
Enterprise Knowledge Graph

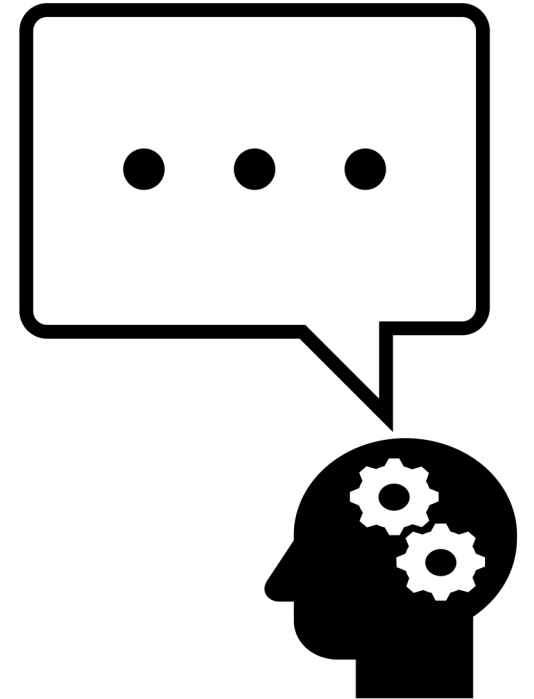
허흥수



데이터?

