Industry-level Knowledge Graph Platform for Large-scale, Diverse and Dynamic Scenarios

Haofen Wang

LLMKG Workshop, VLDB 2024

01 **Knowledge Graph:** An effective way to manage domain knowledge

02 **Opportunities and Challenges:** Knowledge management paradigm shift from binary static to multi-dynamic

03 **SPG:** A novel semantic framework that accelerates data knowledgeization and knowledge symbolization

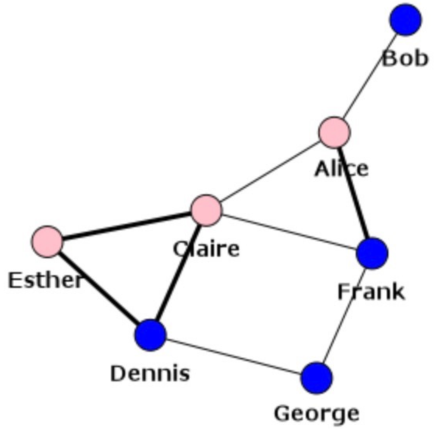
04 **Applications:** SPG-based knowledge graph cases

05 **Future prospects:** The application value brought by SPG and its dual-drive development with LLM in the future

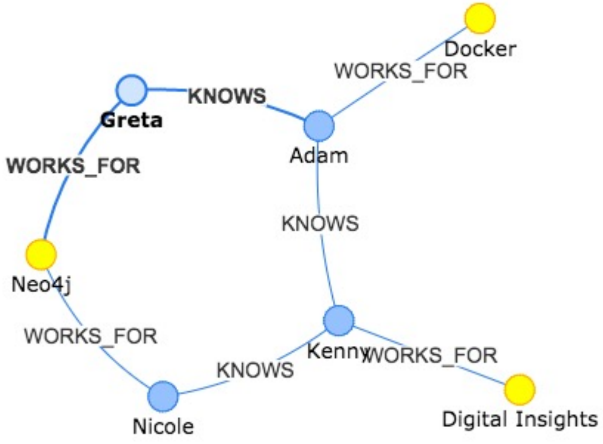
Knowledge Graph: Semantic and High-order Knowledge Management Solution for Data

Graph data structure development trends

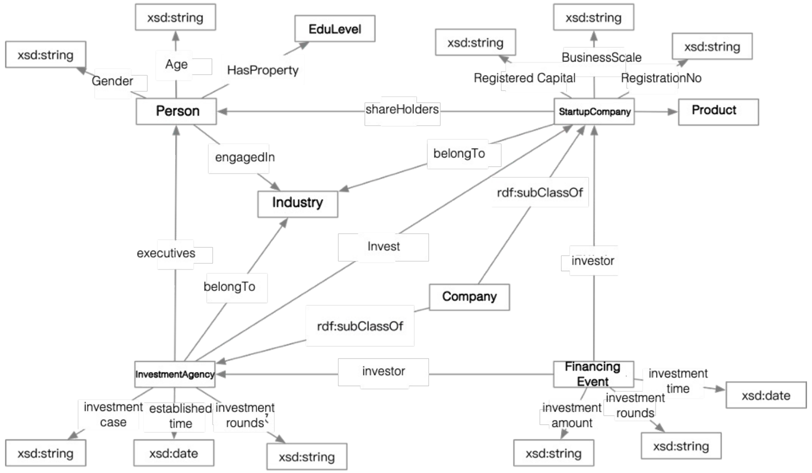
homogeneous graph



heterogeneous graph



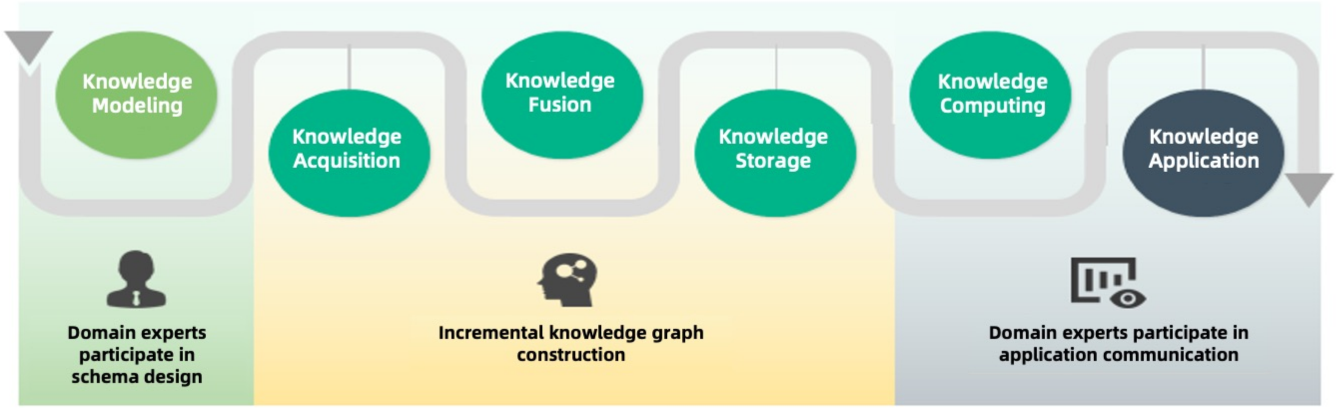
Heterogeneous graphs with rich properties



Characteristics of Knowledge Graph Algorithm

- Knowledge Graph = Graph + Knowledge Base
  - Graph： Learning graph structure
  - Knowledge Base： Learning knowledge semantics
- Rely on NLP and Graph algorithms
- Rely on knowledge of Domain experts

Lifecycle of Domain Knowledge Graph

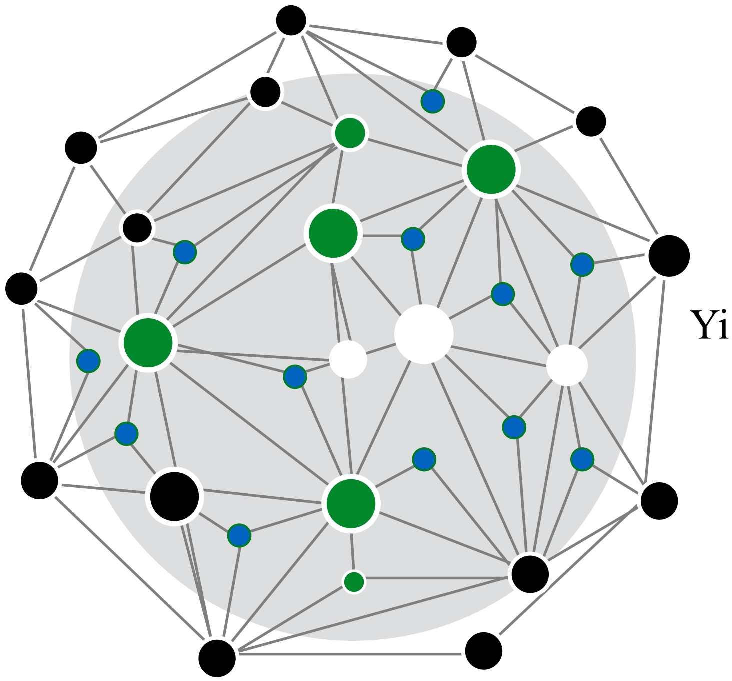


Domain experts participate in schema design

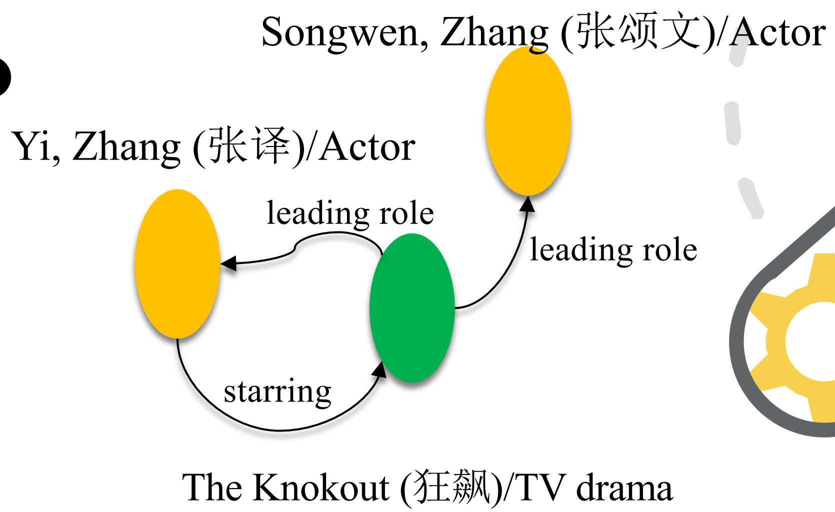
Incremental knowledge graph construction

Domain experts participate in application communication

Knowledge Graph: Semantic and High-order Knowledge Management Solution for Data



Definition and Slogan (Things, not Strings)



Songwen, Zhang (张颂文)/Actor

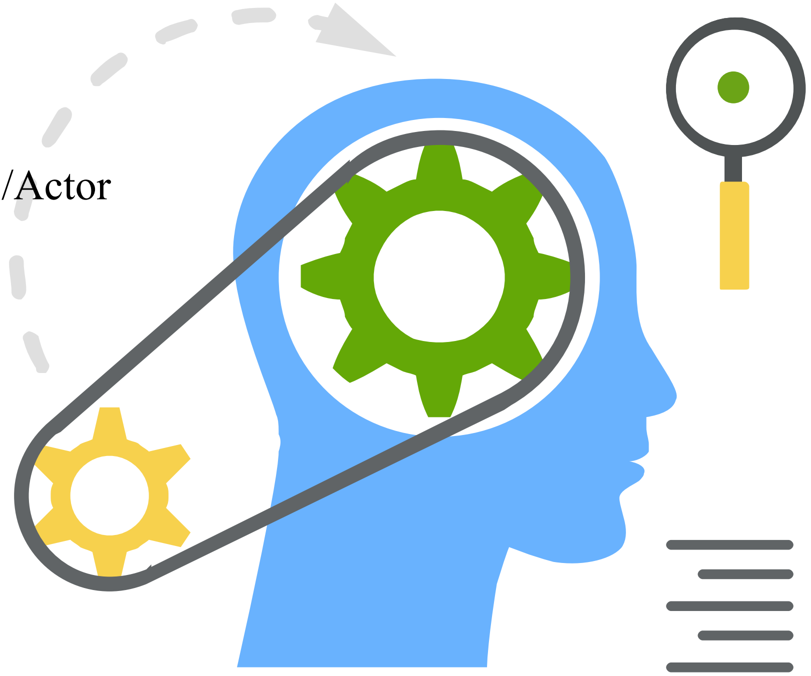
Yi, Zhang (张译)/Actor

leading role

leading role

starring

The Knokout (狂飙)/TV drama



Controllable AI

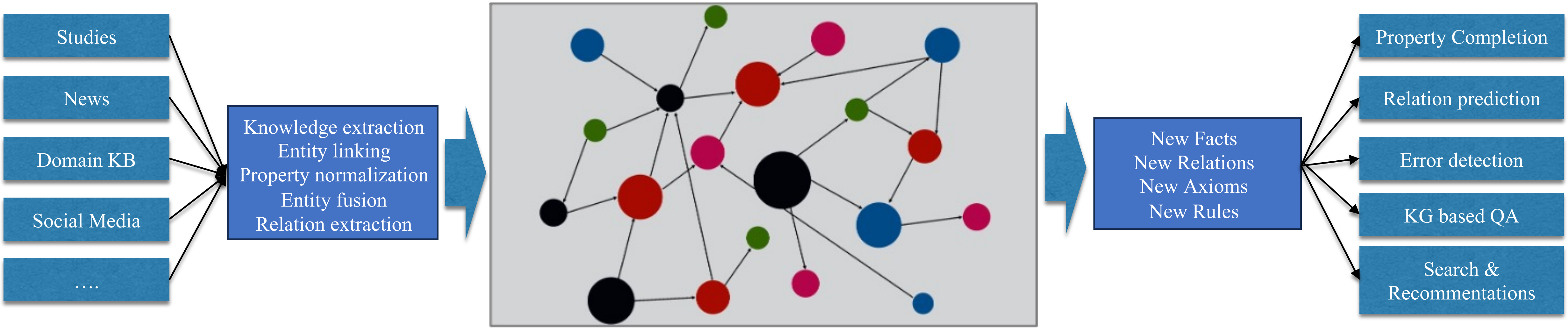
Knowledge graph is a method of modeling the world to achieve knowledge standardization and semantic interconnection of data.

Text: One dimension

KG: Multi dimension

From public documents

Knowledge Graph: Semantic and High-order Knowledge Management Solution for Data



Studies

News

Domain KB

Social Media

….

Knowledge extraction
Entity linking
Property normalization
Entity fusion
Relation extraction

New Facts
New Relations
New Axioms
New Rules

Property Completion

Relation prediction

Error detection

KG based QA

Search & Recommentations

Event, Entity, Concept, Property, Relation, etc.

- **Domain Knowledge Management**：Formalized knowledge representation based on knowledge semantics and graph structure supports the efficient construction of domain knowledge graphs
- **Knowledge Normalization：** Utilize knowledge graph-related technologies to continuously improve the standardization and normalization level of entities, events, concepts, properties, relationships, etc.
- **Cross-knowledge graphs Reuse and Fusion:** Through knowledge fusion capabilities, we can realize cross-graph connection and reuse, connect data silos, and reduce business costs and improve efficiency.
- **Knowledge Reasoning and Discovery：** Discover more rare knowledge based on graph reasoning insights to serve scenarios such as risk control, credit, insurance claims, merchant operations, and marketing recommendation.

01 **Knowledge Graph：** An effective way to manage domain knowledge

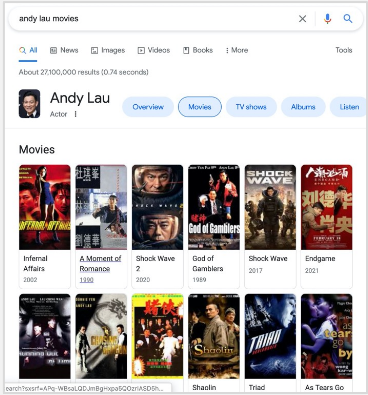
02 **Opportunities and Challenges:** Knowledge management paradigm shift from binary static to multi-dynamic

03 **SPG:** A novel semantic framework that accelerates data knowledgeization and knowledge symbolization

04 **Applications:** SPG-based knowledge graph cases

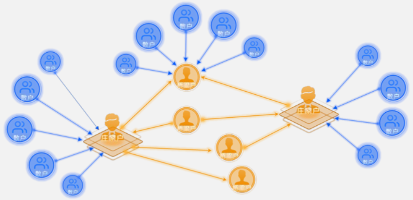
05 **Future prospects:** The application value brought by SPG and its dual-drive development with LLM in the future

Opportunities and Challenges of Knowledge Graph Technology



Commonsense knowledge Graphs (2012)

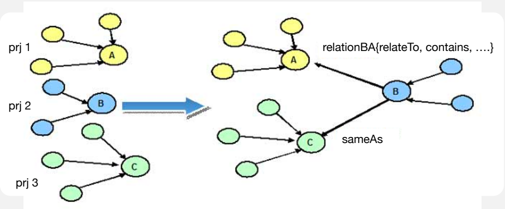
Search engine applications such as Baidu and Google



Domain Knowledge Graphs (2018)

Promote enterprise digitalization

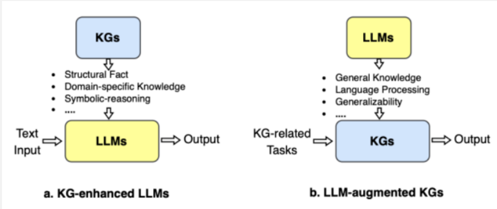
According to the 2022 China Knowledge Graph Market Industry Report by iResearch Consulting, the market space is expected to **grow from 10.7 billion yuan in 2021 to 29 billion yuan in 2026**, with the financial and public security sectors being the major drivers.



Integration and linking across knowledge graphs (2021)

Data Fabric

Enterprise-level knowledge graphs connect data silos



LLMs + KG (2023)

《Unifying Large Language Models and Knowledge Graphs》

Bidirectional drive of LLMs and KGs

Going beyond traditional knowledge graphs, leveraging knowledge graph technology to drive digital transformation in enterprises.

