Question: what is the product capable of, which exactly match the intention of the search query?

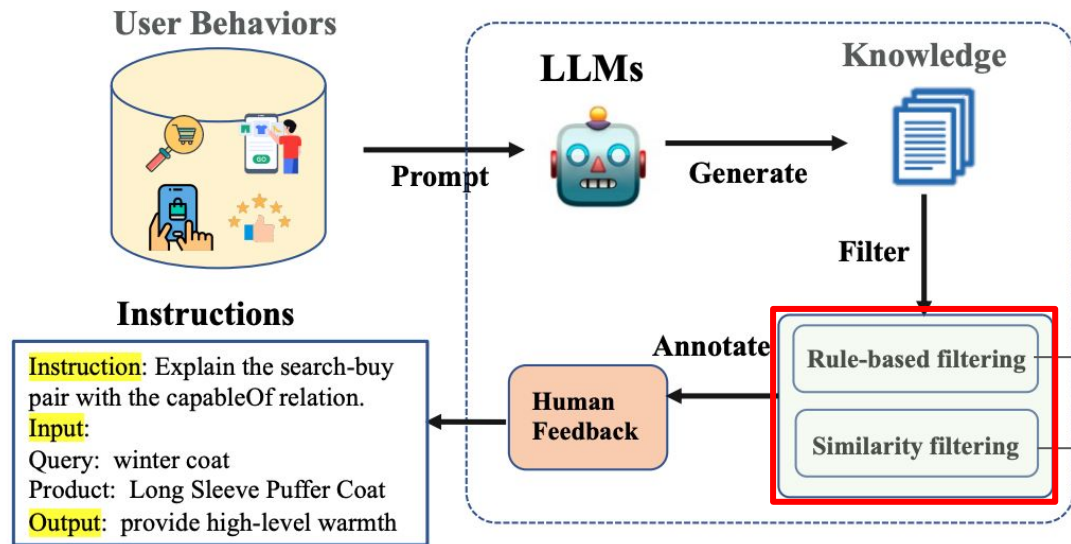
Answer: the query means customers want the product that is capable of  
1.

**Figure 3: Prompts used for generating knowledge candidates.**

# Graph + LLM

User Behavior log를 자동화 & 수동화해서 정교하게 다듬었어요. (자동화)

GS SHOP | Amazon



## Coarse-grained Filtering 이 중요하다.

**Rule-based Filtering.** We first use the sentence segmentation tool from nltk to extract the first sentence from generation. Then we calculate the *perplexity* score based on the GPT-2 language model and tune the threshold to remove incomplete sentences. We also directly filter the generations that are exactly the same as *query*, *product type* or *product title* (or edit distance less than the threshold). For the general knowledge like "used for the same reason", or "used with clothes", we identify those cases by combining frequency and entropy since they co-occur with many products or queries rather than specific ones.

**Similarity Filtering.** To handle the semantic-similar cases that can not easily be handled in the previous step, we use the in-house language model, which was pretrained on the e-commerce corpus including *query*, *product information* etc, to obtain the embeddings for generate knowledge tails, query and product themselves. The similarity between the knowledge embedding and the context embedding (the original query or product embedding) is computed by their cosine similarity:

$$d(k, c) = \cos(E(k), E(c)). \quad (1)$$

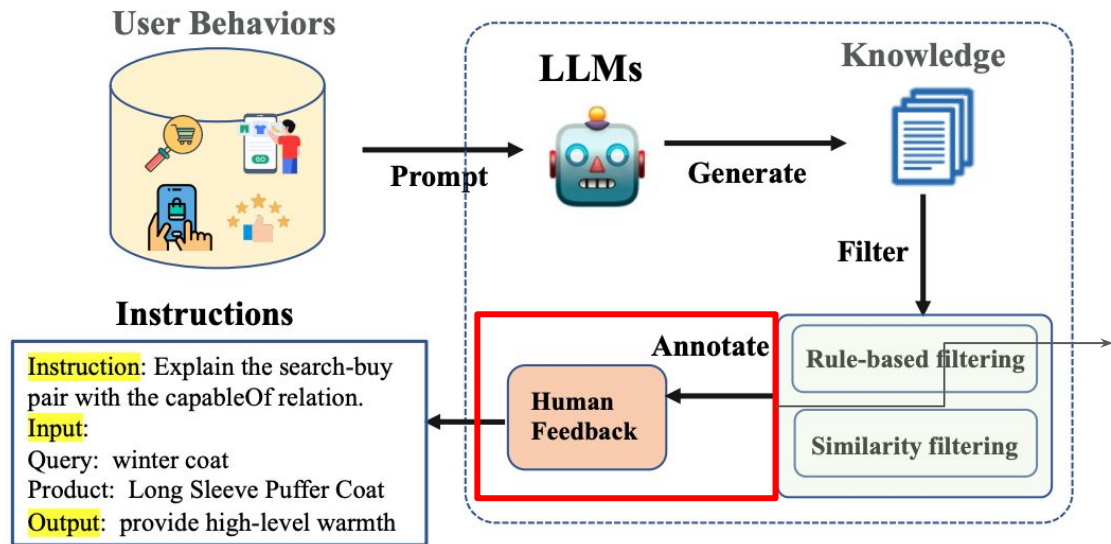
We find that filtered generations are essentially paraphrases of original user behavior contexts with syntactic transformations. By two coarse-grained filtering steps, we are able to remove quite a large amount of noise and keep typical knowledge as much as possible.

**Figure 2: Overall framework of generating high-quality instruction data from massive user behaviors and LLMs.**

# Graph + LLM

User Behavior log를 자동화 & 수동화해서 정교하게 다듬었어요. (수동화)

GS SHOP | Amazon



professional data annotation vendor company 에게 의뢰해 다음 5가지 Checklist 를 기반으로 판단.

1. Is the explanation a complete sentence?
2. Is the explanation relevant?
3. Is the explanation informative?
4. Is the explanation plausible?
5. Is the explanation typical?,

**Figure 2: Overall framework of generating high-quality instruction data from massive user behaviors and LLMs.**

# Graph + LLM

라벨링 일관성 및 편의를 위해 UI까지 만들정도!

GS SHOP | Amazon

The screenshot displays a web browser window titled 'Co-Purchase' with a URL bar showing 'website.com'. The page layout includes a hamburger menu icon on the left and a user profile icon on the right. Two product cards are shown side-by-side:

- PRODUCT A**: Fintie Microsoft Surface 3 Case - Premium PU Leather Folio Stand Cover with Stylus Holder for Microsoft Surface 3 10.8-inch (Digital Device Accessory)
- PRODUCT B**: Mr Shield for Microsoft Surface 3 (10.8 inch 2015 Version) Anti-Glare [Matte] Screen Protector (Screen Protector)

Below the products, a text box states: 'The customer bought Product A and Product B because...'. A blue box contains the generated explanation: 'They are capable of protecting his Surface Pro.'

Five evaluation questions are presented in a grid:

- QUESTION 1**: Is the explanation a complete sentence? (Radio buttons: Yes, No, Not sure)
- QUESTION 2**: Is the explanation relevant? (Radio buttons: Yes, No, Not sure)
- QUESTION 3**: Is the explanation plausible? (Radio buttons: Yes, No, Not sure)
- QUESTION 4**: Is the explanation informative? (Radio buttons: Yes, No, Not sure)
- QUESTION 5**: Is the explanation typical? (Radio buttons: Yes, No, Not sure)

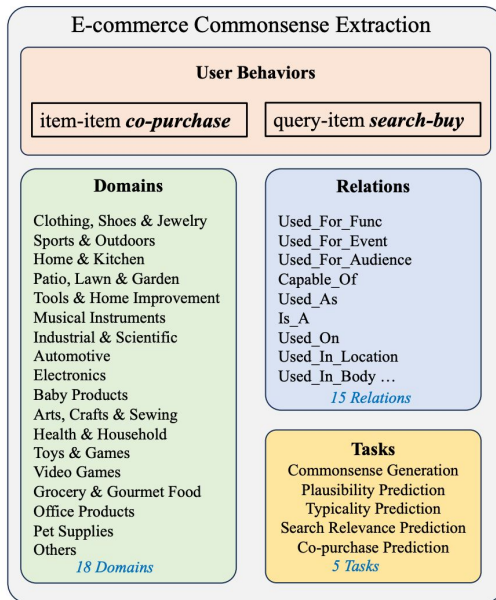
At the bottom right, there is a section labeled 'COMMENTS (OPTIONAL)' with a text area and a right-pointing arrow.