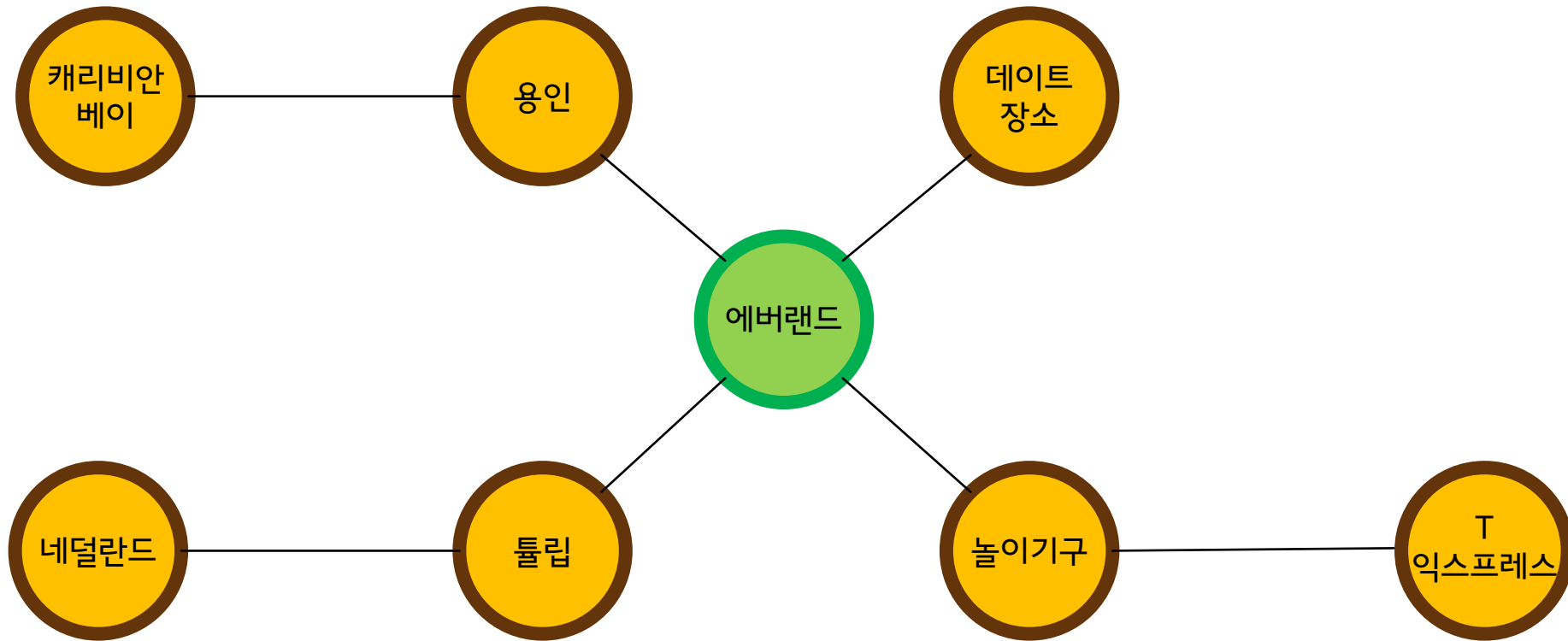


사진 출처 : <http://www.traveli.co.kr/festival/contentsView/3257/1>



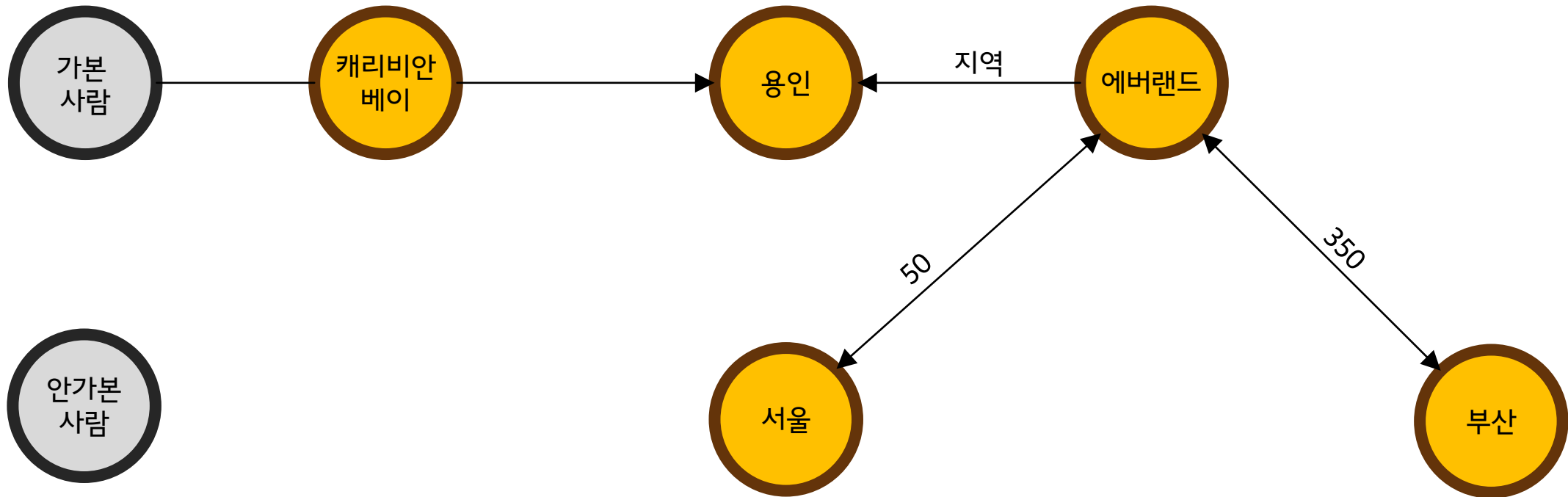
실세계 연상 표현

- 그래프는 상호 연결된 실세계의 사물들을 표현하는 자연스러운 모델



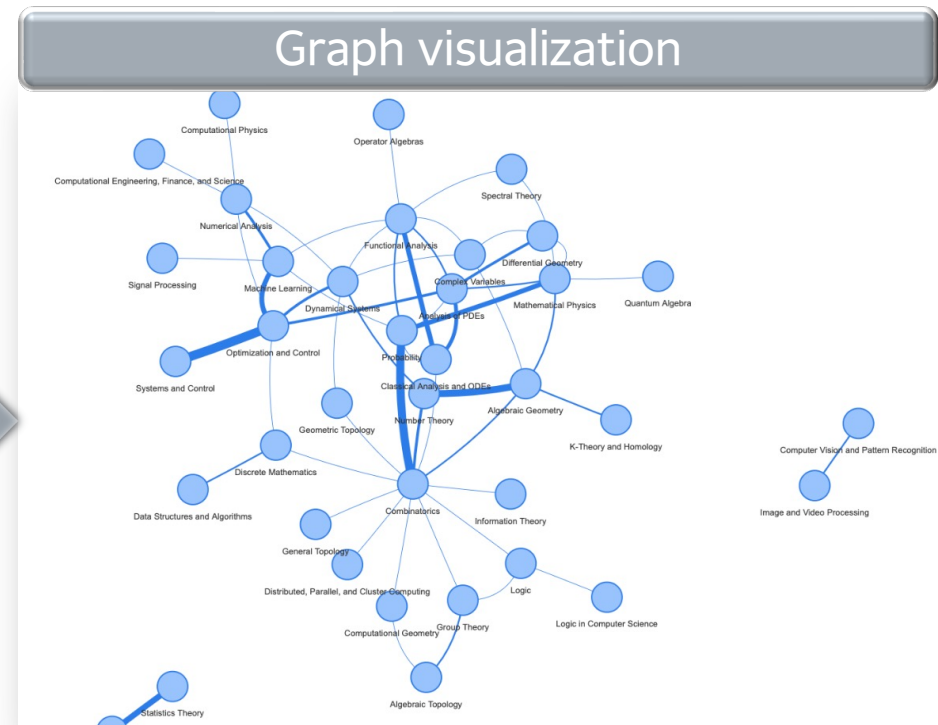
실세계 데이터와 그래프

- 실세계 데이터는 상호 관계를 맺고 있으며, 그래프는 그들이 어떻게 연결되어 있는지 모델링
- 그래프 구분
 - 방향/무방향, 연결/비연결, 가중/비가중, 레이블/비레이블



그래프 표현의 장점

| Table representation | | | |
|----------------------|-----------------------------|--------------------------|--------|
| | label1 | label2 | weight |
| 1 | Statistics Theory | Methodology | 7 |
| 2 | Numerical Analysis | Dynamical Systems | 2 |
| 3 | Numerical Analysis | Computational Physics | 2 |
| 4 | Numerical Analysis | Optimization and Control | 2 |
| 5 | Numerical Analysis | Machine Learning | 4 |
| ... | ... | ... | ... |
| 108 | Number Theory | Algebraic Geometry | 8 |
| 109 | Classical Analysis and ODEs | Combinatorics | 2 |
| 110 | Classical Analysis and ODEs | Probability | 2 |
| 111 | Classical Analysis and ODEs | Functional Analysis | 6 |
| 112 | Classical Analysis and ODEs | Analysis of PDEs | 5 |



관계 추적의 용이성

- 인간의 단기 기억 한계를 보완
- 대량의 연결을 추적하며 파악 가능

시각적 민감성 활용

- 관계에 따라 표현되는 시각적 효과 활용
- 클러스터링, 굵기, 길이 등을 통한 주요 부분 알림

간접적 관계 파악

- 직접적인 관계가 없어도 연결된 관계를 통해 유사성/근접성 파악

통찰력 제공

- 데이터 상으로 파악이 필요한 부분을 빠르게 조명
- 추가적인 조사와 탐색에 도움

그래프 모델

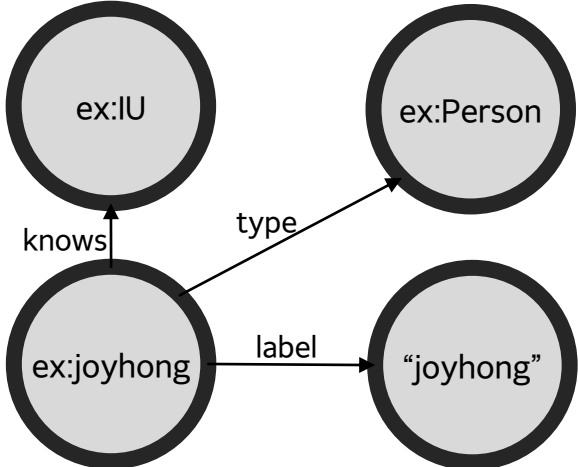
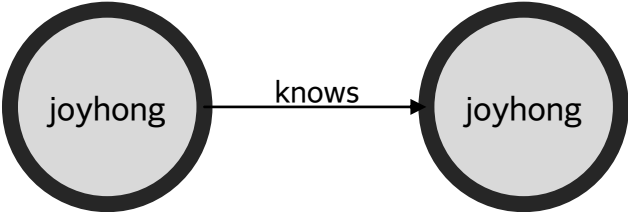
- 두가지 주요 그래프 모델

| RDF | 구분 | Property Graph |
|--|--------------|---|
| W3C의 RDF와 관련 표준을 기반으로 한 그래프 | 개요 | 아파치 ThinkerPop과 Gremlin 에 기반을 둔 노드와 엣지에 속성을 추가한 그래프 |
| 데이터 공유와 연결을 위한 자원 기술 | 설계 목적 | 분석 목적의 그래프 표현 |
| 추상화되고 유연한 스키마 구조로 구조화된 분석용 데이터 제공 | 활용 강점 | 그래프 순회 다양한 그래프 분석과 라이브러리 |
| 글로벌 식별자, 표준을 통한 상호운용성 그래프에 대한 스키마, 의미적 추론 | 데이터 관리 강점 | 그래프 데이터를 간결히 표현 짧은 러닝 커브 |
| Data-driven 아키텍처 지식 표현, 데이터 통합, 메타데이터 관리 | 주 활용사례 | 그래프 분석 경로 탐색 |
| SPARQL | 질의 언어 | Cypher, GQL, GSQL, PGQL |

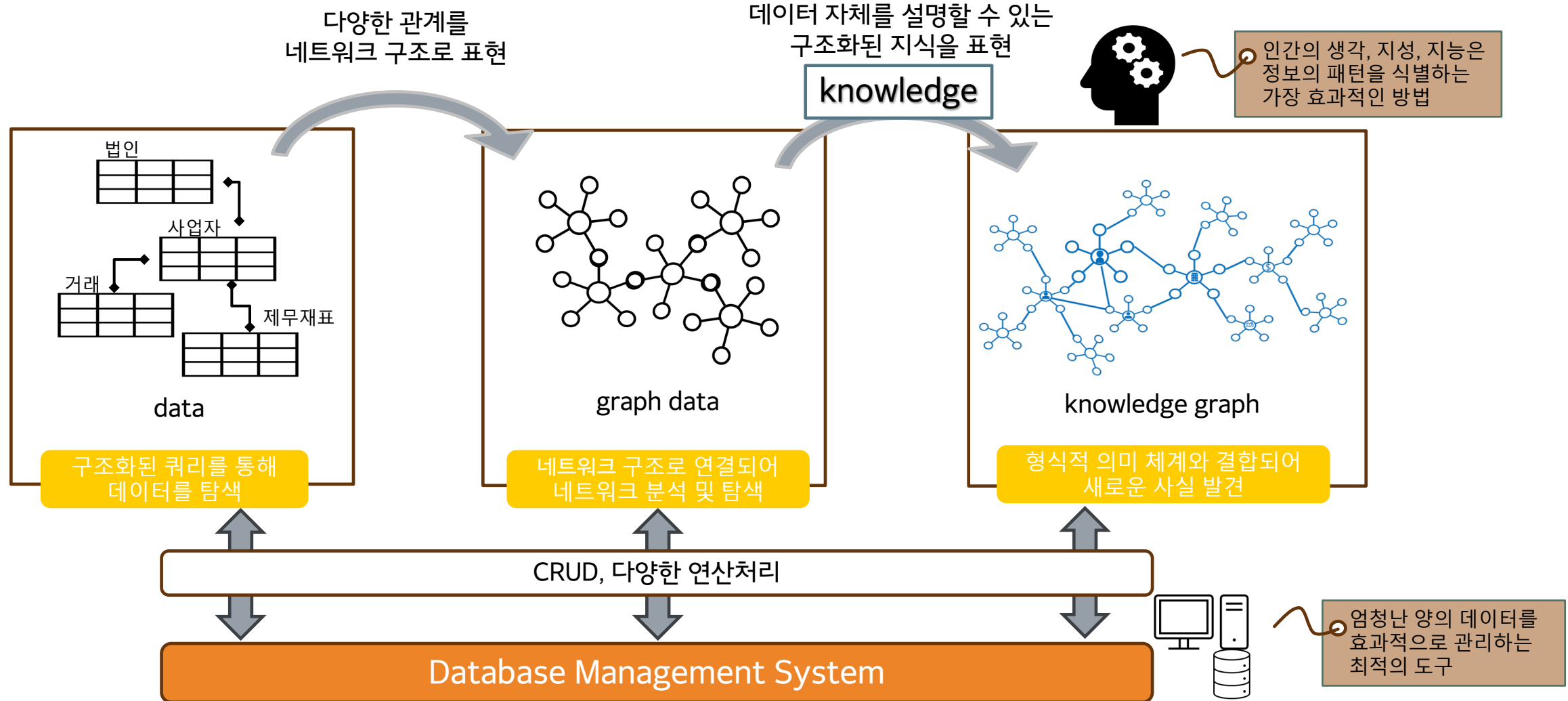
- ThinkerPop은 그래프 데이터베이스, 그래프 분석 시스템 둘 다를 지원하는 그래프 컴퓨팅 프레임워크
- Gremlin은 아파치 ThinkerPop의 그래프 순회(traversal) 언어