Commonsense, Accuracy tolerance

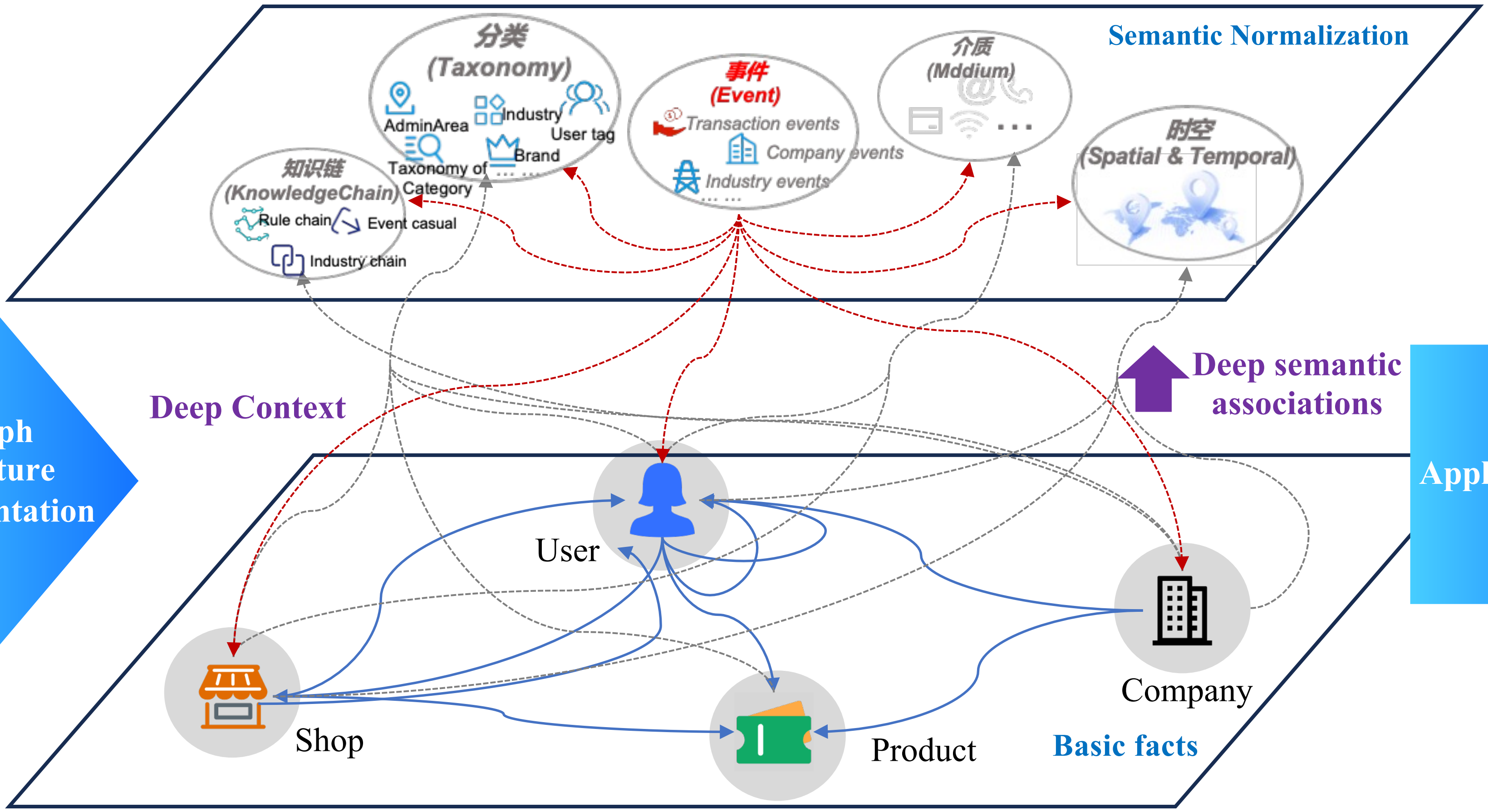
Deep-context aware, Interpretable

Take Ant Group Knowledge Graph Application as an Example

Financial scenarios

Graph Structure Representation

- Operating data
- External risk events
- Heterogeneous dynamic information
- Complicated domain rules



AI Applications

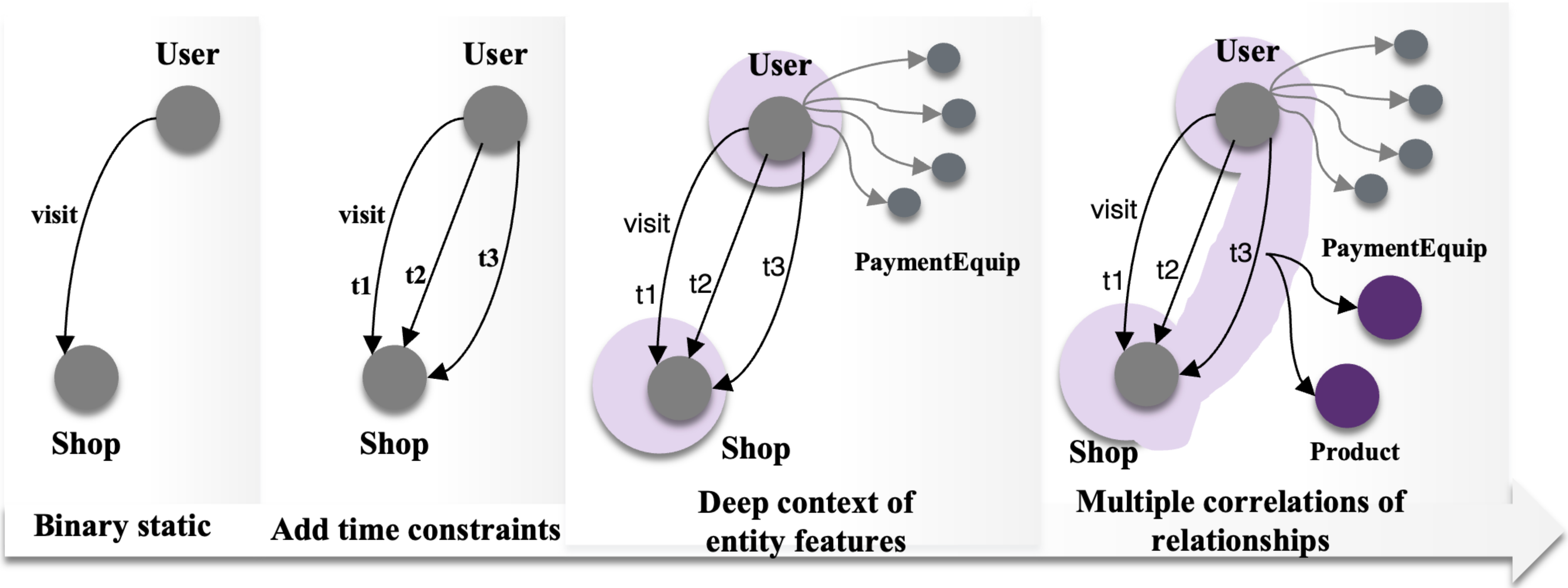
Business Growth

- Recommendation
- Fortune Q&A
- Merchant promotion
- … …

Risk Control

- Anti-money laundering
- Fund tracking
- eKYB / eKYC
- … …

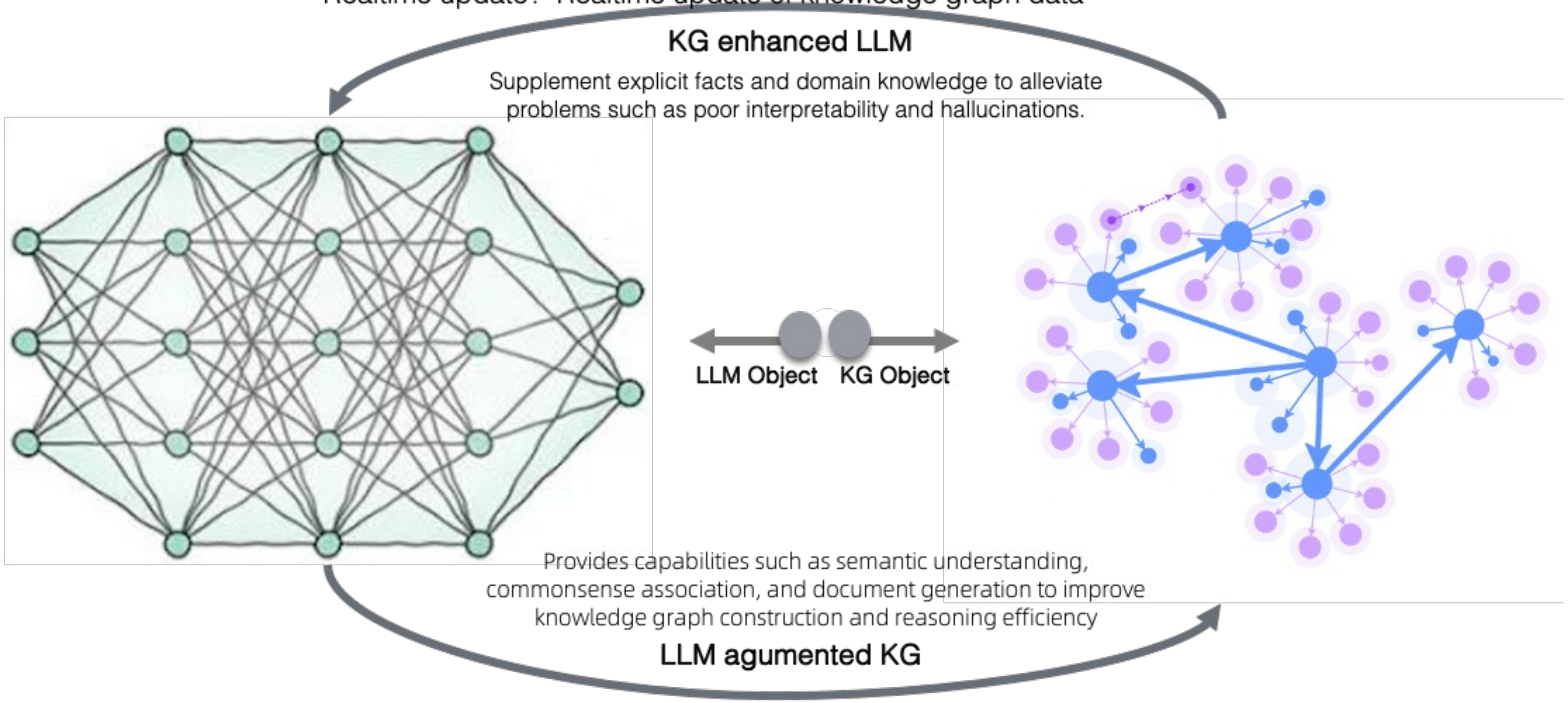
Paradigm Shift in Knowledge Representation from Binary Statics to Multi Dynamics



- The common sense knowledge graph only uses conceptual level induction, which cannot perceive individual differences and cannot achieve individual-oriented reasoning and judgment.
- Industrial-level knowledge management requires strong context awareness to achieve operational understanding and risk insights of thin customer groups.
- Knowledge-enhanced LLM also require domain knowledge graphs to cover more domain common sense knowledge, entities and events

Dual-drive Enhancement of LLM and KG in Enterprise Digital Scenarios

- Factual knowledge: Provide structured, semantic, factual and interpretable factual knowledge
- Domain expert knowledge： Provide domain-specific business expert knowledge
- Realtime update： Realtime update of knowledge graph data



KG enhanced LLM

Supplement explicit facts and domain knowledge to alleviate problems such as poor interpretability and hallucinations.

LLM Object KG Object

Provides capabilities such as semantic understanding, commonsense association, and document generation to improve knowledge graph construction and reasoning efficiency

LLM agumented KG

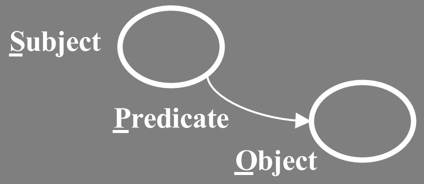
- **Semantic understanding:** document structure and entity word understanding to assist knowledge extraction
- **Commonsense knowledge:** Knowledge completion and commonsense knowledge association based on LLM

Dual-drive Enhancement of LLM and KG in Enterprise Digital Scenarios

Scenarios and applications		LLM only	KG enhanced LLM	LLM augmented KG	KG only
Business Growth	Interactive applications	Chat, write poems and songs	Knowledge Q&A, service retrieval, report analysis, etc.	-	Marketing recommendation, event context, marketing decision-making, etc.
	Marketing Recommendation	-	Data report query, crowd label selection, intelligent copywriting, etc.	-	Event analysis, materials analysis, crowd analysis, etc.
Risk Control	Risk forecasting and control	-	Explanatory message generation, waking up the robot, etc.	-	Clues tracking, events transmission, rule based claims, corporate credit, ultimate beneficiaries, equity penetration, etc.
Knowledge Construction	Knowledge extraction	-	-	Document element extraction, event extraction, entity linking, etc.	Knowledge construction based on structured business data
	Knowledge completion	-	-	Obtain the entity LLM embedding representation, extract and supplement the missing knowledge in the knowledge graph from the LLM	Relationships mining, properties prediction, groups mining, rules mining, etc.

Opportunities and Challenges of Knowledge Graph Technology Development

Lack of unified knowledge modeling method


Subject

Predicate

Object

RDF : Used for knowledge exchange, strong semantics, high threshold, **few industrial applications**

LPG : Used for graph storage and query, weak semantics, low threshold**, many industrial applications**

High cost of knowledge construction and acquisition

Knowledge extraction lacks a unified paradigm

Expert knowledge is **complex and lacking in reuse**

Lack of unified technology framework

There are already many tools for modeling, construction, storage, and reasoning **but lack of framework support, making cross-scenario reuse costly.**

The development of the knowledge graph's own technical system needs to keep pace with the times

01 **Knowledge Graph：** An effective way to manage domain knowledge

02 **Opportunities and Challenges:** Knowledge management paradigm shift from binary static to multi-dynamic

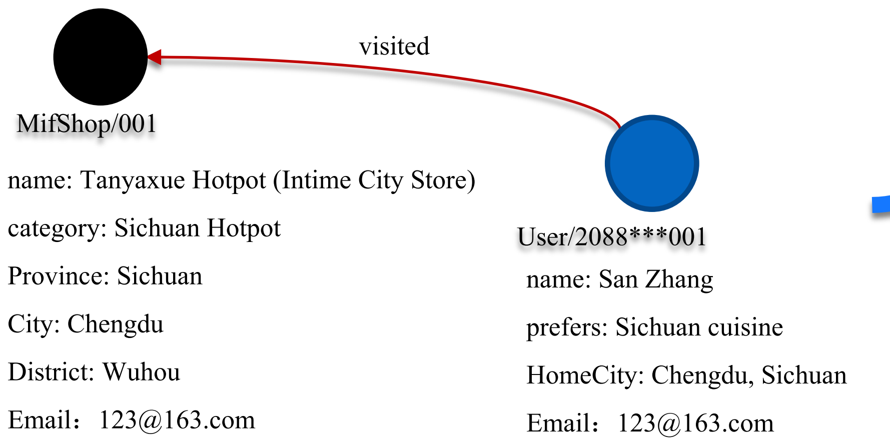
03 **SPG:** A novel semantic framework that accelerates data knowledgeization and knowledge symbolization

04 **Applications:** SPG-based knowledge graph cases

05 **Future prospects:** The application value brought by SPG and its dual-drive development with LLM in the future

SPG: Semantic-enhanced Programmable Graph (Schematic Diagram)

Factual entities and relationships

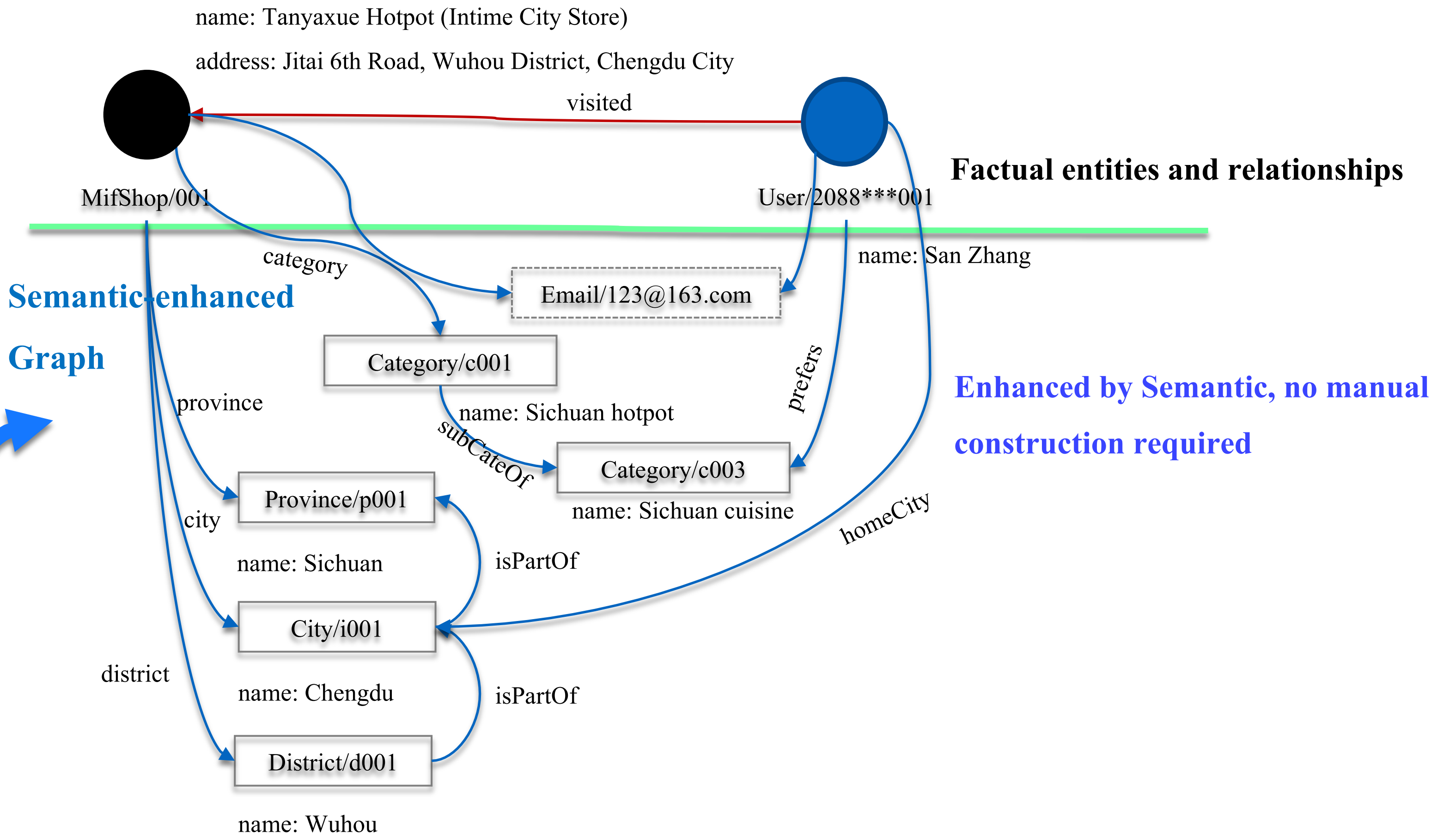


name: Tanyaxue Hotpot (Intime City Store)
category: Sichuan Hotpot
Province: Sichuan
City: Chengdu
District: Wuhou
Email：123@163.com

name: San Zhang
prefers: Sichuan cuisine
HomeCity: Chengdu, Sichuan
Email：123@163.com