**Figure 11: Screenshot of data annotation interface.**

# Graph + LLM

구축하기 까다롭지만, 한 번 구축해 놓으면 응용할 분야가 산더미!

GS SHOP | Amazon

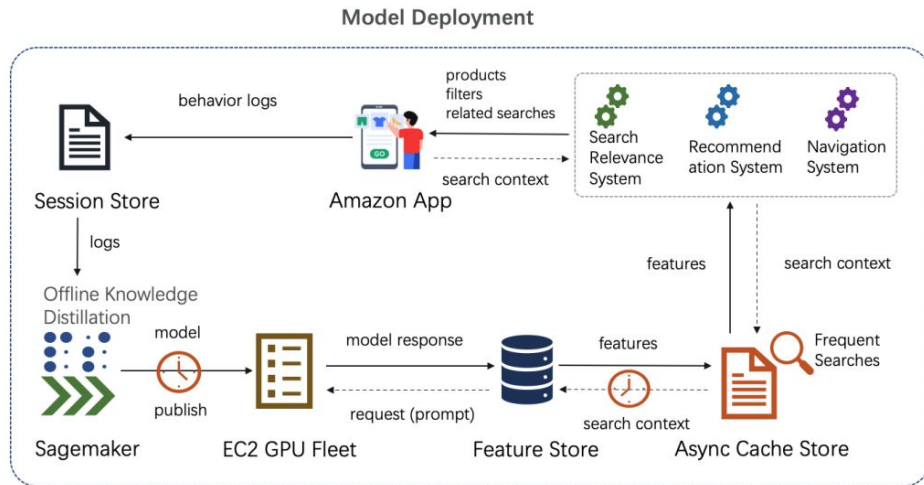


**Figure 4: Illustration of finetuning COSMO-LM to generate e-commonsense knowledge from two typical user behaviors. We scale up product domains, relation types and tasks.**

# Graph + LLM

Online(실시간) 배포까지!

GS SHOP | Amazon



**Figure 5: Illustration of COSMO-LM deployment, featuring the Asynchronous Cache Store and Feature Store as central components. It depicts the efficient processing of user queries and dynamic daily updates, crucial for meeting Amazon's search latency requirements.**

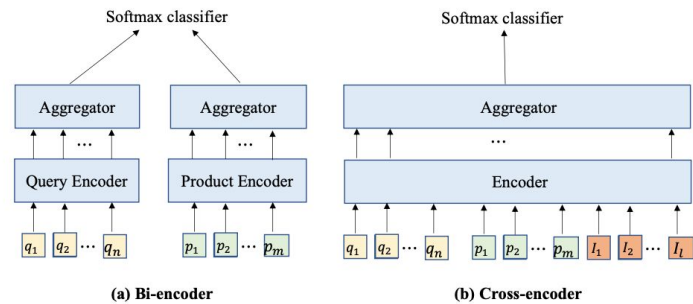


Figure 6: Illustration of Search Relevance Models.

Table 6: Experimental Result of Public ESCI English subset.

Method	Fixed Encoder		Trainable Encoder	
	Macro F1	Micro F1	Macro F1	Micro F1
Bi-encoder	25.52	65.49	47.96	70.23
Cross-encoder [49]	28.44	66.84	57.49	74.23
Cross-encoder w/ Intent	<b>45.52</b>	<b>86.40</b>	<b>73.48</b>	<b>90.78</b>
$\Delta$	60.06%	29.26%	27.81%	22.30%

\*\* Yu, Changlong, et al. "COSMO: A large-scale e-commerce common sense knowledge generation and serving system at Amazon." Companion of the 2024 International Conference on Management of Data. 2024.

Table 8: Experimental Results of Session-based Recommendations.