그래프 모델

■ 두가지 주요 그래프 모델의 차이점

| RDF | Property Graph |
|--|---|
| <p>생성 쿼리</p> <pre>PREFIX ex: <http://joyhong.tistory.com/> INSERT DATA { ex:joyhong rdf:type ex:Person . ex:joyhong rdfs:label "joyhong" . ex:joyhong rdfs:label "joyhong" . ex:joyhong ex:knows ex:IU . }</pre> | <p>생성 쿼리</p> <pre>CREATE (n1:Person{name: "joyhong"}) CREATE (n2:Person{name: "joyhong"}) CREATE (n1)-[rel:knows]->(n2) RETURN n1, n2</pre> |
| <p>표현 결과</p>  <pre>graph TD exIUIU((ex:IU)) exPerson((ex:Person)) exJoyhong((ex:joyhong)) joyhongLabel(("joyhong")) exJoyhong -- knows --> exIUIU exJoyhong -- type --> exPerson exJoyhong -- label --> joyhongLabel</pre> <p>ex:joyhong rdf:type ex:Person . ex:joyhong rdfs:label "joyhong" . ex:joyhong ex:knows ex:IU .</p> | <p>표현 결과</p>  <pre>graph LR joyhong1((joyhong)) -- knows --> joyhong2((joyhong))</pre> <pre>[{ "identity": 1, "labels": ["Person"], "properties": { "name": "joyhong" } }, { "identity": 2, "labels": ["Person"], "properties": { "name": "joyhong" } }]</pre> |

데이터, 그래프, 지식그래프

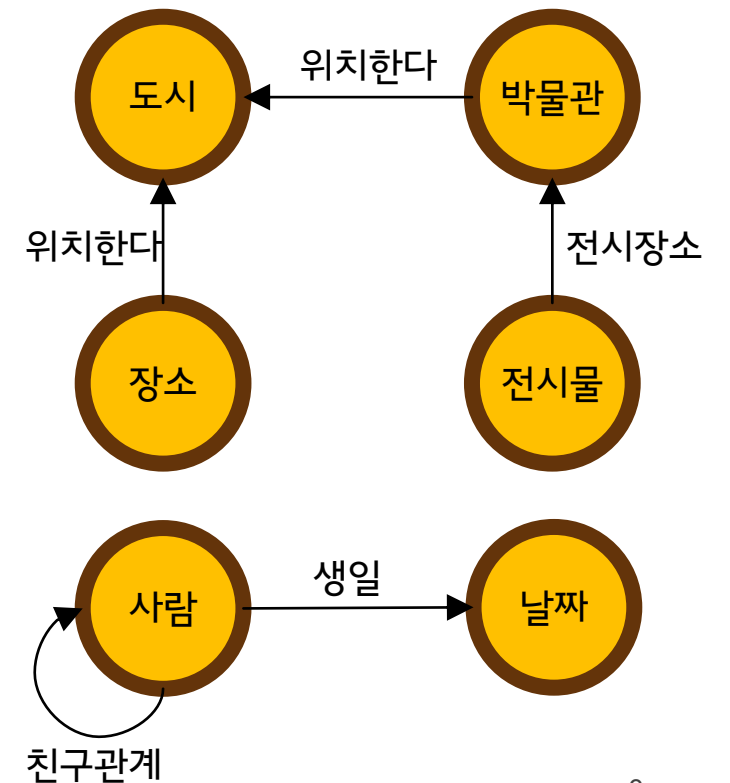
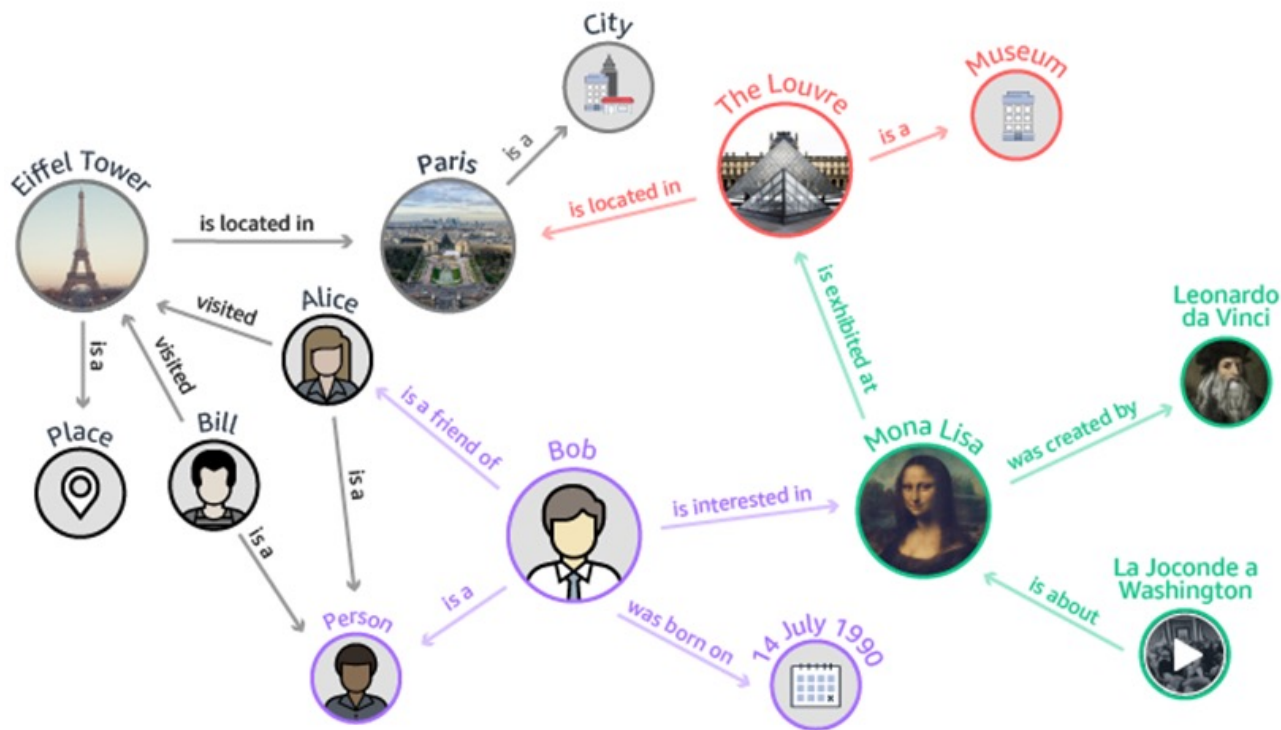


지식

- 지식이란?
 - 정보를 보다 체계화하고 개념화 한 것
 - 데이터에 대한 주요 내용, 데이터 상호간에 어떻게 관련되는지 와 같은 배경정보를 제공
- 지식의 표현
 - 정보로부터 지식을 획득하고, 다시 이 지식을 데이터로 표현 (메타데이터)
 - 마크업 언어로 표현
 - 메타데이터 자체가 또 다른 하나의 데이터