Property Graph

Slogan of KG: Things, not Strings


name: Tanyaxue Hotpot (Intime City Store)
address: Jitai 6th Road, Wuhou District, Chengdu City

Factual entities and relationships

Semantic-enhanced Graph

Enhanced by Semantic, no manual construction required

Three distinctive characteristics of knowledge under the SPG framework:

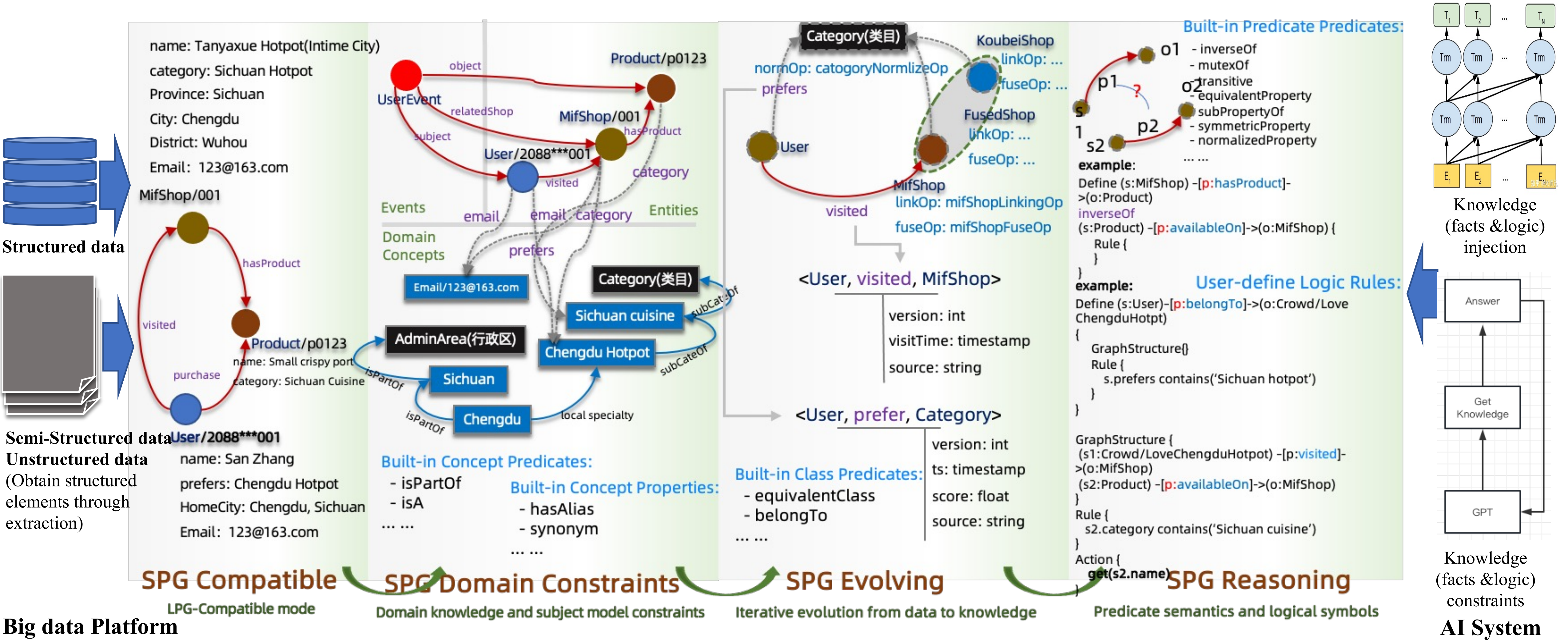
1. Every Thing must have a Class, every object in the real world belongs to at least one classification

2. Each instance is unique within a Class, SPG uses NLP algorithms to build capabilities such as entity linking, property normalization, and entity fusion.

3. No Thing exists in isolation, define knowledge element dependencies through predicates and logic.

SPG: Semantic-enhanced Programmable Graph (L1 – L3)

Connecting big data and AI technology systems to help machines better understand the world



Structured data

Semi-Structured data

Unstructured data

(Obtain structured elements through extraction)

Big data Platform

name: Tanyaxue Hotpot(Intime City)
category: Sichuan Hotpot
Province: Sichuan
City: Chengdu
District: Wuhou
Email：123@163.com

MifShop/001

Product/p0123
name: Small crispy port
category: Sichuan Cuisine

User/2088***001
name: San Zhang
prefers: Chengdu Hotpot
HomeCity: Chengdu, Sichuan
Email：123@163.com

SPG Compatible

LPG-Compatible mode

Built-in Concept Predicates:
- isPartOf
- isA
… …

Built-in Concept Properties:
- hasAlias
- synonym
… …

SPG Domain Constraints

Domain knowledge and subject model constraints

<User, visited, MifShop>
version: int
visitTime: timestamp
source: string

<User, prefer, Category>
version: int
ts: timestamp
score: float
source: string

Built-in Class Predicates:
- equivalentClass
- belongTo
… …

SPG Evolving

Iterative evolution from data to knowledge

Built-in Predicate Predicates:

- inverseOf
- mutexOf
- transitive
- equivalentProperty
subPropertyOf
- symmetricProperty
- normalizedProperty
… …

```
example:
Define (s:MifShop) –[p:hasProduct]-
>(o:Product)
inverseOf
(s:Product) –[p:availableOn]->(o:MifShop) {
    Rule {
    }
}
```

User-define Logic Rules:

```
example:
Define (s:User)–[p:belongTo]->(o:Crowd/Love
ChengduHotpt)
{
    GraphStructure{}
    Rule {
       s.prefers contains('Sichuan hotpot')
    }
}

GraphStructure {
 (s1:Crowd/LoveChengduHotpot) –[p:visited]-
>(o:MifShop)
 (s2:Product) –[p:availableOn]->(o:MifShop)
}
Rule {
  s2.category contains('Sichuan cuisine')
}
Action {
   get(s2.name)
}
```

SPG Reasoning

Predicate semantics and logical symbols

Knowledge (facts &logic) injection

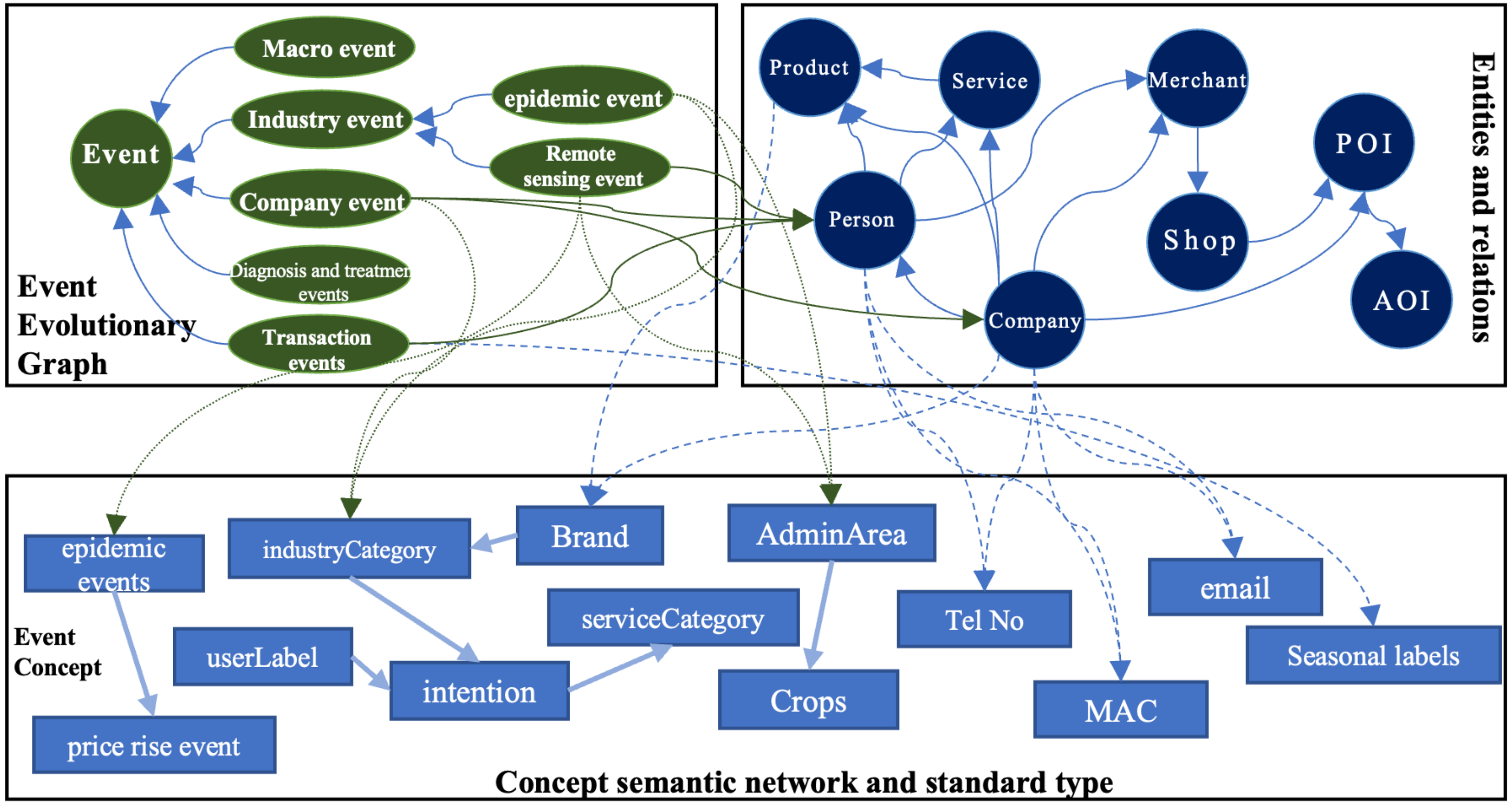
Knowledge (facts &logic) constraints

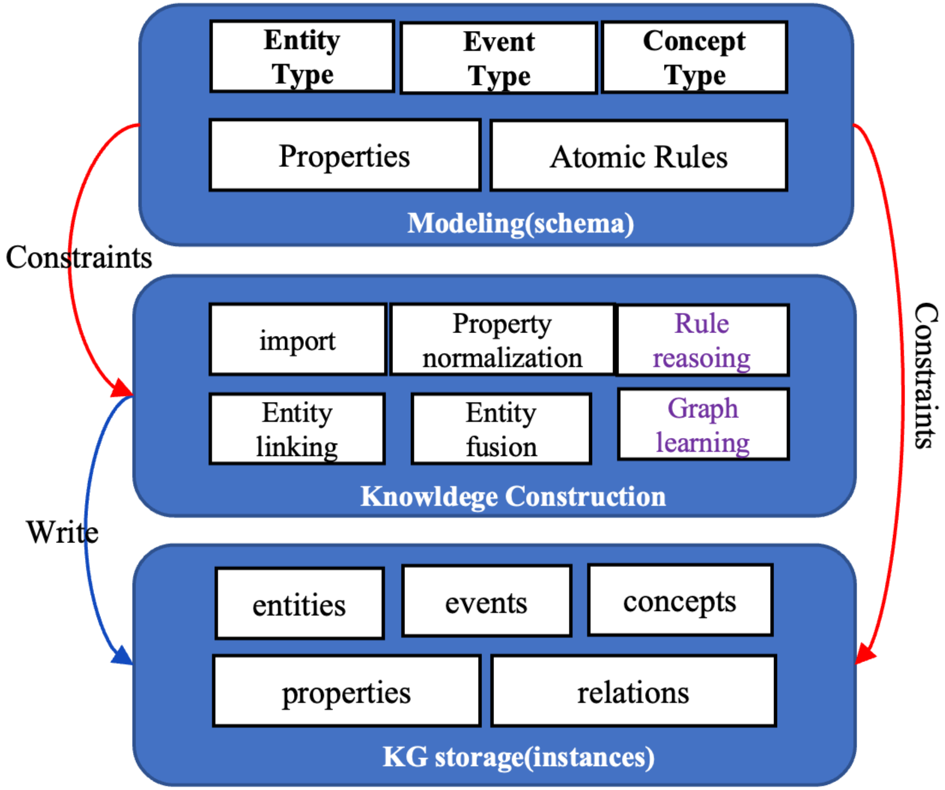
AI System

SPG: Subject/Object Type Definition (Class–Instance Paradigm)

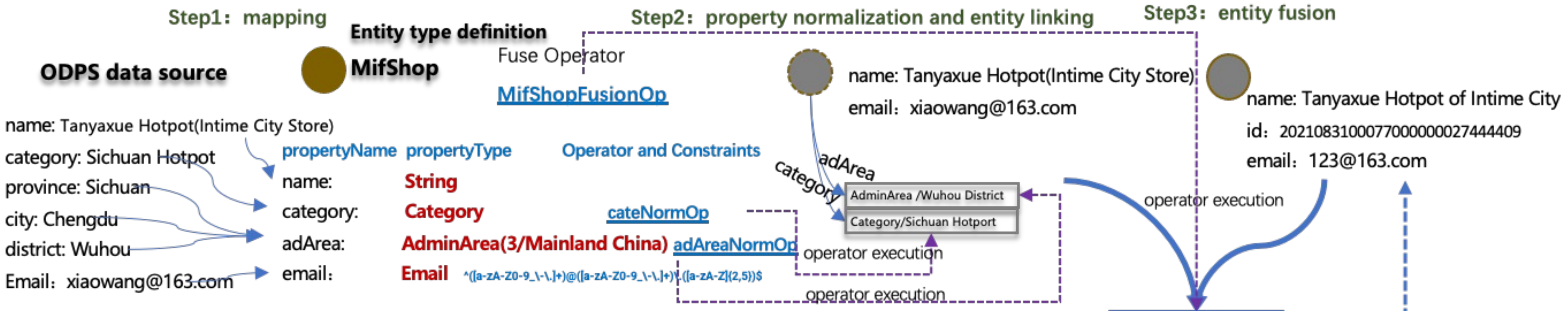
No clear boundary between the definitions of entities, events and concepts in the industry. SPG’s definition of knowledge types is:

- **Entity Type:** Objective objects that have strong business relevance and are described by multiple elements. Multiple elements are described through entity Properties (attributes and relationships), such as users, enterprises, merchants, etc.
- **Concept Type:** The inductive abstraction of entity objects from specific to general expresses a set of entities, which is a classification system in an inductive sense. Relatively static and highly reusable, such as crowd tags, domain standard types, semantic vocabulary (such as HowNet), etc.
- **Event type:** Add time, space, target and other constraints, such as industry events, corporate events, diagnosis and treatment events extracted through NLP or CV





SPG: Knowledge Construction based on Structured data (Programmable)



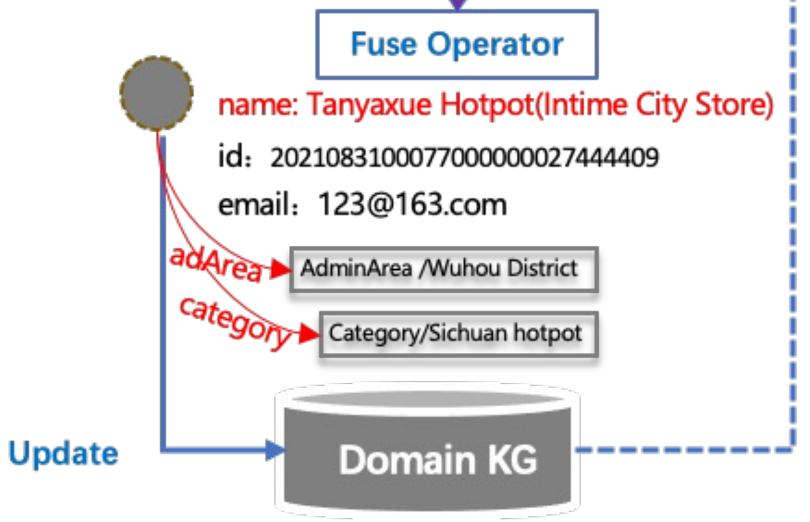
propertyName	propertyType	Operator and Constraints
name:	String	
category:	Category	cateNormOp
adArea:	AdminArea(3/Mainland China)	adAreaNormOp
email:	Email	^([a-zA-Z0-9_\-\.]+)@([a-zA-Z0-9_\-\.]+)\.([a-zA-Z]{2,5})$

Problem solved:

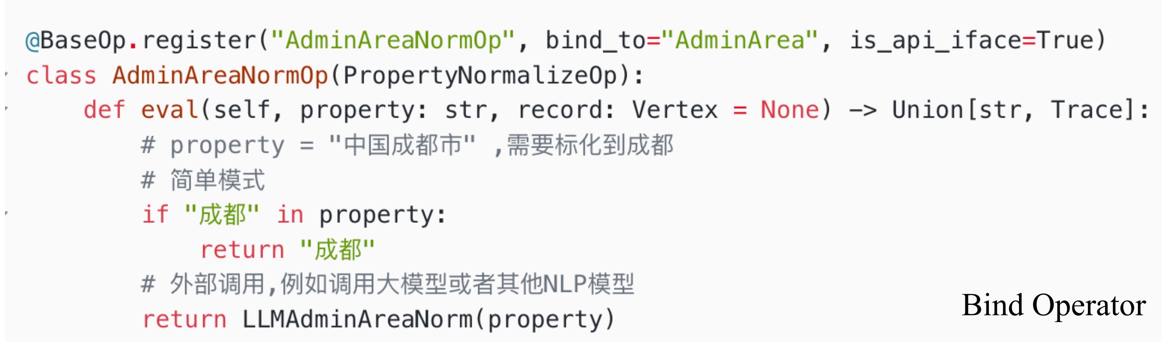
1) Obtain the structured expression of the target Entity and Concept type.

2) Map structured data to the target property and associated operators.

3) Run operator to implement entity linking, property normalization and entity fusion.