Method	clothing			electronics		
	Hits@10	NDCG@10	MRR@10	Hits@10	NDCG@10	MRR@10
FPMC	62.16	45.07	39.60	21.79	16.01	14.18
GRU4Rec	83.20	63.37	56.94	49.53	33.99	29.06
STAMP	81.34	61.32	54.86	56.96	38.74	32.92
CSRM	82.31	65.59	60.25	61.66	46.63	41.83
SRGNN	85.82	69.68	64.45	67.83	55.23	51.22
GC-SAN	84.43	68.96	63.93	66.88	55.87	<b>52.34</b>
GCE-GNN	86.67	69.35	63.79	70.13	55.17	50.37
COSMO-GNN	<b>90.18</b>	<b>72.30</b>	<b>67.08</b>	<b>74.21</b>	<b>56.26</b>	50.67
$\Delta$	4.05%	3.76%	4.08%	5.82%	0.70%	-3.19%

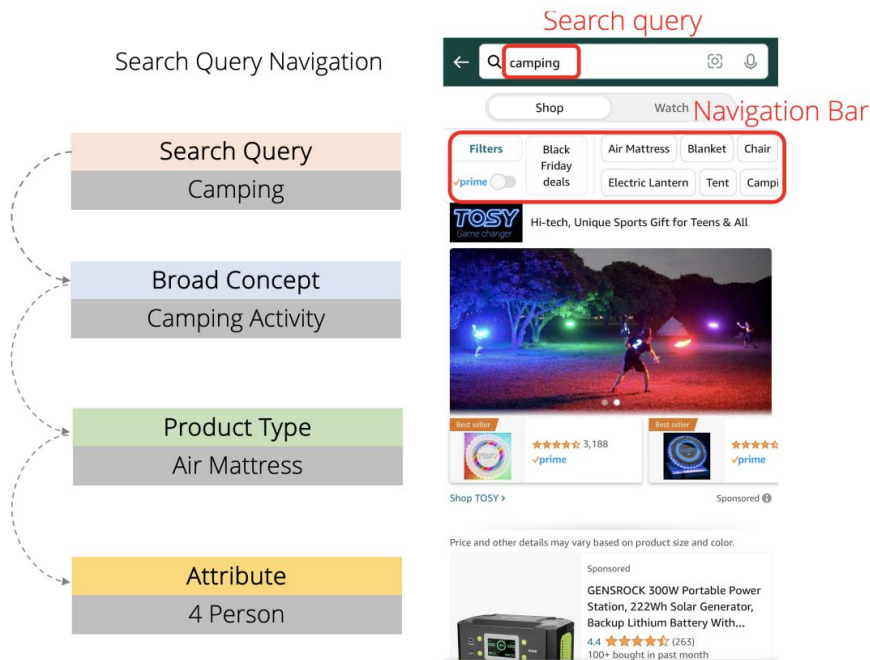
\*\* Yu, Changlong, et al. "COSMO: A large-scale e-commerce common sense knowledge generation and serving system at Amazon." Companion of the 2024 International Conference on Management of Data. 2024.



# Graph + LLM

검색 엔진에서 한 단계 더 나아가, 유저 체류시간을 늘리기 위해  
관심있을만한 상품들까지 접근하게끔 Navigation Bar 까지 개선

GS SHOP | Amazon



**Figure 9: Search navigation experience using COSMO**

# Graph + LLM

---

비즈니스 관점에서 ROI 성공적?! - 1.traffic 증가 , 2.revenue(product sale) 증가 , 3.navigation engagement 증가

GS SHOP | Amazon

This conclusion is drawn from meticulously conducted Amazon **online A/B tests carried out over several months in total, targeting approximately 10% of Amazon's U.S. traffic.**

These well-structured tests revealed **a notable 0.7% relative increase in product sales within this segment, translating to hundreds of million dollars in annual revenue surge.**

**an 8% increase in navigation engagement rate was observed within the same traffic segment,** highlighting improved customer interaction and satisfaction

The logo consists of a solid purple square. Inside the square, the word "Kurlly" is written in a white, elegant, cursive script font.

*Kurlly*

감사합니다  
여러분