지식그래프

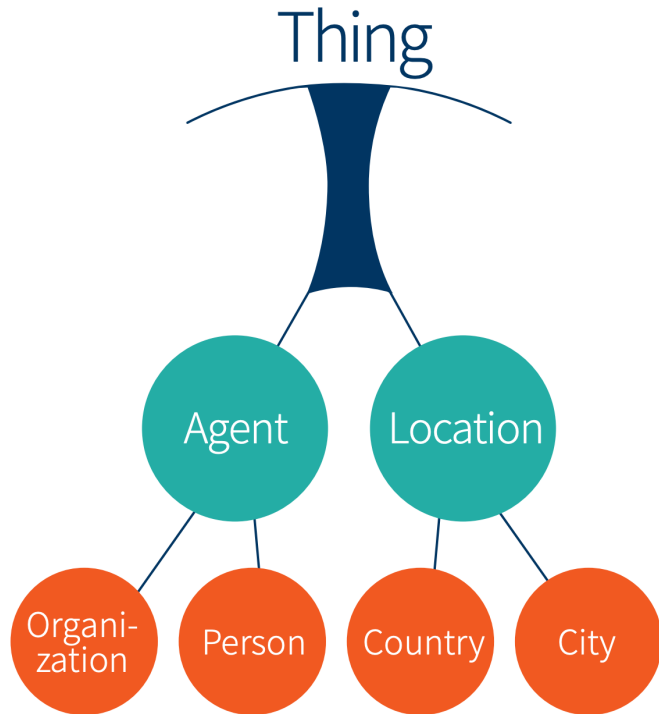
- 그래프 데이터에 메타데이터(지식)을 표현한 데이터
- 데이터를 이루고 있는 다양한 개체의 정의, 속성, 그리고 개체들 사이의 관계를 그래프 형태로 표현



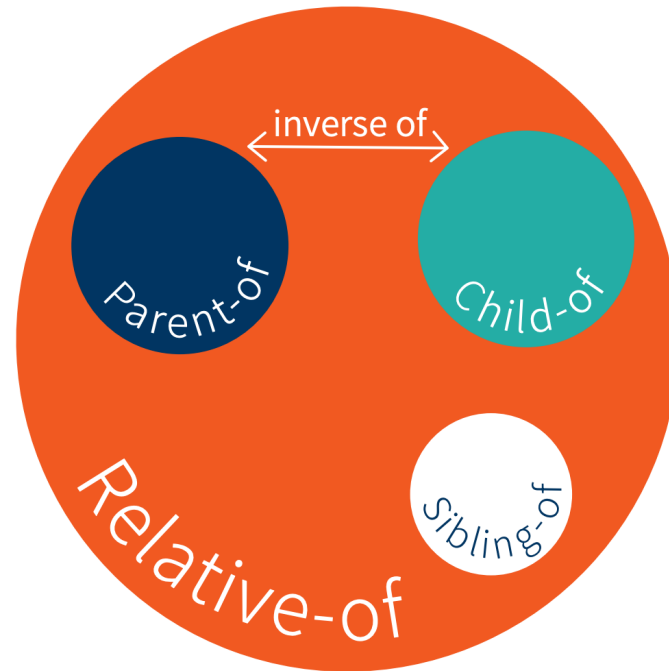
지식그래프 표현 요소

- 지식 그래프의 데이터를 표현하는 방법들

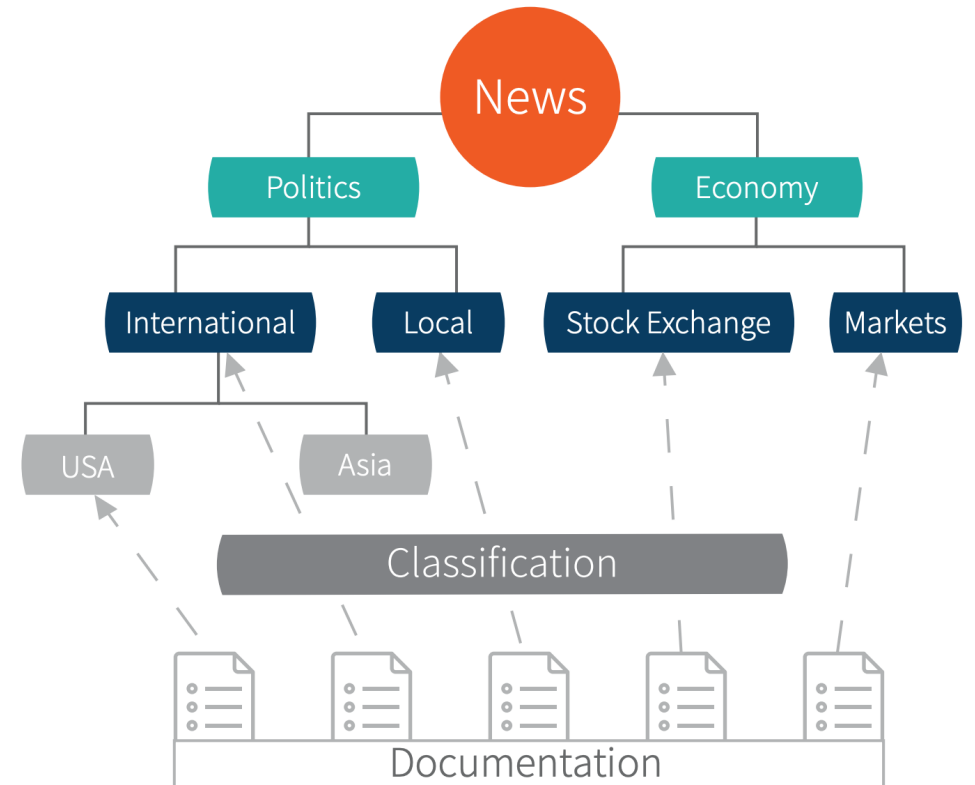
클래스



관계 유형

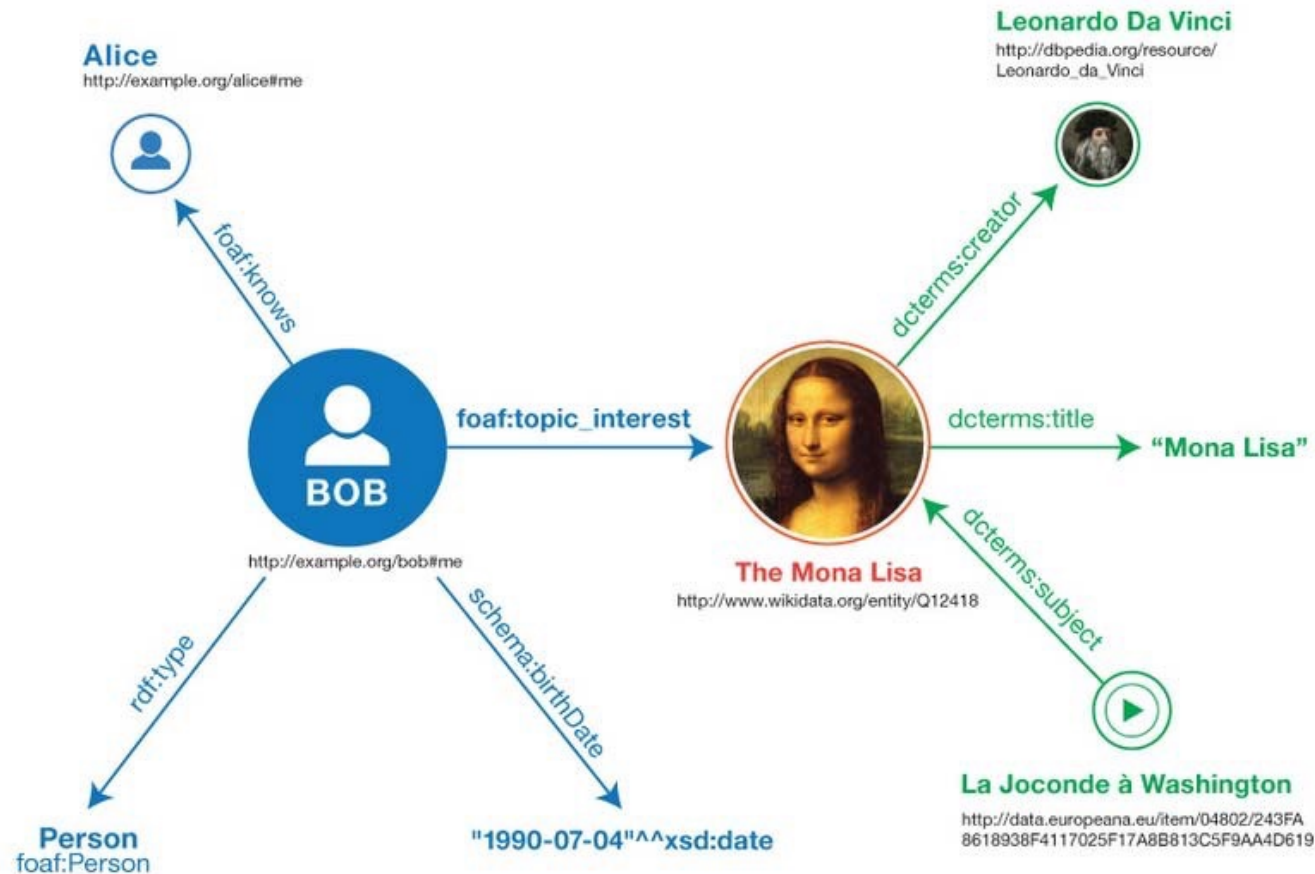


카테고리



지식그래프 표현 언어 - RDF

- RDF는 Resource Description Framework의 약자로 W3C(World Wide Web Consortium)에서 개발 및 표준화한 웹 자원 및 데이터 교환을 기술하는 표준