Fuse Operator
name: Tanyaxue Hotpot(Intime City Store)
id：20210831000770000000027444409
email：123@163.com
adArea
AdminArea /Wuhou District
category
Category/Sichuan hotpot
Update
Domain KG

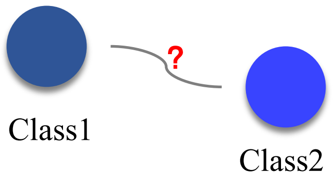


```
@BaseOp.register("AdminAreaNormOp", bind_to="AdminArea", is_api_iface=True)
class AdminAreaNormOp(PropertyNormalizeOp):
    def eval(self, property: str, record: Vertex = None) -> Union[str, Trace]:
        # property = "中国成都市" ,需要标化到成都
        # 简单模式
        if "成都" in property:
            return "成都"
        # 外部调用,例如调用大模型或者其他NLP模型
        return LLMAdminAreaNorm(property)
```

Bind Operator

SPG: Predicate Semantics and Logical Symbols

Semantics of Types



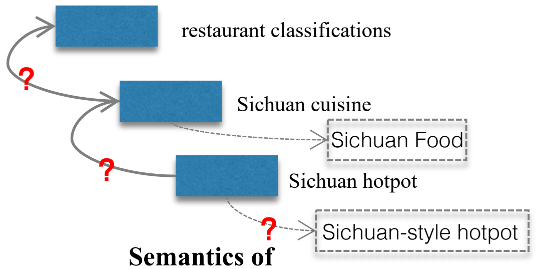
Built-in predicates:

- equivalentClass
- belongTo
- sameAs

Pseudo code:

```
Create EntityType FusedPOI equivalentClass (
   fuse(AmapPOI, AlipayPOI)
   .withLinkFunction(samePoiSimilarityFunc)
   .withFuseStrategy(){
      FusedPOI.attr1 = isNotBlank(AmapPOI.attr1) ?
AmapPOI.attr1 : AlipayPOI:attr1
      FusedPOI.attrx = （AmapPOI.attrx1 > AlipayPOI.attrx2
）
          … …
   }
)
```

Semantics of Concept hypernym



Semantics of Property extension

Built-in hypernym predicates :

- isPartOf
- subCategoryOf
- isA
- …

Built-in property predicates:

- hasAlias
- synonym
- …

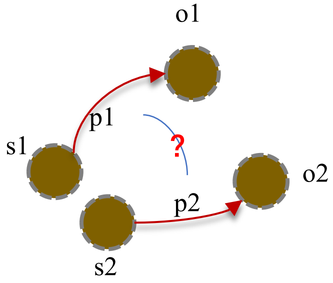
Examples of semantic-enhanced Prompt:

getPrompts(‘Sichuan hotpot’)

hypernym: Sichuan Food

synonyms: Sichuan-style hotpot, Chongqing hotpot, Basu hotpot

Semantics of Relations and Properties



Built-in predicates:

- inverseOf
- mutexOf
- transitive
- equivalentProperty
- subPropertyOf
- symmetricProperty
- normalizedProperty

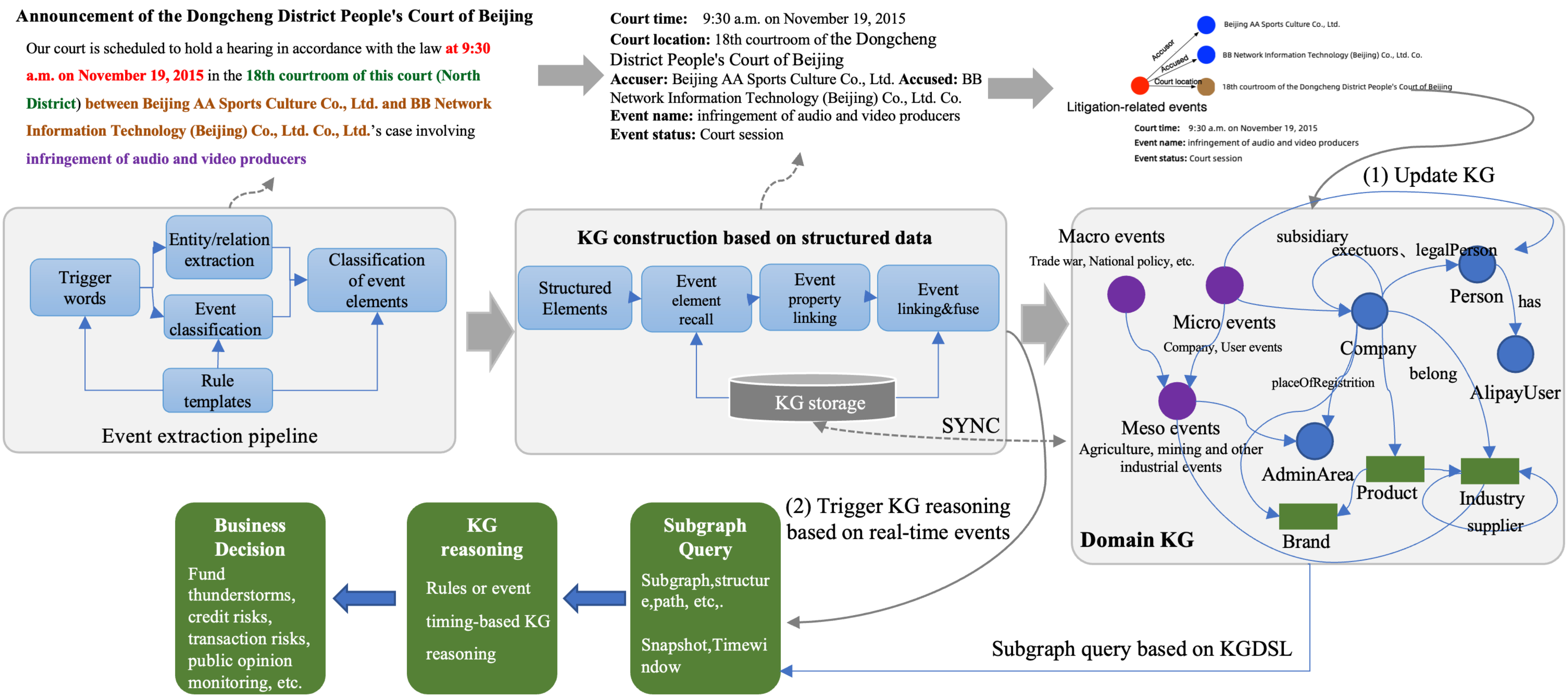
… …

```
Define (s:MifShop) –[p:hasProduct]->(o:Product)
inverseOf
(s:Product) –[p:availableOn]->(o:MifShop) {
   Rule { }
}

Define (s:User)–[p:belongTo]->(o:Crowd/LoveChengduHotpot)
{
   GraphStructure{}
   Rule {
      s.prefers contains(‘Sichuan hotpot’)
   }
}

GraphStructure {
  (s1:Crowd/ LoveChengduHotpot) –[p:visited]->(o:MifShop)
  (s2:Product) –[p:availableOn]->(o:MifShop)
}
Rule {
   s2.category contains(‘Sichuan cuisine’)
}
Action {
  get(s2.name)
}
```

SPG: Event Extraction based on Knowledge Construction Pipeline



Announcement of the Dongcheng District People's Court of Beijing

Our court is scheduled to hold a hearing in accordance with the law **at 9:30 a.m. on November 19, 2015** in the **18th courtroom of this court (North District)** **between Beijing AA Sports Culture Co., Ltd. and BB Network Information Technology (Beijing) Co., Ltd. Co., Ltd.**'s case involving **infringement of audio and video producers**

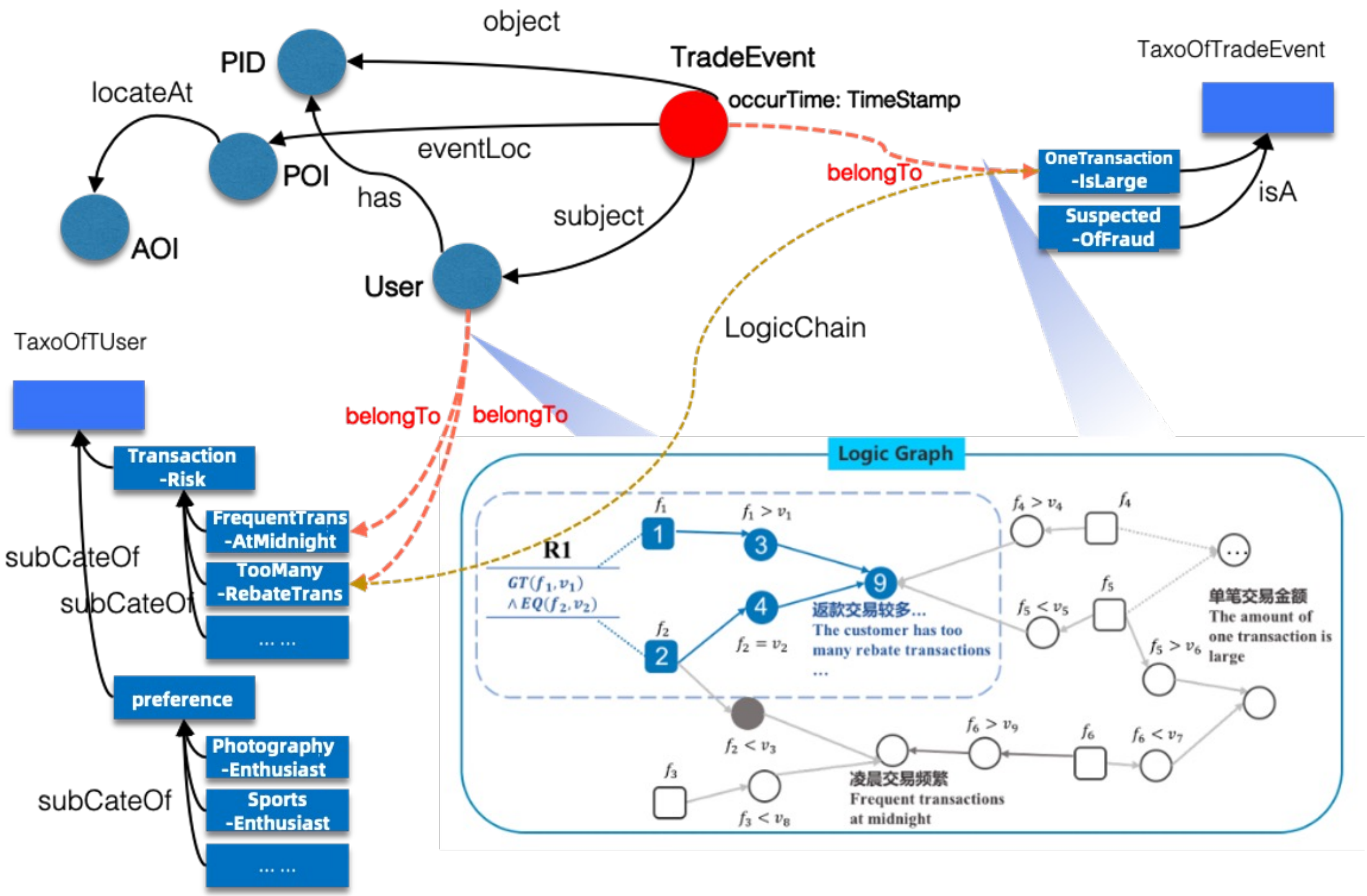
Court time: 9:30 a.m. on November 19, 2015
Court location: 18th courtroom of the Dongcheng District People's Court of Beijing
Accuser: Beijing AA Sports Culture Co., Ltd. **Accused:** BB Network Information Technology (Beijing) Co., Ltd. Co.
Event name: infringement of audio and video producers
Event status: Court session

Accusor
Accused
Court location
Beijing AA Sports Culture Co., Ltd.
BB Network Information Technology (Beijing) Co., Ltd. Co.
18th courtroom of the Dongcheng District People's Court of Beijing
Litigation-related events

Court time: 9:30 a.m. on November 19, 2015
Event name: infringement of audio and video producers
Event status: Court session

Trigger words
Entity/relation extraction
Event classification
Classification of event elements
Rule templates
Event extraction pipeline
KG construction based on structured data
Structured Elements
Event element recall
Event property linking
Event linking&fuse
KG storage
SYNC
(1) Update KG
Macro events
Trade war, National policy, etc.
Micro events
Company, User events
Meso events
Agriculture, mining and other industrial events
subsidiary
exectuors、legalPerson
Person
has
Company
belong
AlipayUser
placeOfRegistrition
AdminArea
Product
Industry
supplier
Brand
Domain KG
(2) Trigger KG reasoning based on real-time events
Subgraph query based on KGDSL
Subgraph Query
Subgraph,structure,path, etc,.
Snapshot,Timewindow
KG reasoning
Rules or event timing-based KG reasoning
Business Decision
Fund thunderstorms, credit risks, transaction risks, public opinion monitoring, etc.

SPG: Implementing LogicChain based on Semantic Logic



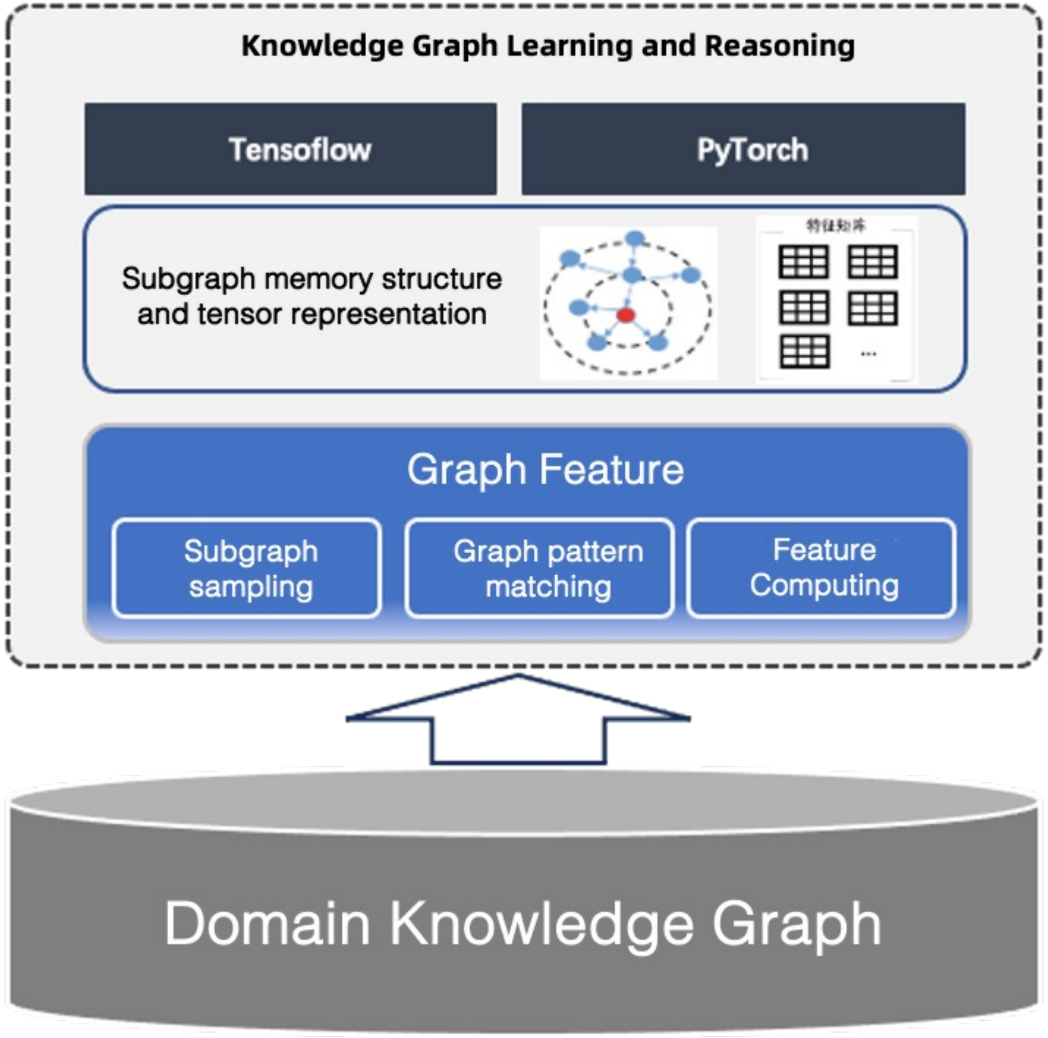
```
Define (e:TradeEvent)–[p:belongTo]->(o:TaxoOfTradeEvent/OneTransactionIsLarge)
{
    GraphStructure{}
    Rule {
        e.amount > 500
    }
}
Define (s:User)–[p:belongTo]->(o:TaxoOfUser/TransactionRisk/TooManyRebateTrans)
{
    GraphStructure{
      (e1:TradeEvent)-[ps1:subject]->(su1:User)
      (e1:TradeEvent)-[pp1:object]->(sp1:PID)
      (e2:TradeEvent)-[ps2:subject]->(su2:User)
      (e2:TradeEvent)-[pp2:object]->(sp2:PID)
      (su1)-[has]->(sp2)
      (su2)-[has]->(sp1)
      (e2)-[pb:belongTo]->(o:/TaxoOfTradeEvent/OneTransactionIsLarge)
    }
    Rule {
        s.id == su1.id
        e1.ts < e2.ts and hour(current_time()) - hour(e1.ts) < 24
        group(s).count() > 10
    }

}
Define (s:User)–[p:belongTo]->(o:TaxoOfUser/ TransactionRisk /FrequentTransAtMidnight)
{
    GraphStructure{
      (e1:TradeEvent)-[ps1:subject]->(su1:User)
}
    Rule {
        s.id == su1.id
        hour(e1.occurTime) between(0, 4)
    }
}
```

SuspectedOfProstitution: FrequentTransAtMidnight & OneTransactionIsLarge

LargeCashOutUsers: TooManyRebateTrans

SPG: Graph Learning Subgraph Sampling based on GNN (Programmable)



Knowledge Graph Learning and Reasoning

Tensoflow

PyTorch

Subgraph memory structure and tensor representation

Graph Feature

Subgraph sampling

Graph pattern matching

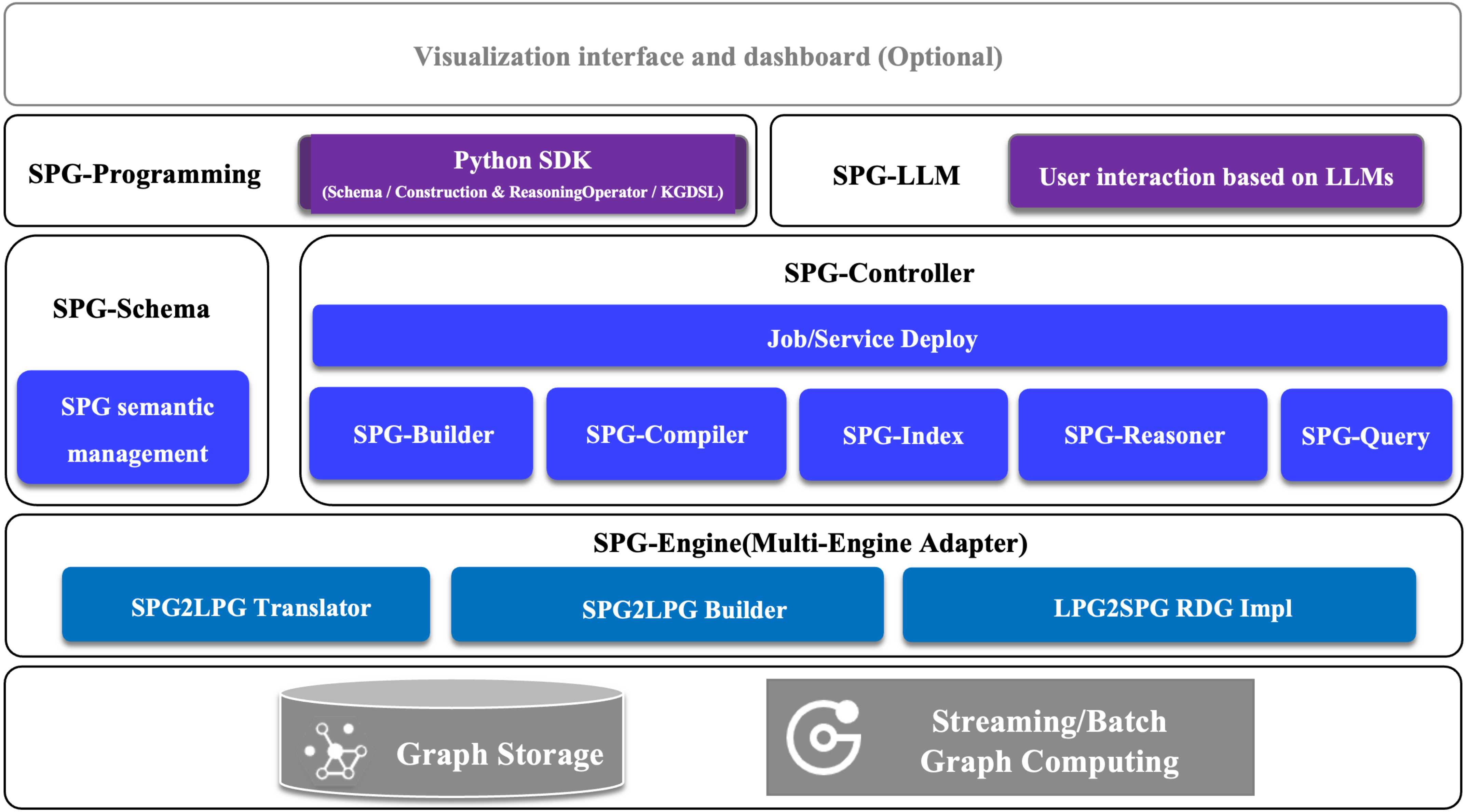
Feature Computing

Domain Knowledge Graph

```
# -*- coding: utf-8 -*-
import libkg_client
from kgrl.conf import KgrlConstants # noqa
from kgrl.data import KGExpression # noqa
from kgrl.data.sampler import KGStateCacheBaseSampler

in_degree = KGExpression.SourceNodeInDegreeKey()
out_degree = KGExpression.SourceNodeOutDegreeKey()
node_version = KGExpression.SourceNodeVersionKey()
edge_version = KGExpression.EdgeVersionKey()
v_begin = 30
v_end    = 40
def get_filters(v_begin, v_end):
   return {
      KgrlConstants.NEIGHBORHOOD_SAMPLING_FILTER_NAME: f"{edge_version}<{v_begin} and       {edge_version}>{v_end}",
      KgrlConstants.NODE_SAMPLING_FILTER_NAME: f"{node_version}==0",
      KgrlConstants.EDGE_SAMPLING_FILTER_NAME: f"{edge_version}<{v_begin} and {edge_version}>{v_end}",
   }
def get_weights(v_begin, v_end):
   return {
      KgrlConstants.NEIGHBORHOOD_SAMPLING_WEIGHT_NAME: f"abs({edge_version}- {v_begin})*log2({edge_version}+{v_end})",
      KgrlConstants.NODE_SAMPLING_WEIGHT_NAME: f"({out_degree}+{in_degree})",
      KgrlConstants.EDGE_SAMPLING_WEIGHT_NAME: f"abs({edge_version}-{v_begin})*log2({edge_version}+{v_end})",
   }
sampler_conf = {
    "client_conf": {...},
    "gen_data_conf": {
         "random": True, "fanouts": [50, 20], "buffer_size": 2, "filters": get_filters(10, 20), "weights": get_weights(10, 20),
   },
}
sampler = NodeSubGraphSampler.from_params(sampler_conf)
```

SPG: Architecture of SPG



Visualization interface and dashboard (Optional)

SPG-Programming

Python SDK
(Schema / Construction & ReasoningOperator / KGDSL)

SPG-LLM

User interaction based on LLMs

SPG-Schema

SPG semantic management

SPG-Controller

Job/Service Deploy

SPG-Builder

SPG-Compiler

SPG-Index

SPG-Reasoner

SPG-Query

SPG-Engine(Multi-Engine Adapter)

SPG2LPG Translator

SPG2LPG Builder

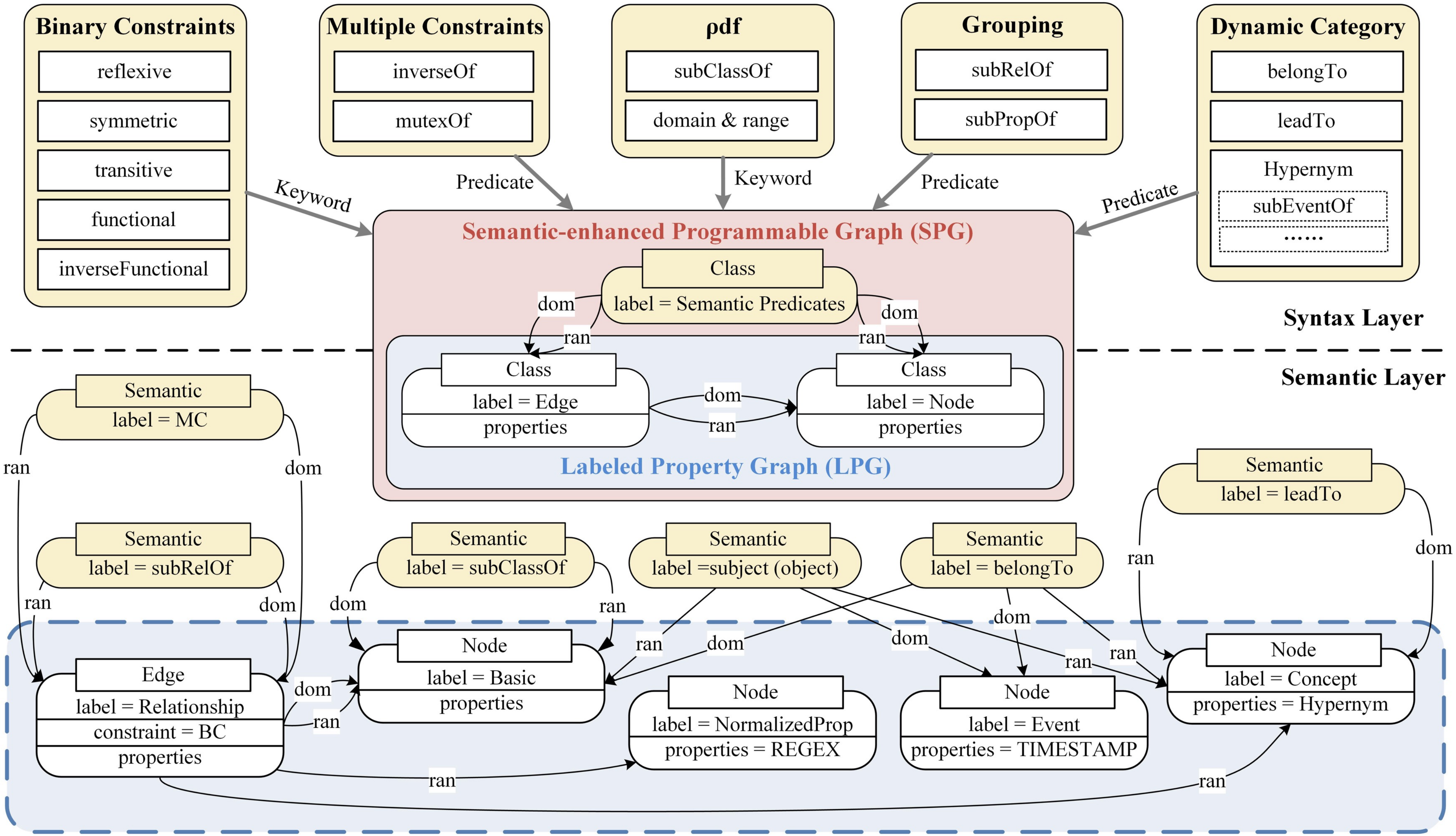
LPG2SPG RDG Impl

Graph Storage

Streaming/Batch
Graph Computing

Official website： https://spg.openkg.cn/

SPG: SPG-Schema



Binary Constraints
reflexive
symmetric
transitive
functional
inverseFunctional
Keyword

Multiple Constraints
inverseOf
mutexOf
Predicate

ρdf
subClassOf
domain & range
Keyword

Grouping
subRelOf
subPropOf
Predicate

Dynamic Category
belongTo
leadTo
Hypernym
subEventOf
……
Predicate

Semantic-enhanced Programmable Graph (SPG)
Class
label = Semantic Predicates
dom
ran
dom
ran
Class
label = Edge
properties
dom
ran
Class
label = Node
properties
Labeled Property Graph (LPG)

Syntax Layer

Semantic Layer

Semantic
label = MC
ran
dom

Semantic
label = subRelOf
ran
dom

Semantic
label = subClassOf
dom
ran

Semantic
label =subject (object)
ran
dom

Semantic
label = belongTo
dom
ran
dom

Semantic
label = leadTo
ran
dom

Edge
label = Relationship
constraint = BC
properties
dom
ran

Node
label = Basic
properties

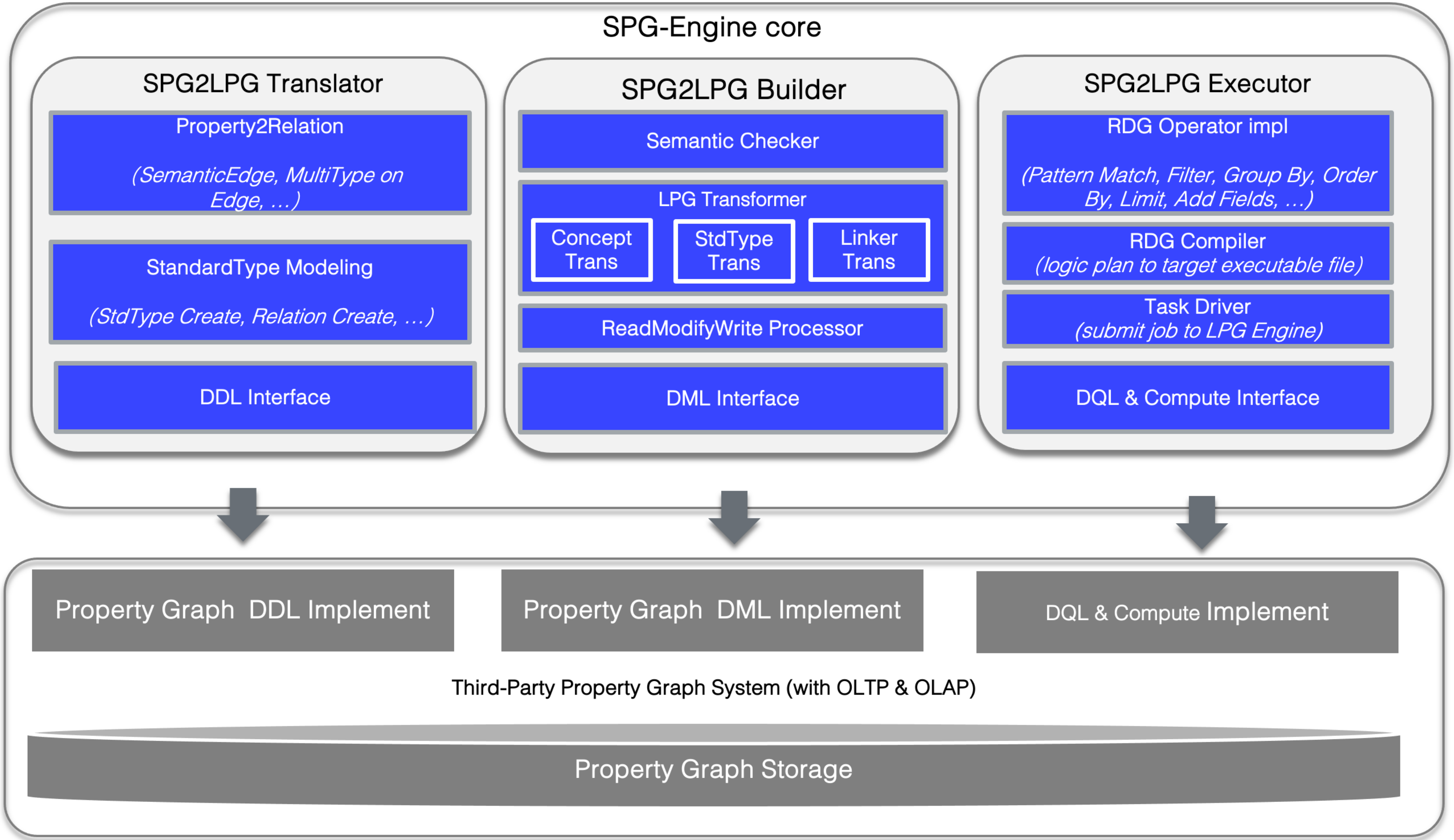
Node
label = NormalizedProp
properties = REGEX

Node
label = Event
properties = TIMESTAMP

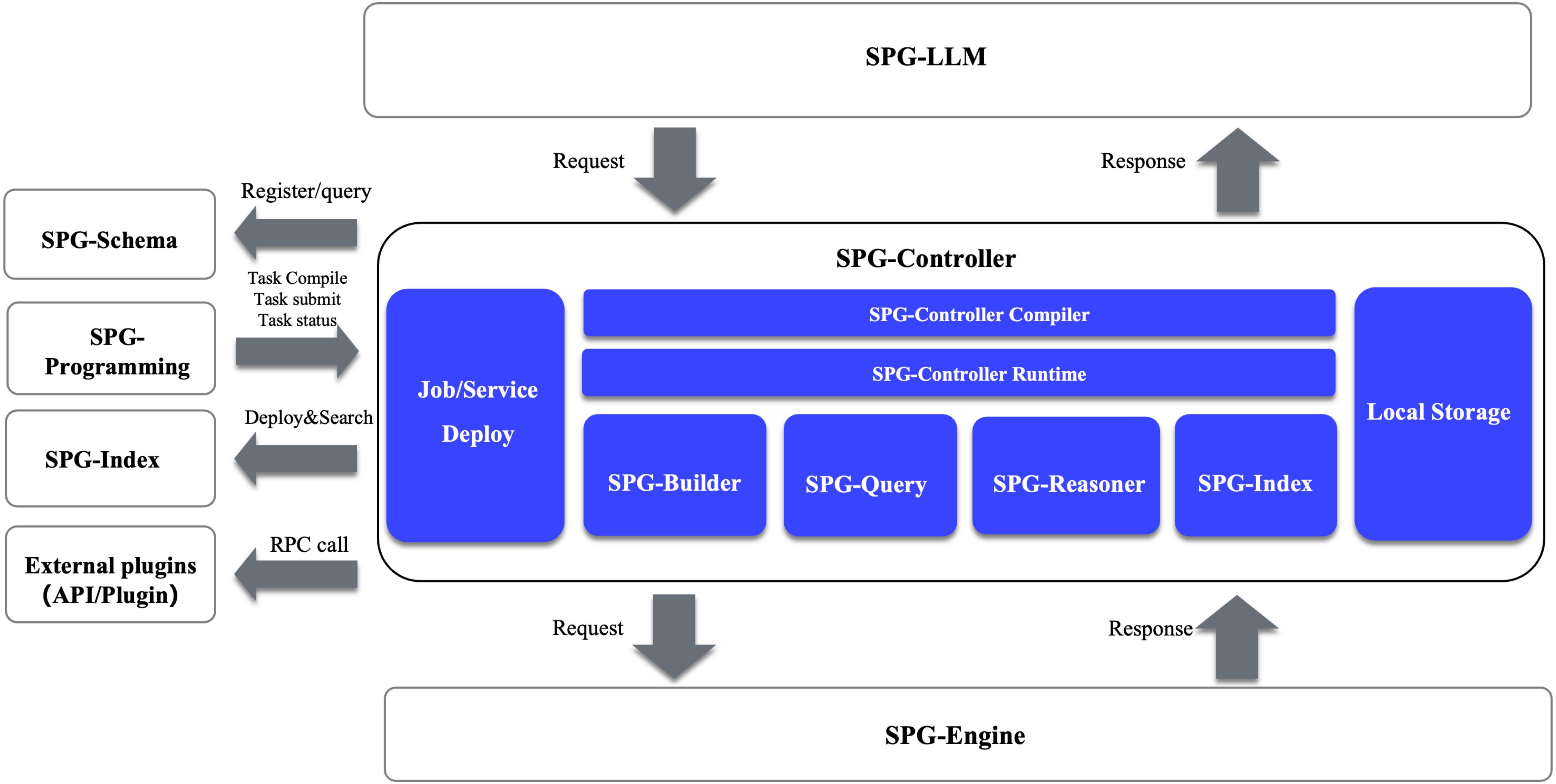
Node
label = Concept
properties = Hypernym

ran
ran

SPG: SPG-Engine


SPG-Engine core
SPG2LPG Translator
Property2Relation
(SemanticEdge, MultiType on Edge, …)
StandardType Modeling
(StdType Create, Relation Create, …)
DDL Interface
SPG2LPG Builder
Semantic Checker
LPG Transformer
Concept Trans
StdType Trans
Linker Trans
ReadModifyWrite Processor
DML Interface
SPG2LPG Executor
RDG Operator impl
(Pattern Match, Filter, Group By, Order By, Limit, Add Fields, …)
RDG Compiler
(logic plan to target executable file)
Task Driver
(submit job to LPG Engine)
DQL & Compute Interface
Property Graph DDL Implement
Property Graph DML Implement
DQL & Compute Implement
Third-Party Property Graph System (with OLTP & OLAP)
Property Graph Storage