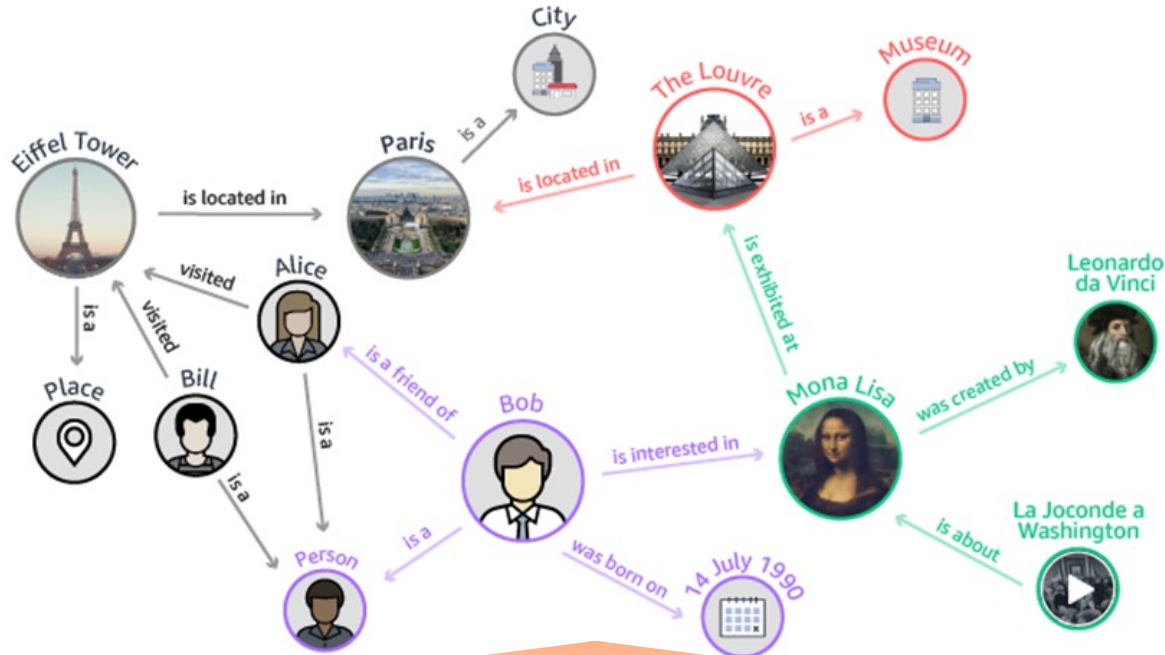


지식그래프 구성 특징

지식의 명시적이고 공유된 표현

데이터 통합

- 그래프 모델의 구조적인 특성으로 인해 다양한 출처의 데이터를 통합할 수 있음
- 실제 사물들이 다른 다양한 사물들과 가지는 연결 관계를 개념적으로 표현
- 용어에 대한 목록 생성 및 용어의 의미 또는 정의를 정리한 집합 생성



데이터 추론(지식발견)

- 메타데이터(지식)의 명시적인 표현으로 인해 새로운 관계 또는 정보를 발견
- 개념의 정의, 속성, 개념 간의 관계, 논리적 제약 조건을 통해 의미 체계를 나타냄
- 모델에 내장된 논리를 통해 모델에 포함된 정보를 이해하는데 도움



지식그래프의 장점

단절된 데이터 소스를 결합

- 서로 다른 데이터 사일로를 결합하는데 도움 제공
- 부서 또는 조직 전반에 걸쳐 지식의 전반적인 개요 제공

정형과 비정형 데이터 결합

- 다양한 유형의 데이터를 의미 있는 방식으로 연결 가능
- 연결된 정보 안에서 패턴을 추출하고 발견하는데 사용

정보에 입각한 의사 결정 지원

- 데이터들이 연결되고 어떤 관계를 가지고 있는지를 표현하는 모델을 제공하기 때문에 큰 그림을 조망하는데 용이함
- 관련 없는 정보 검색 결과가 아니라 질문에 대한 관련 사실과 문맥화된 답변 제공 가능

계층적 / 연결 흐름에 따른 통찰력 제공

- 계층적 데이터 또는 연결 흐름에 따른 데이터로부터 통찰력을 얻기에 용이
- 더 많은 데이터가 축적될수록 통찰력이 높아지게 되며, 이와 더불어 개별 객체를 식별하고, 그들 사이의 관계를 이해하며, 이를 다른 통합 데이터 세트와 비교 가능

직관적 표현과 시각화

- 비즈니스 프로세스의 정보흐름을 시각화하고, 군집화를 통해 정보를 제공
- 그래프 데이터베이스는 전통적인 방법이 할 수 없는 방식으로 연결된 데이터를 통해 보다 더 간단하고 직관적인 표현이 가능

지식그래프 구분

- 오픈 지식그래프 (Open Knowledge Graph)
 - 웹에 게시된 개방형 지식그래프
 - 많은 개방형 지식그래프가 Linked Open Data 형태로 출판
 - DBpedia, YAGO, Freebase, Wikidata
 - 대부분 Linked Data 원칙에 따라 게시되는 RDF로 모델링을 하고 SPARQL을 통해 접근을 제공
 - 도메인별 오픈 지식그래프
 - 뉴스, 텔레비전, 라디오 관련된 미디어 - BBC
 - 데이터 공개와 관련된 정부 - 미국과 영국 정부
 - 학술 문헌과 관련된 간행물 - OpenCitations, SciGraph, Microsoft 학술 지식그래프
 - 관심 장소 및 지역과 관련된 지리분야 - LinkedGeoData
 - 단백질, 유전자, 약물, 질병 관련된 생명과학 - Bio2RDF
 - 문화유산, 음악, 법, 신학, 관광업

지식그래프 구분

- 엔터프라이즈 지식그래프 (Enterprise Knowledge Graph)
 - 광범위한 응용 프로그램을 위해 다양한 산업 분야의 회사에서 만드는 지식그래프

다양한 목표들

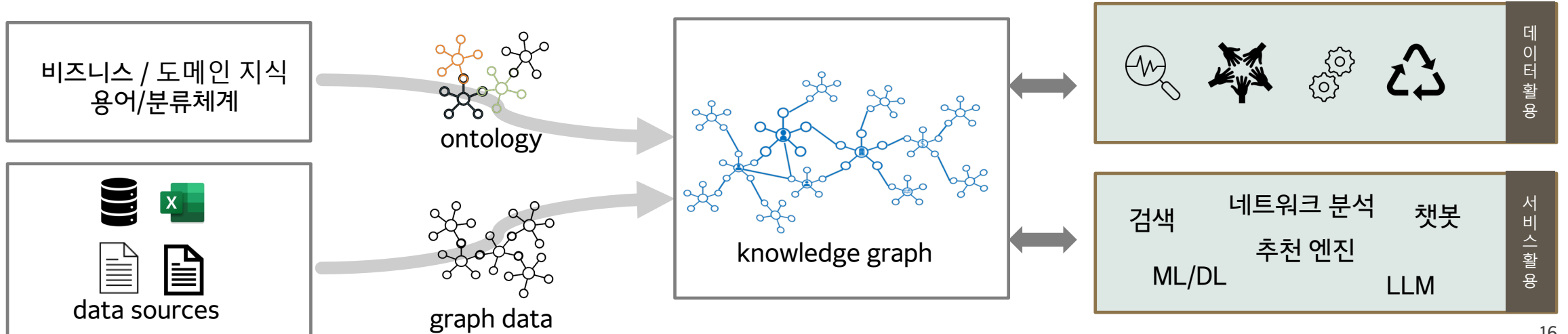
- 검색 기능 개선
- 사용자 추천 제공
- 대화형/개인 에이전트 구현
- 타켓 광고 강화
- 비즈니스 분석 강화
- 사용자 연결