SPG: SPG-Controller



01 **Knowledge Graph：** An effective way to manage domain knowledge

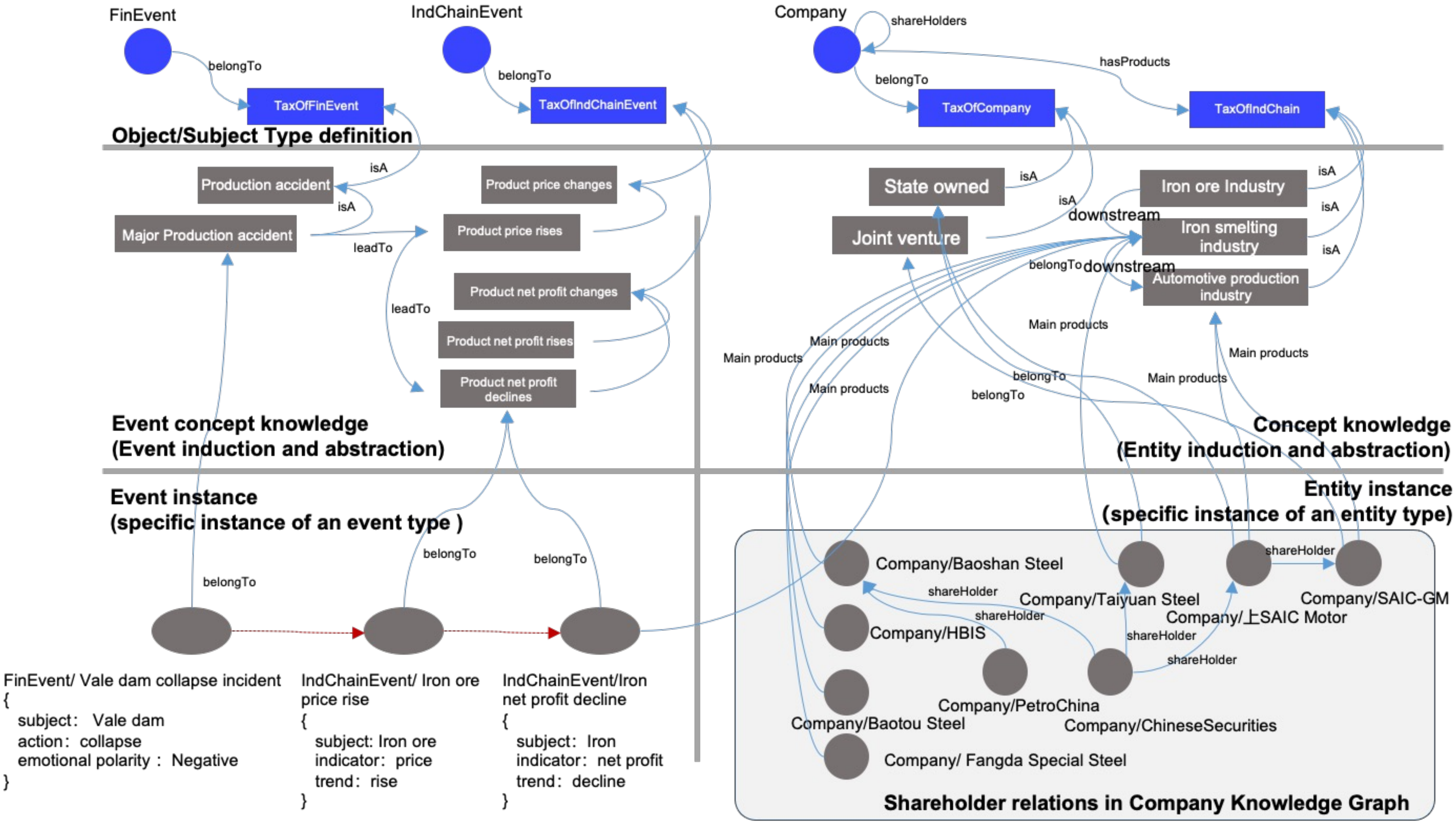
02 **Opportunities and Challenges:** Knowledge management paradigm shift from binary static to multi-dynamic

03 **SPG:** A novel semantic framework that accelerates data knowledgeization and knowledge symbolization

04 **Applications:** SPG-based knowledge graph cases

05 **Future prospects:** The application value brought by SPG and its dual-drive development with LLM in the future

Four-quadrant Model of Event Evolutionary KG from the Perspective of SPG



FinEvent

belongTo

TaxOfFinEvent

IndChainEvent

belongTo

TaxOfIndChainEvent

Company

shareHolders

hasProducts

belongTo

TaxOfCompany

TaxOfIndChain

Object/Subject Type definition

isA

Production accident

isA

Major Production accident

leadTo

leadTo

Product price changes

Product price rises

Product net profit changes

Product net profit rises

Product net profit declines

Event concept knowledge
(Event induction and abstraction)

State owned

isA

Joint venture

isA

downstream

Iron ore Industry

isA

Iron smelting industry

isA

belongTo

downstream

Automotive production industry

isA

Main products

Main products

Main products

Main products

belongTo

belongTo

Main products

Main products

Concept knowledge
(Entity induction and abstraction)

Event instance
(specific instance of an event type)

belongTo

belongTo

belongTo

FinEvent/ Vale dam collapse incident
{
subject: Vale dam
action: collapse
emotional polarity : Negative
}

IndChainEvent/ Iron ore price rise
{
subject: Iron ore
indicator: price
trend: rise
}

IndChainEvent/Iron net profit decline
{
subject: Iron
indicator: net profit
trend: decline
}

Entity instance
(specific instance of an entity type)

Company/Baoshan Steel

shareHolder

Company/Taiyuan Steel

shareHolder

Company/SAIC-GM

Company/上SAIC Motor

shareHolder

shareHolder

shareHolder

Company/HBIS

Company/PetroChina

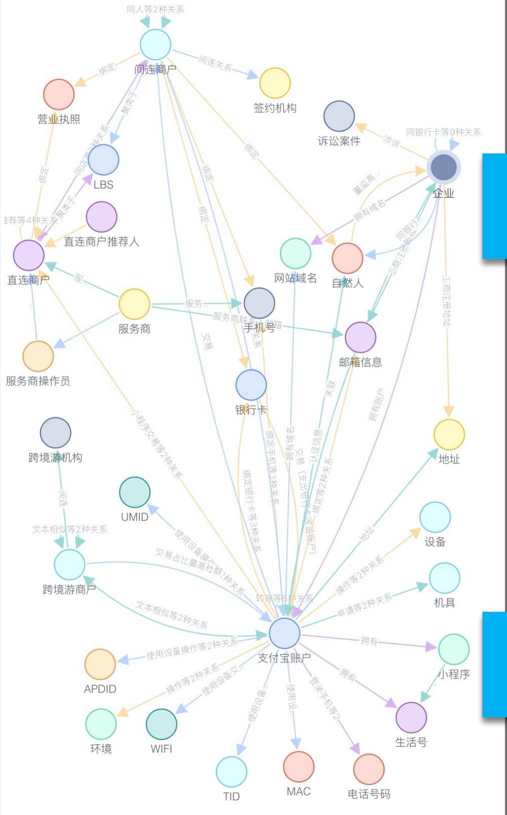
Company/ChineseSecurities

Company/Baotou Steel

Company/ Fangda Special Steel

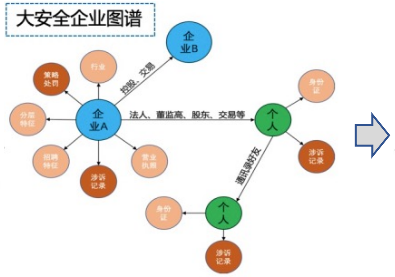
Shareholder relations in Company Knowledge Graph

Interpretable Reasoning based on the Integration of Rules and Graph Models

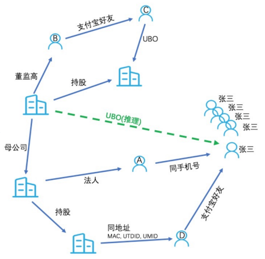


Risk Control Knowledge Graph

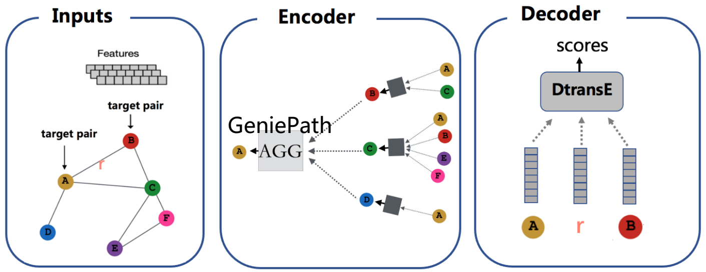
Prediction Model: KARI + XGB



Construct a Risk Control KG including entities such as User, Company, Medium, Account, and their relations



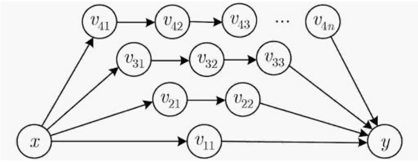
Subgraph pattern matching and subgraph sampling



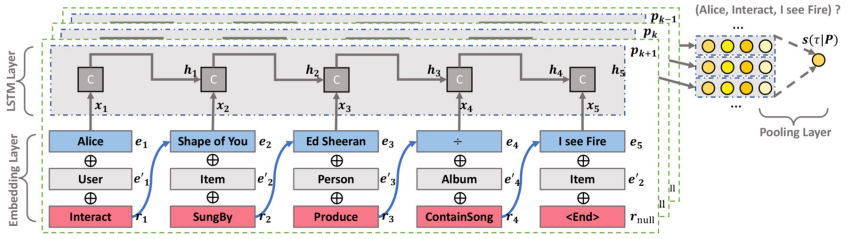
Problem solved

- ✓ Breaking down data barriers by constructing KG
- ✓ Established tens of billions of graph representation learning and reasoning capabilities

Interpretation Model: KGDSL + KPRN



Path generation with KGDSL



KPRN

Problem solved

- ✓ For each link prediction result, path weights can be generated as an interpretable representation of the results

Business Target

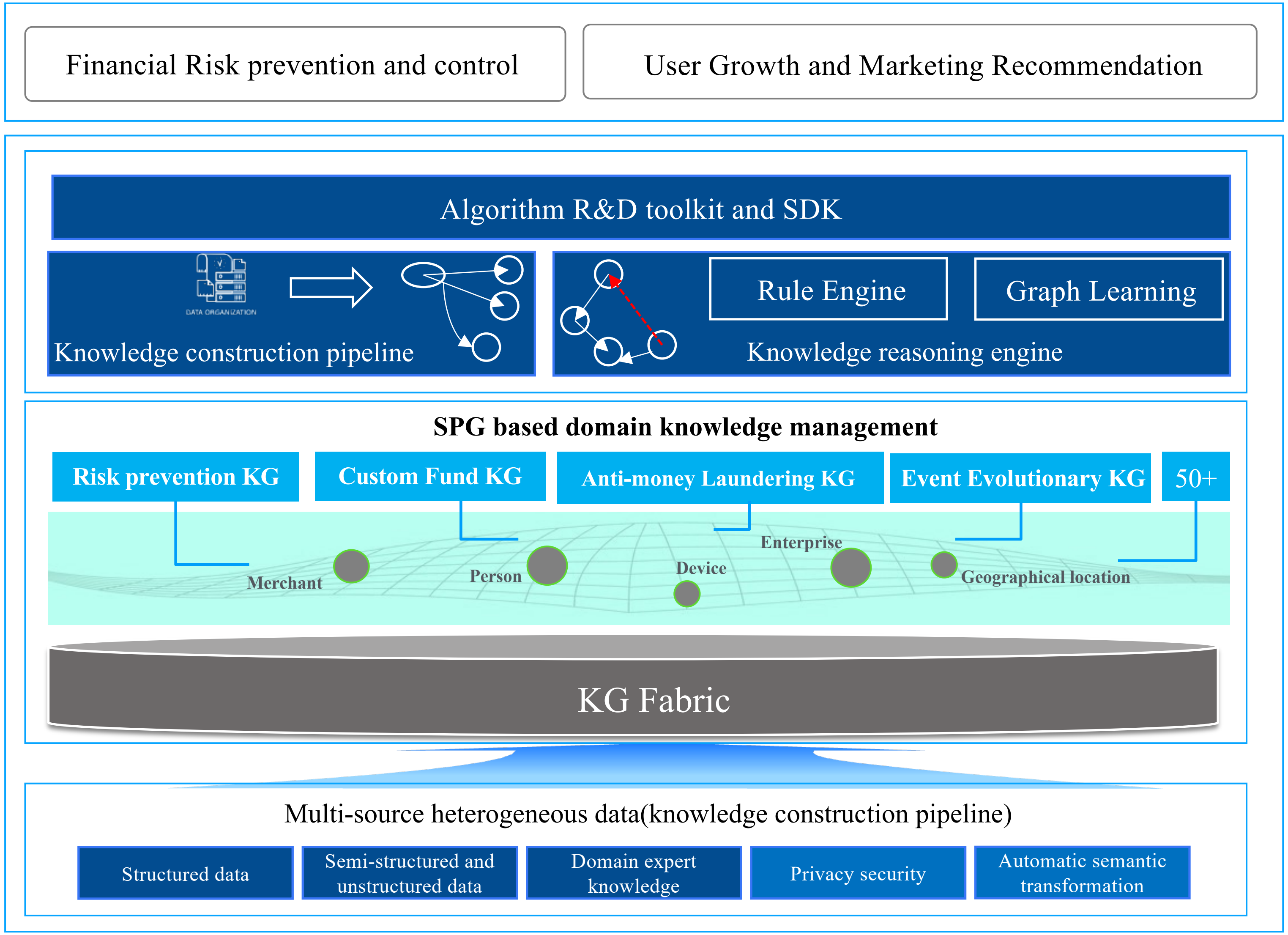
- The PBOC requires completion of the ultimate beneficiary (UBO) information and a reasonable explanation

Issues

- Business data are independent, and their definitions and representations are not unified, they cannot be deeply related.
- The scale of business graph data is large, ranging from tens of billions to hundreds of billions
- Heterogeneous knowledge graph with complex entities, relationships, and properties

- KARI：An encoder-decoder framework that supports heterogeneous graph learning, relying on the graph KGDSL to obtain multi-dimensional subgraph features, such as degree, page rank, neighbor statistics, rules, etc.
- KGDSL +KPRN： Interpretable results based on graph association paths and rule learning.

Ant Knowledge Graph: Novel Knowledge Graph Engine based on SPG



Financial Risk prevention and control

User Growth and Marketing Recommendation

Algorithm R&D toolkit and SDK

Knowledge construction pipeline

Rule Engine

Graph Learning

Knowledge reasoning engine

SPG based domain knowledge management

Risk prevention KG

Custom Fund KG

Anti-money Laundering KG

Event Evolutionary KG

50+

KG Fabric

Multi-source heterogeneous data(knowledge construction pipeline)

Structured data

Semi-structured and unstructured data

Domain expert knowledge

Privacy security

Automatic semantic transformation

First SPG-based KG Engine

Quickly incubate knowledge graphs in new fields

KG Fabric

Linear scaling up to 1.3 trillion with zero-copy and sub-minute latency

Knowledge Reasoning Engine

Rule-guided learning and interpretable reasoning at a scale of 100 billions

Algorithm Framework and SDK

SPG Modeling, KG Construction, KG Reasoning

(python + KGDSL/GQL)

We have been granted 35 patents, filed for over 140 patents, participated in the development of 18 standards, and received recognition such as the Zhejiang Science and Technology Progress Second Prize and the Guiyang Big Data Expo Excellent Achievement Award.

01 **Knowledge Graph：** An effective way to manage domain knowledge

02 **Opportunities and Challenges:** Knowledge management paradigm shift from binary static to multi-dynamic

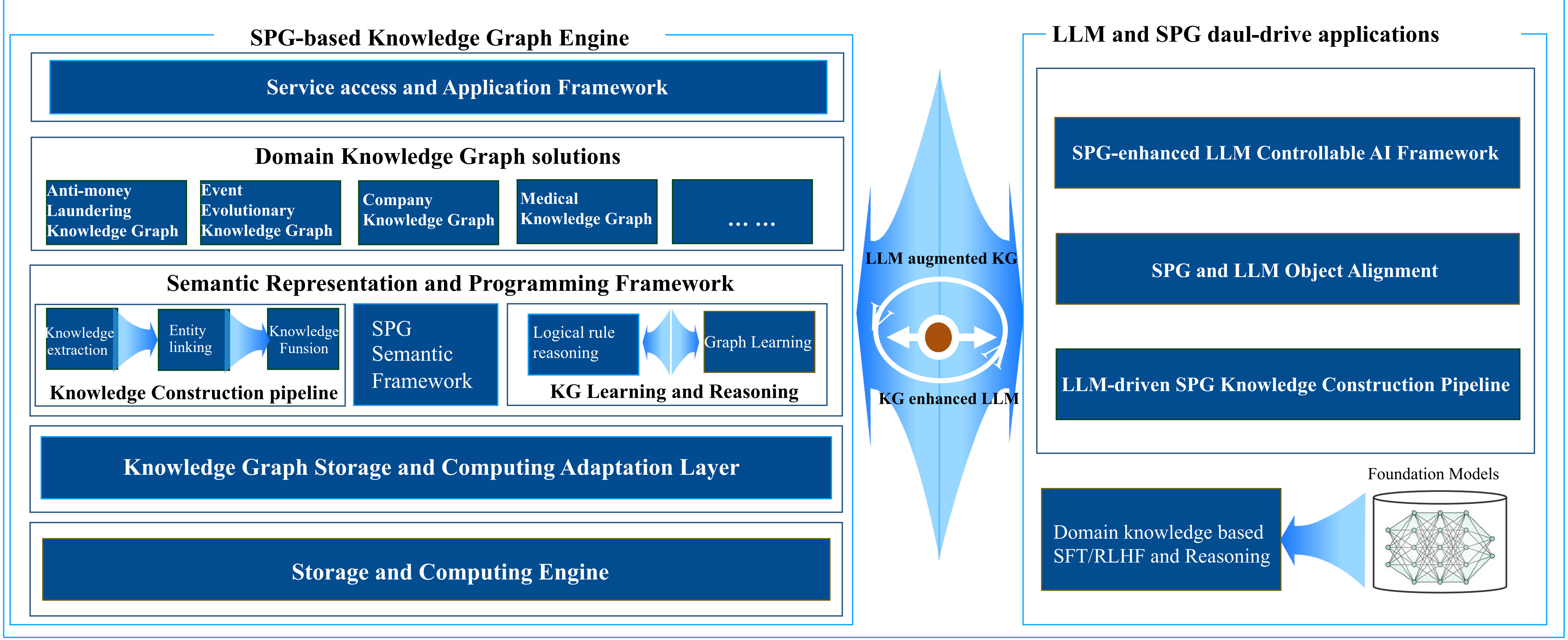
03 **SPG:** A novel semantic framework that accelerates data knowledgeization and knowledge symbolization

04 **Applications:** SPG-based knowledge graph cases

05 **Future prospects:** The application value brought by SPG and its dual-drive development with LLM in the future

SPG and LLM Bidirectionally Driven Controllable AI

Building next-generation industrial-level cognitive engine



SPG-based Knowledge Graph Engine

Service access and Application Framework

Domain Knowledge Graph solutions

- Anti-money Laundering Knowledge Graph
- Event Evolutionary Knowledge Graph
- Company Knowledge Graph
- Medical Knowledge Graph
- … …

Semantic Representation and Programming Framework

Knowledge extraction → Entity linking → Knowledge Funsion

Knowledge Construction pipeline

SPG Semantic Framework

Logical rule reasoning ↔ Graph Learning

KG Learning and Reasoning

Knowledge Graph Storage and Computing Adaptation Layer

Storage and Computing Engine

LLM augmented KG

KG enhanced LLM

LLM and SPG daul-drive applications

SPG-enhanced LLM Controllable AI Framework

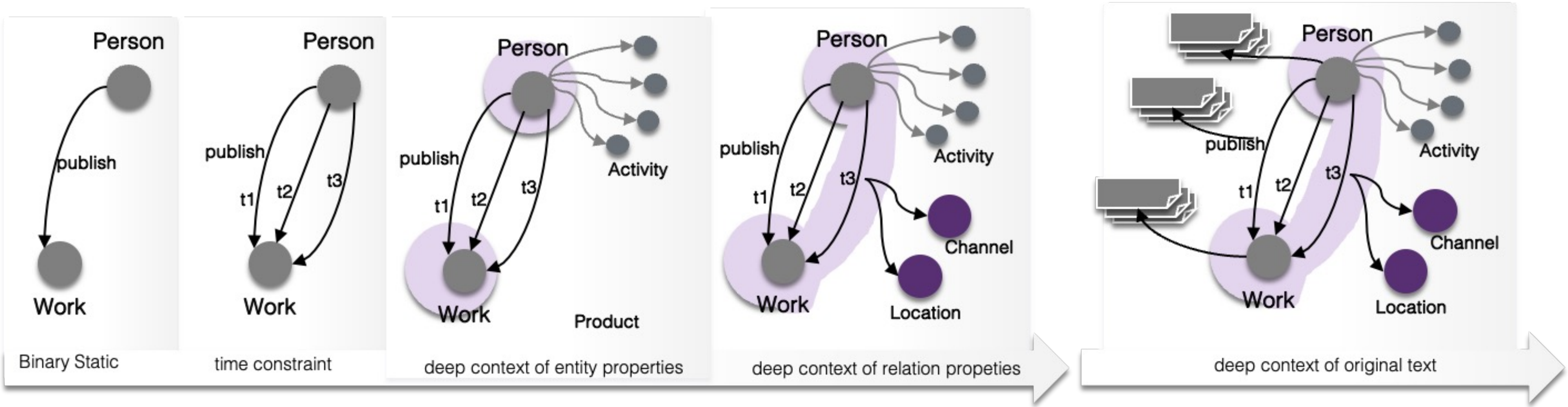
SPG and LLM Object Alignment

LLM-driven SPG Knowledge Construction Pipeline

Foundation Models

Domain knowledge based SFT/RLHF and Reasoning

Knowledge Representation from Big Data Era to the LLM Era



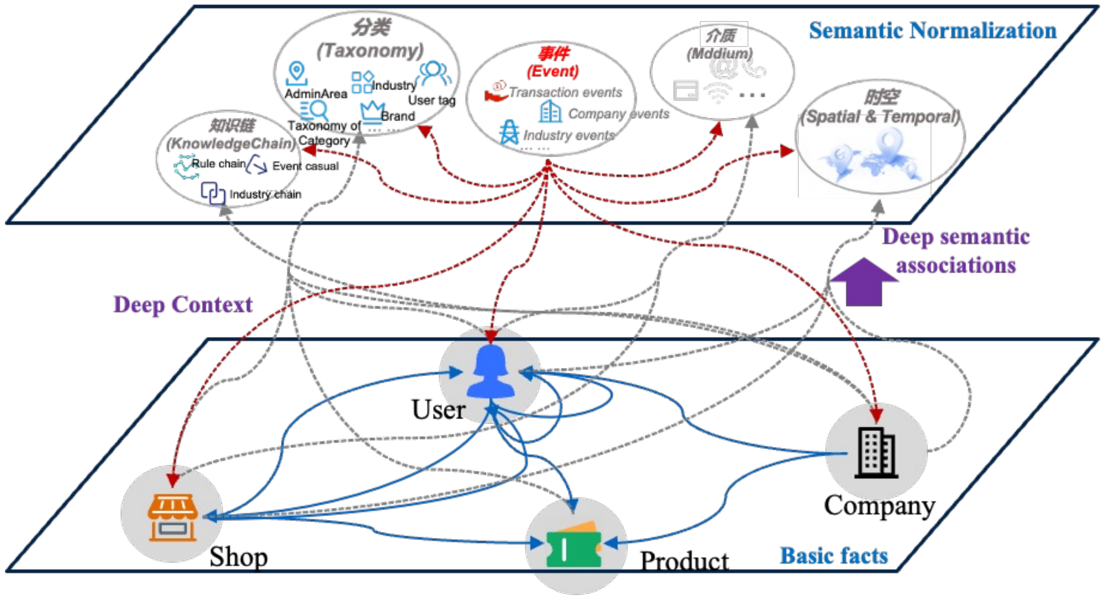
Big Data Era

From Binary-static to multiple-dynamic

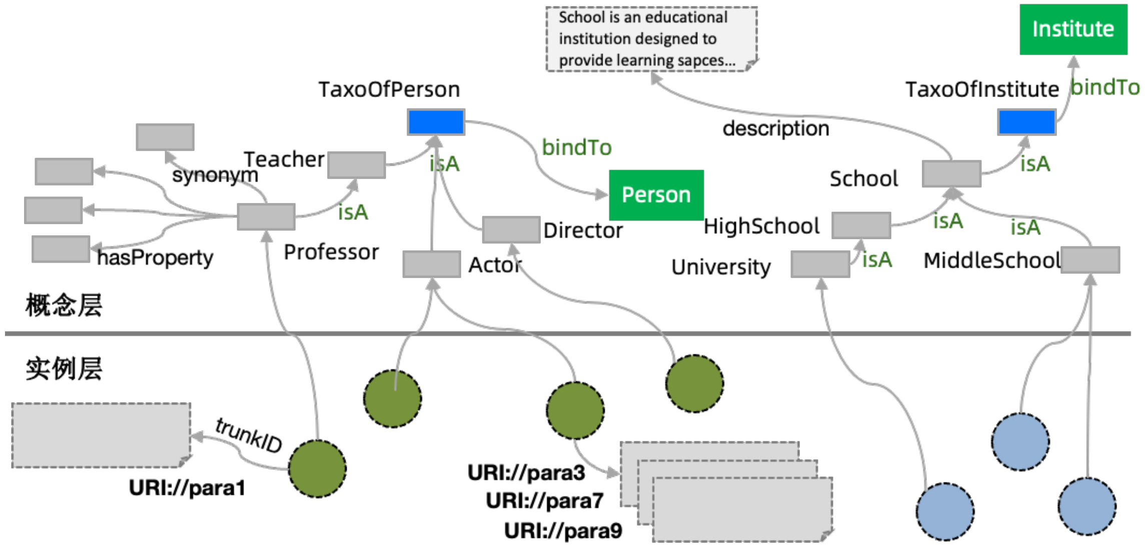
LLM Era

More traceable original text

KGs Need to Continuously Upgrade Semantic Representation



upgrade



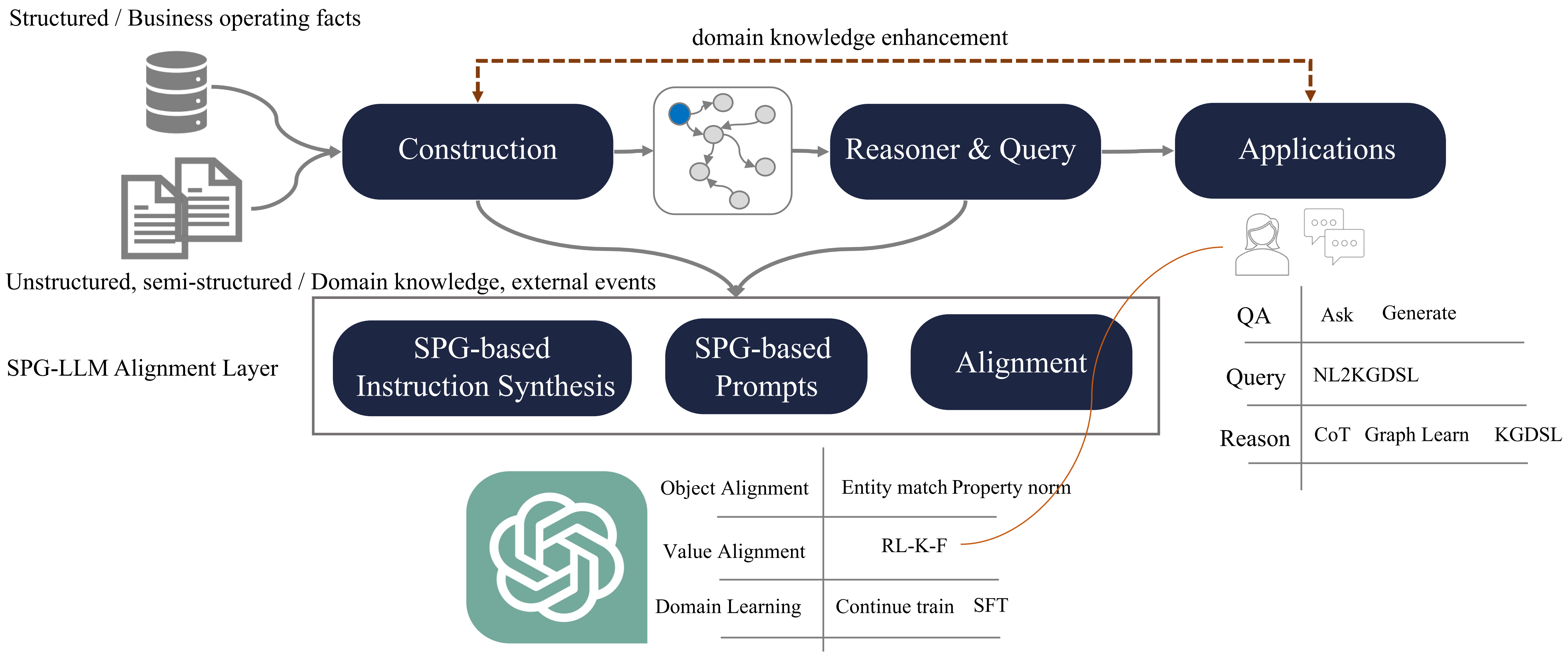
BigData Friendly (Structured Deep Context)

1. Hierarchical and categorical representation of knowledge
2. Logical and programmable reasoning between concepts
3. Integration of logical rules and factual knowledge

LLM Friendly (Raw text Deep Context)

1. Hierarchical representation of instances and concepts
2. Align LLM with instance through concept layer
3. Mutual indexing representation of original text and structured knowledge

Build an AI framework based on the OpenSPG Knowledge Engine



Structured / Business operating facts

Unstructured, semi-structured / Domain knowledge, external events

domain knowledge enhancement

Construction

Reasoner & Query

Applications

SPG-LLM Alignment Layer

SPG-based Instruction Synthesis

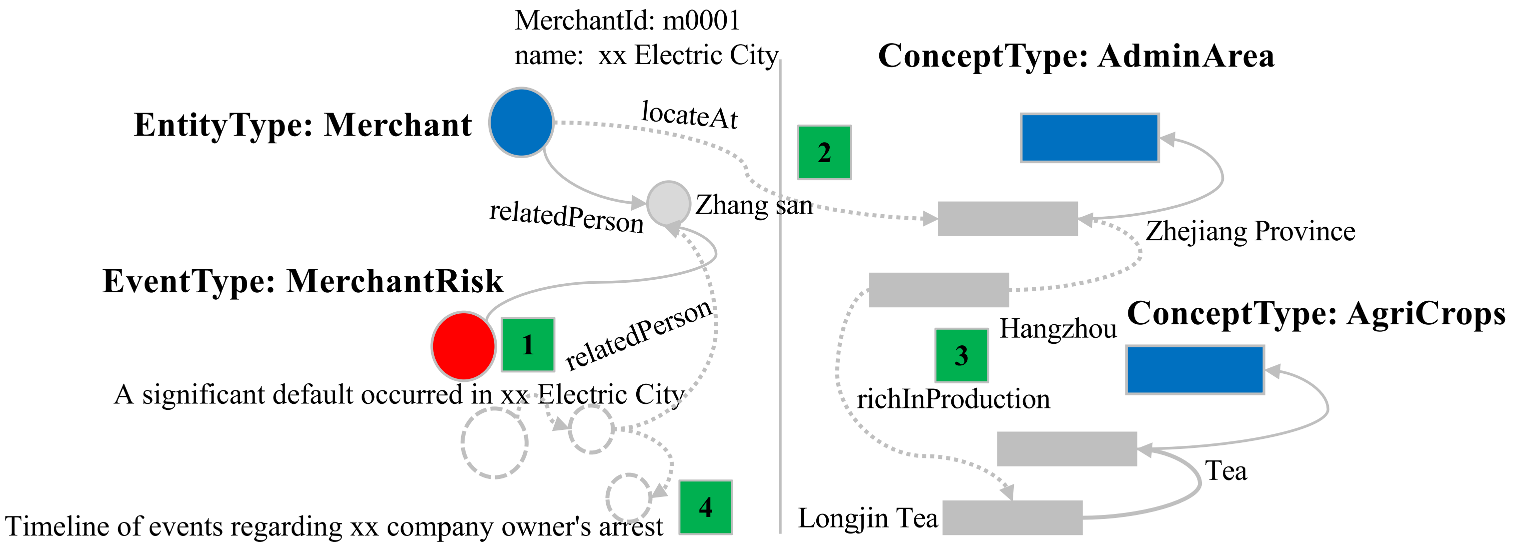
SPG-based Prompts

Alignment

QA	Ask Generate
Query	NL2KGDSL
Reason	CoT Graph Learn KGDSL

Object Alignment	Entity match Property norm
Value Alignment	RL-K-F
Domain Learning	Continue train SFT