유사한 점들

- 데이터는 일반적으로 다양한 내부/외부로부터 지식그래프 통합
- 엔터프라이즈 지식그래프는 데이터의 양이 커서 확장성 측면에서 문제를 제기
- 초기 지식그래프의 개선은 품질 개선을 중요하게 생각
- 도메인과 함께 지식그래프를 최신 상태로 유지하는 기술이 중요
- 온톨로지 및 기계 학습 표현의 결합이 사용되는 경우가 많음
- 사용되는 온톨로지는 가벼운 경향이 있으며, 단순한 분류법의 계층 구조를 가짐

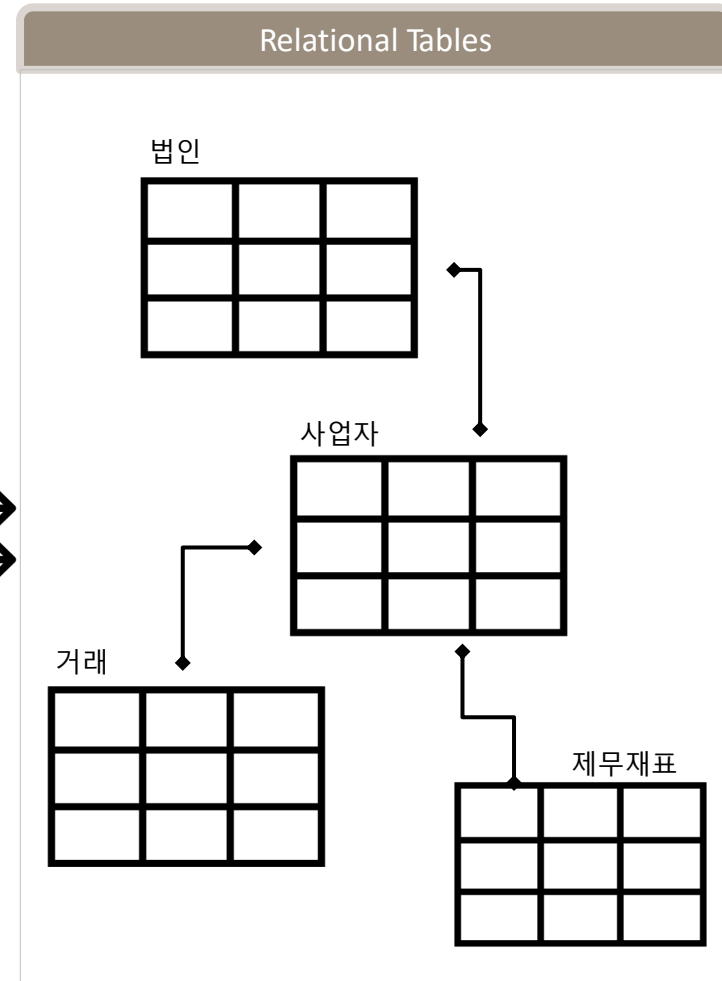
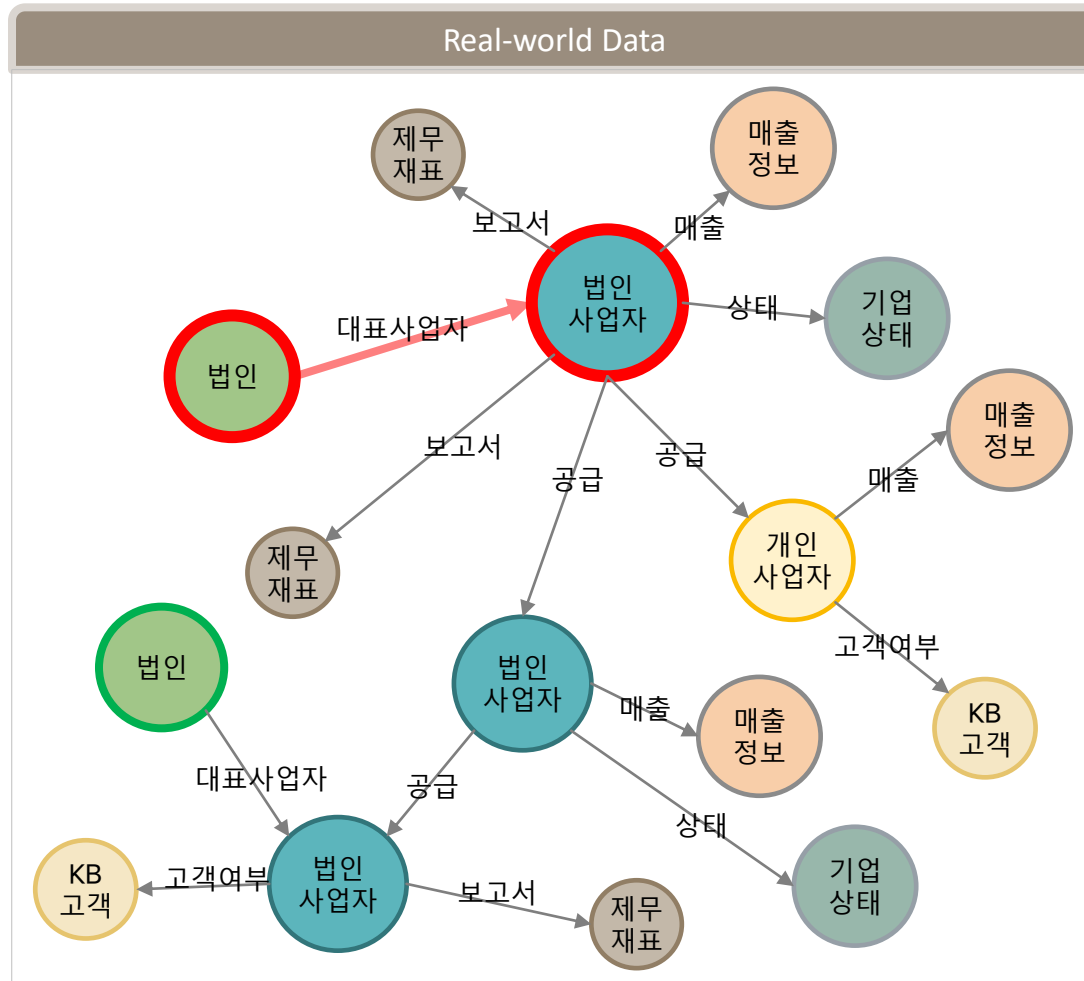
엔터프라이즈 지식그래프

- Enterprise의 주요 관심 사항
 - 조직에서 활용할 수 있는 가치 있는 콘텐츠 혹은 정보를 발견하는 것이 핵심
- Enterprise Knowledge Graph
 - 조직의 도메인 지식과 정보들을 인간과 기계가 모두 이해할 수 있도록 표현
 - 그래프 형태의 데이터 모델을 활용하여 사람, 장소, 사물, 그리고 그것들이 어떻게 관련되어 있는지를 설명하는 조직의 지식 자산, 콘텐츠 및 데이터 모음을 제공
 - 구조화된 정보와 구조화되지 않은 정보를 연결하고 조직에 대한 사실을 발견하는 데 매우 유용한 도구로 활용



왜 지식그래프를 도입하는가?

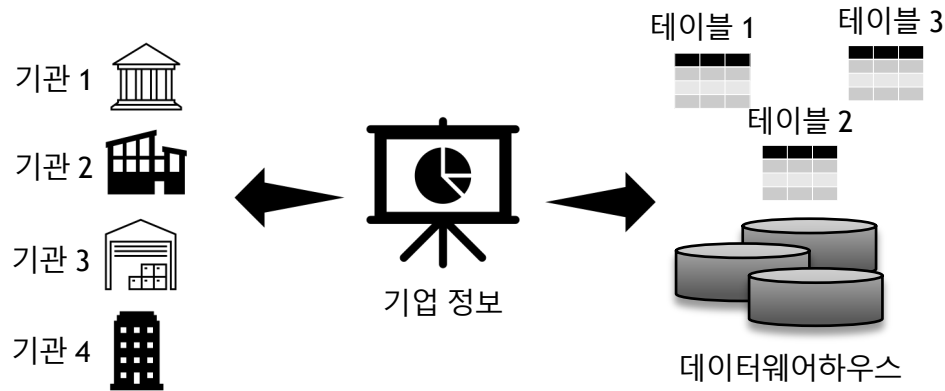
- 실제 데이터 VS 대부분의 데이터 구조



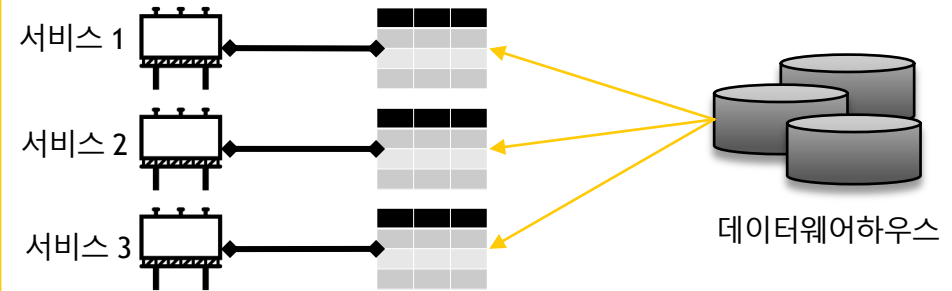
왜 지식그래프를 도입하는가?

■ 데이터 관리 측면에서의 어려움

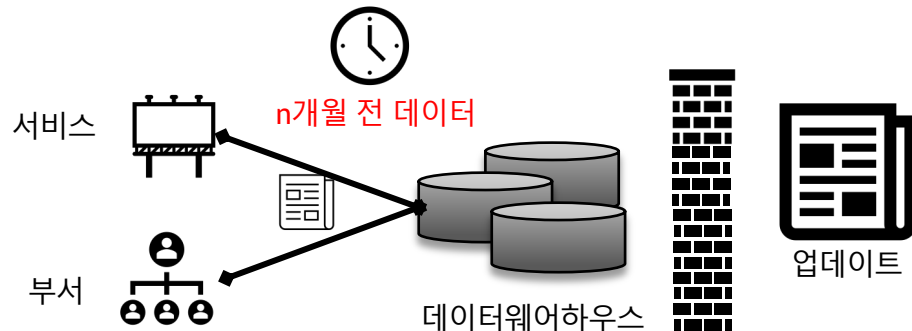
1 다양한 출처의 데이터 및 분산 존재



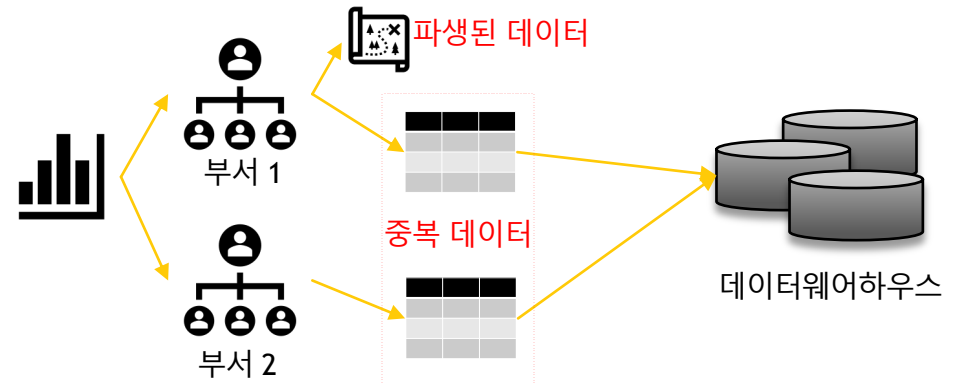
2 서비스 종속적으로 구축되는 데이터



3 적시성이 떨어지는 현상



4 데이터 중복 적재 및 고립된 관리



지식그래프로 해결하고자 하는 것

As-Is

데이터 소스 및
유형의 다양성

초기 설계된 목적에 상이한 데이터 소스 및
유형의 출현

실제 세계와 데이터
사이의 비연결성

소프트웨어 및 하드웨어의 기술 및 개념에
맞도록 데이터를 구성

제한된 의미 부여

특정 시스템에만 유용하도록 모델링 된 데이터

유연하지 않은
스키마

비즈니스 및 요구사항이 변경되는 경우 스키마
변경이 쉽게 이루어지지 못함

데이터 사일로(silo)

데이터간 서로 단절되어 존재하는 경우가
많으며, 통합은 비용과 시간이 많이 소요됨

데이터 관리 비용

애플리케이션에 따라 데이터 재구성, 또는 모델
관리에 비용이 발생

지식그래프

추상화를 통해 데이터 표현하며, 생성된
데이터는 형식과 사용 목적이 분리되어 존재

인간이 정보를 표현하고 소비하는 방식으로
구성

인간과 기계가 모두 해석할 수 있는 형식적
의미체계를 통해 생성

데이터를 동일하게 유지하면서 스키마를
변경하거나 구성이 가능

의미 있고 원활한 방식으로 데이터를 병합할 수
있고, 개방형 표준을 기반으로 구축되어 독점
포맷 또는 벤더 종속 회피에 용이

데이터 재사용, 유연한 스키마를 통해 비용이
절감되고 관리가 용이

엔터프라이즈 지식그래프 활용 기업

■ 분야별 엔터프라이즈 지식그래프 (1/2)

| 분야 | 기업 | 설명 | 주요 데이터 |
|---------|----------|---|--|
| 웹 검색 | 구글 지식그래프 | 검색 엔진은 이제 더이상 “문자열 일치”가 아닌 엔티티를 식별하려고 시도 지식그래프 자체는 엔티티와 상호 관계를 설명 | 검색에서 찾고 있는 것으로 보이는 엔티티 목록, 이미지, 속성-값 쌍 |
| 커머스(상업) | 아마존 | 제품에 대한 의미적 검색 기능을 활성화하고 온라인 마켓플레이스 사용자에게 제품 추천 개선 | 상품 |
| | eBay | 사용자가 자연어 인터페이스를 통해 관련 제품을 찾는데 도움이 되는 대화형 에이전트를 구동하는데 사용 | 제품 설명 및 쇼핑 행동 패턴 |
| | Airbnb | 주요 데이터에 대한 숙박 시설을 지식그래프로 생성하여 특정 주택 근처에서 임대할 수 있는 명소, 이벤트 및 활동에 대한 잠재 고객 추천을 제공하는데 사용 | 임대, 장소, 이벤트, 경험, 이웃, 사용자, 태그 |
| | Uber | 우버 이츠 배달 서비스를 통해 정확히 어떤 종류의 음식을 찾고 있는지 확실하지 않은 사용자에게 시맨틱 검색 기능과 권장 사항을 제공 | 음식, 레스토랑 |
| 소셜 네트워크 | Facebook | 사람들을 연결하기 위해 사용자에게 대한 소셜 데이터 뿐만 아니라 사용자가 관심 있는 엔티티를 설명하는 지식그래프를 통해 사용자의 관심사를 이해하고 권장 사항을 제공 | 소셜 데이터, 유명인, 장소, 영화, 음악 |
| | LinkedIn | 분류 스키마가 정의된 데이터로 지식그래프를 생성하고 이를 통해 중요한 개념의 다국어 번역을 제공하고, 타겟 광고를 개선, 직업 검색 및 사람 검색을 위한 고급 기능을 제공하고, 사람에게 일치하는 직업을 추천하는데 사용 | 사용자, 직업, 기술, 회사, 장소, 학교 |

엔터프라이즈 지식그래프 활용 기업

■ 분야별 엔터프라이즈 지식그래프 (2/2)

| 분야 | 기업 | 설명 | 주요 데이터 |
|-----------|---------------|---|---------------------------------------|
| 금융 | Bloomberg | 기업의 정서 분석, 질문 응답 서비스, 주식 가치에 영향을 미칠 수 있는 새로운 이벤트 감지를 포함하여 재무 데이터 분석을 지원하는 지식그래프를 제안 | 최신 뉴스 보고서 및 트윗 |
| | Refinitiv | 금융 생태계를 표현하는 지식그래프를 발표했으며 이를 통해 일부 응용 프로그램에서 공급망 모니터링, 위험 평가 및 투자 조사를 제공 | 사람, 조직, 지분 상품, 산업 분류, 합작 투자 및 제휴, 공급망 |
| | Bank of Italy | 규칙 기반 추론을 사용하여 다양한 이해관계자의 회사 소유 비율을 결정하는 등 연역적 추론에 지식 그래프를 사용 | 회사 소유 비율 |
| | Capital One | 올바른 목적을 위해 올바른 데이터를 사용하고 액세스할 수 있도록 지식그래프 | |
| | Wells Fargo | 비즈니스 조직에서 사일로된 데이터의 통합을 목적으로 활용 데이터 식별하고 지식 레이어를 구축하여 요구사항, 사양, 개념 등의 표준화를 수행 | 기업, 법인, 파생상품 등 |
| 헬스케어 | IBM | 의약품 발견과 패키지 삽입물에서 정보 추출을 위한 사용 사례를 탐구 | |
| | AstraZeneca | 유전체학 연구와 질병 이해를 진전시키기 위해 지식 그래프를 사용 | |
| 운송 | Bosch | 자동화 주행을 위한 장면 및 위치에 대한 지식 그래프를 탐구 | |
| oil & gas | Maana | 유정 및 시추와 관련된 위험 완화를 위한 데이터 통합을 수행하기 위해 지식 그래프를 사용 | |

지식그래프 활용 트렌드 & 활용 사례

소매

마케팅 담당자가 고객 행동, 다른 고객과의 관계, 사람들이 함께 구매할 수 있는 제품 종류 파악에 활용.
다양한 제품 간의 관계에 대한 지식을 사용하여 새로운 제품 출시

금융

고객 알기 및 자금 세탁 방지에 지식그래프를 활용
은행 및 기타 금융 기관은 고객 기반 전체의 자금 흐름을 추적하고 규정을 준수하지 않는 고객을 탐지

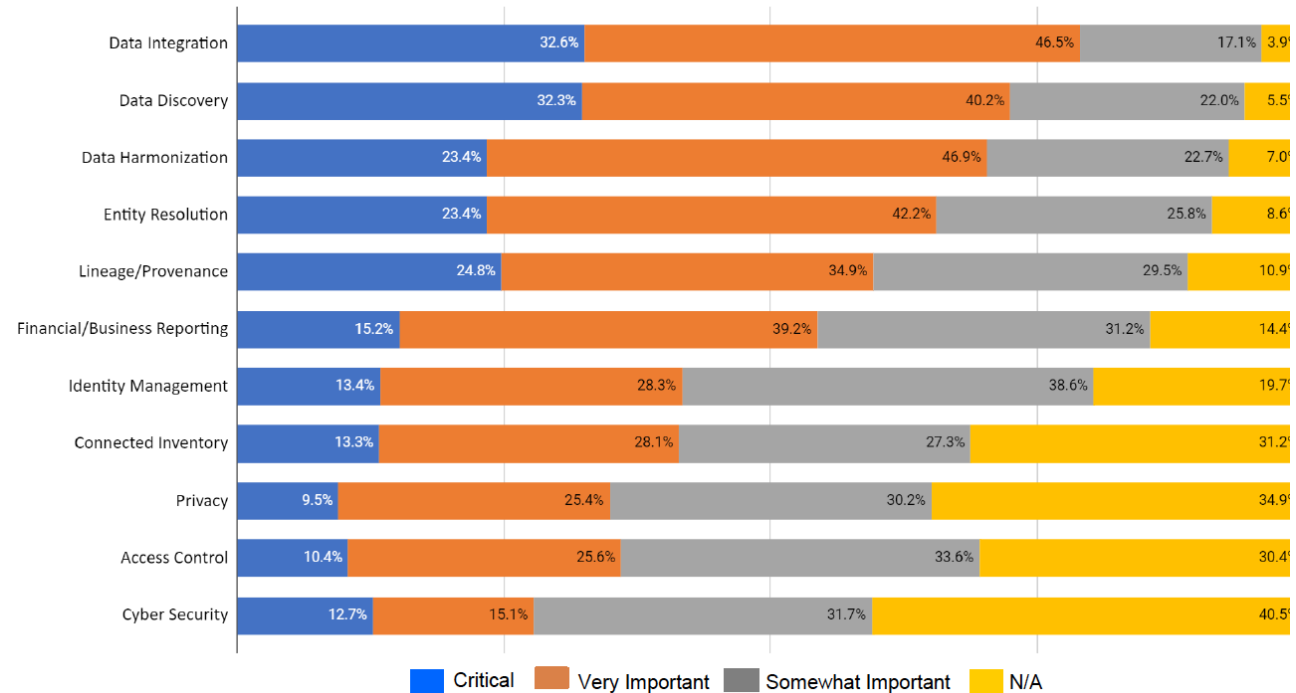
보건 의료

의료 연구 내에서 데이터 세트의 복잡한 관계를 구성하고 분류
의료 서비스 제공자가 진단을 확인하고 개인의 필요에 맞는 치료 계획을 식별하는데 도움이 되는 정보를 제공

엔터테인먼트

Netflix, Amazon, Facebook 등과 같은 콘텐츠 플랫폼은 지식 그래프를 사용하여 인공지능 기반 추천 엔진을 구축
온라인 참여 행동의 데이터를 사용하여 새로운 콘텐츠 추천

지식그래프 활용 사례

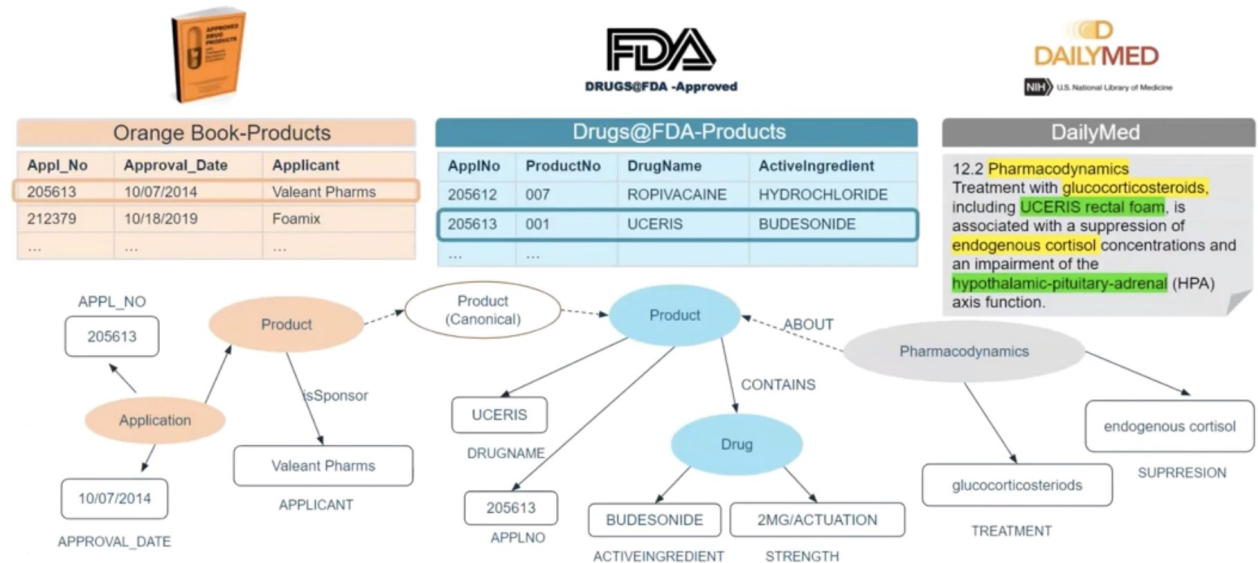


지식그래프 활용 기관 사례

- FDA



Knowledge graphs are connected graphs of data and metadata that richly model real-world entities.



마무리

- 이제 데이터는 넘쳐 있는 시대
- 필요한 데이터를 어떻게 잘 엮어서 가치있는 정보를 사용할 것인가
- 지식그래프는 지식과 데이터의 결합
- 지식그래프 기술은 데이터 통합과 새로운 지식의 발견에 장점을 제공
- 데이터 통합, 데이터 발견이 선행되어야 의사 결정과 비즈니스 혁신이 일어날 수 있다.



감사합니다

Blog : <https://joyhong.tistory.com/>

Mail : su4620@gmail.com