LLM4SPG Enhanced Knowledge Construction



MerchantId: m0001
name: xx Electric City

EntityType: Merchant

locateAt

relatedPerson

Zhang san

EventType: MerchantRisk

1

relatedPerson

A significant default occurred in xx Electric City

4

Timeline of events regarding xx company owner's arrest

ConceptType: AdminArea

2

Zhejiang Province

Hangzhou

3

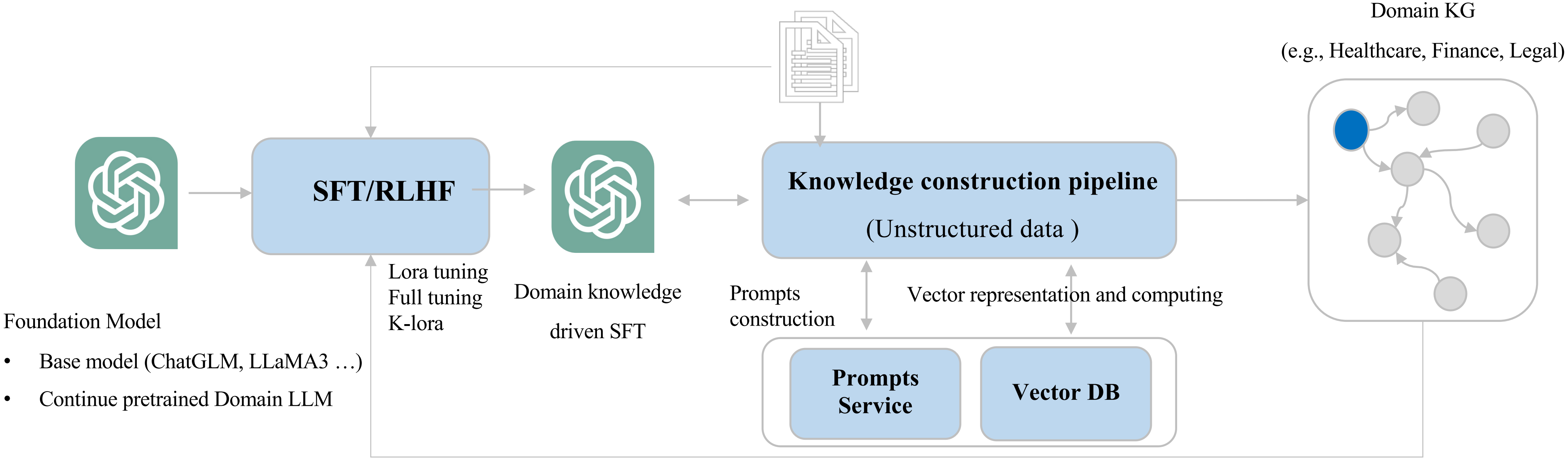
richInProduction

ConceptType: AgriCrops

Tea

Longjin Tea

1	Knowledge extraction (entity/event/relation/property)
2	Entity linking/property normalization
3	Commonsense knowledge completion
4	Document-level subgraph extraction



SFT/RLHF

Lora tuning
Full tuning
K-lora

Domain knowledge
driven SFT

Knowledge construction pipeline
(Unstructured data)

Prompts
construction

Vector representation and computing

Prompts
Service

Vector DB

Domain KG
(e.g., Healthcare, Finance, Legal)

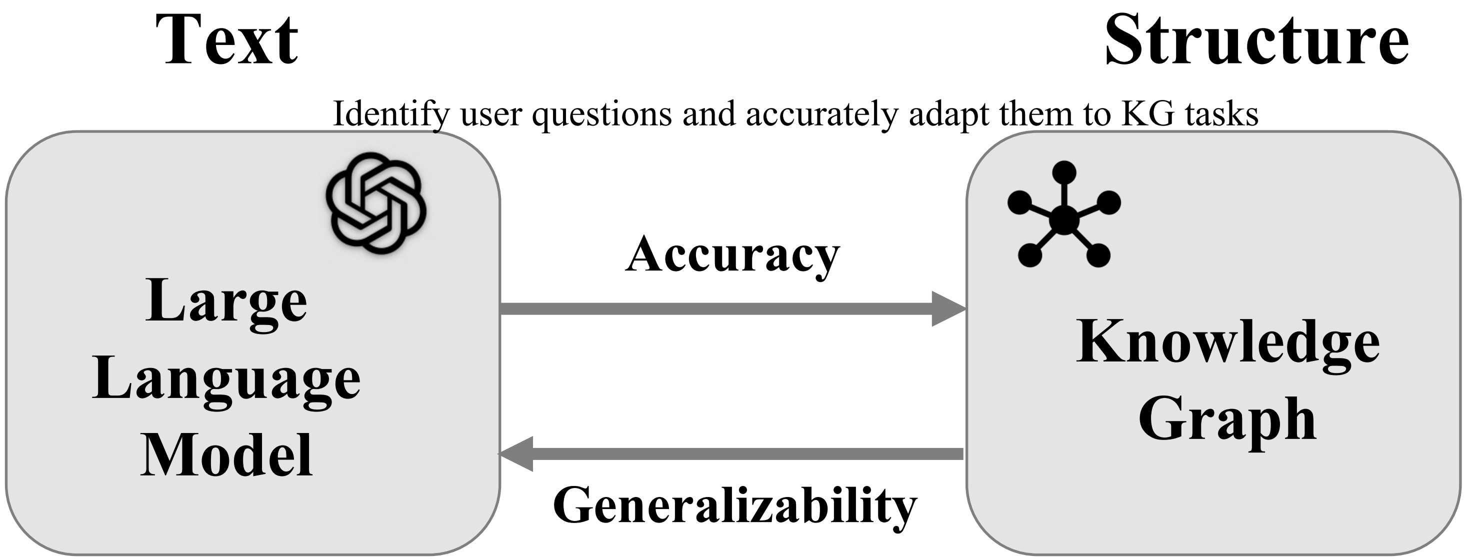
Foundation Model

- Base model (ChatGLM, LLaMA3 …)
- Continue pretrained Domain LLM

KGs are Better Instruction Synthesizers for LLMs

KG as a guide: synthesize professional corpus, guide LLM to more accurately identify KG tasks through SFT/RLHF

KG as an executor: execute KG tasks identified by LLM to complete knowledge query and reasoning



Text

Structure

Identify user questions and accurately adapt them to KG tasks

Large Language Model

Accuracy

Generalizability

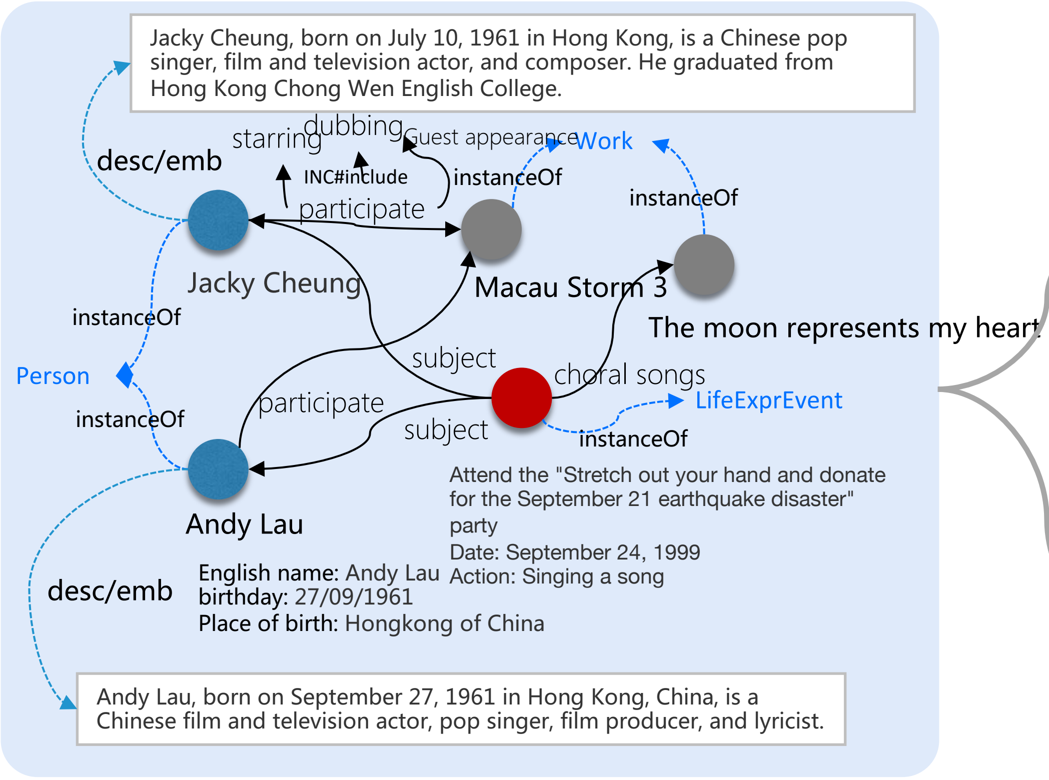
Knowledge Graph

Relying on a large-scale KG with precise semantics, rigorous logic, and complete context, it can construct complex, diverse, and large-scale instruction data

Instruction synthesis injects knowledge so that LLM can understand the KG task structure

KGs need to enhance knowledge semantic representation to complete missing context

Instruction Synthesis based on SPG-based KGs

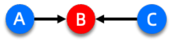


Jacky Cheung, born on July 10, 1961 in Hong Kong, is a Chinese pop singer, film and television actor, and composer. He graduated from Hong Kong Chong Wen English College.

Attend the "Stretch out your hand and donate for the September 21 earthquake disaster" party
Date: September 24, 1999
Action: Singing a song

English name: Andy Lau
birthday: 27/09/1961
Place of birth: Hongkong of China

Andy Lau, born on September 27, 1961 in Hong Kong, China, is a Chinese film and television actor, pop singer, film producer, and lyricist.

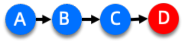


Andy Lau-**participate**->(**Movie**)<-**participate**-(Jacky Cheung)

[Constructed Question] What movies have Andy Lau and Jacky Cheung participated in together?

[CoT]

1. Determine which films Andy Lau has appeared in;
2. Determine which films Jacky Cheung has appeared in;
3. Find out the common works of the two people in the film



The Macau Storm 3-**starring**->(Actor)-**participate**->(**Movie**)-**producer**->(Institution)

[Constructed Question] Which film and television productions have the starring actors of Macau Storm 3 participated in?

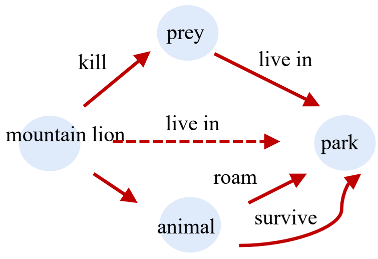
[CoT]

1. Determine who are the leading actors in Macau Storm 3
2. Determine the film and television works in which these actors have participated
3. Determine the producers of these film and television works

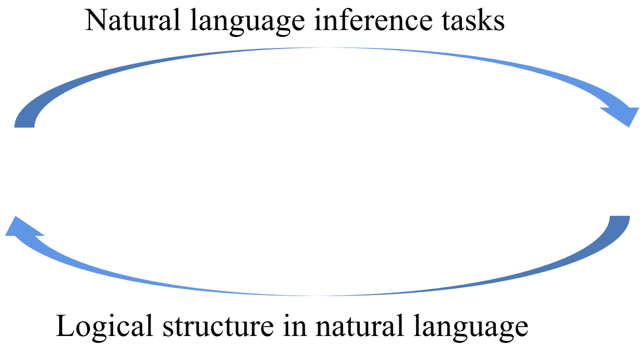
Person(人物): EntityType
desc: Social roles, individuals with unique personalities, experiences and influence
properties:
name(姓名) : Text
enName(英文名) : Text
placeOfBorn(出生地): Text
dateOfBorn(出生日期): Date

```
{
    “instruction”: “You are an expert in entity knowledge extraction for  knowledge graphs... ...。",
    "schema": [
        {
            "entity_type": ”Person",
            "attributes": [
                ”name“, ”english name“, ”birthday“,”place of birth“]
        }
    ],
    "input": "刘德华（Andy Lau），... ...。”
}
```

Synthetic Reasoning Instructions



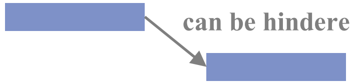
Inductive link prediction


Natural language inference tasks
Logical structure in natural language

context:
mountain lion -> kill -> prey ~ prey -> live in -> park
mountain lion -> be classify as -> animal ~ animal -> roam -> park
mountain lion -> be classify as -> animal ~ animal -> survive -> park
statement: mountain lion -> live in -> park

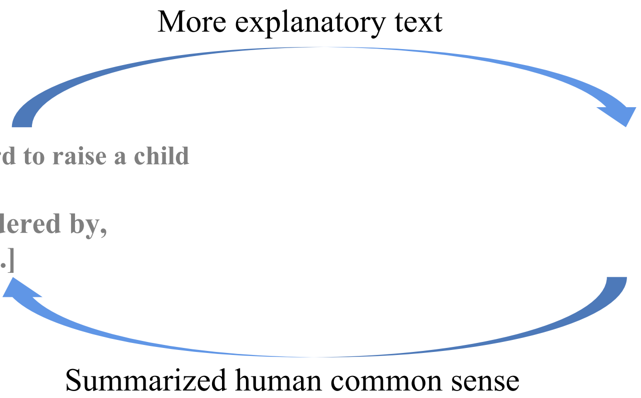
Inductive Process:
1） Mountain lions are known to be predators that kill the prey. The prey, in turn, lives in parks. This correlation suggests that mountain lions may also live in parks, as they are likely to follow their prey to their habitats.
2） Additionally, mountain lions are classified as animals, and animals are known to roam freely in parks. This roaming behavior is essential for their survival, which implies that mountain lions may also live in parks to ensure their survival.
3） …

In summary, the evidence suggests that mountain lions live in parks due to their predatory behavior, roaming habits, and need for survival.



[PersonX adopts a baby, can be hindered by, PersonX can't afford to raise a child.]

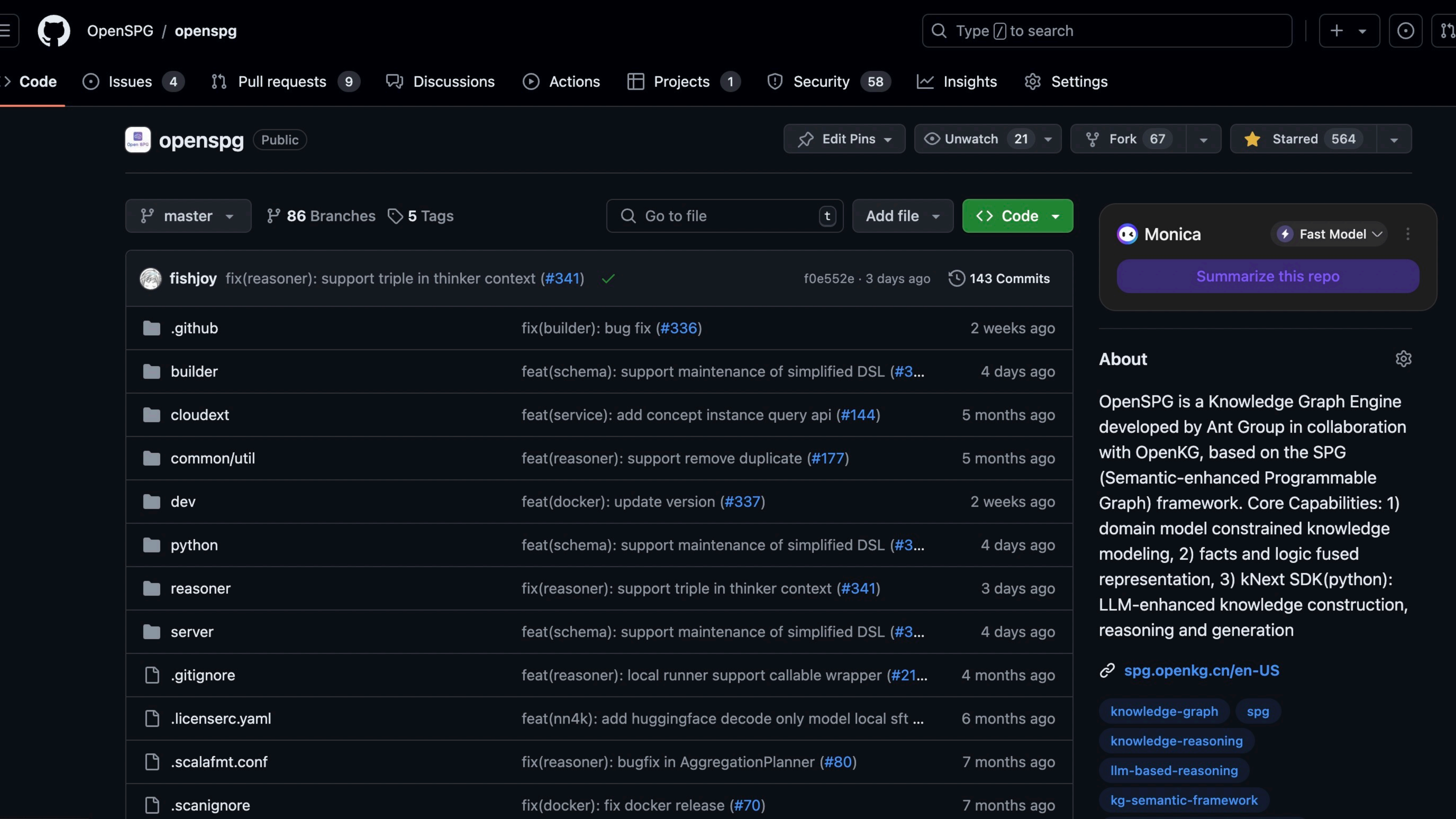
Commonsense reasoning



[Sentence1: After years of saving, Sarah finally adopts a baby, bringing immense joy to her once quiet home.]

[Sentence2: If Sarah hadn't been able to afford the expenses, the dream of filling her life with a child's laughter would have remained unfulfilled.]

Combined with the mutual indexing structure of text and structure, a large amount of inductive, deductive, and abductive reasoning corpus can be synthesized. In addition, there are more than 30 conceptual semantics such as succession, hypernym, inclusion, hindered by, etc.



fishjoy fix(reasoner): support triple in thinker context (#341) ✓ f0e552e · 3 days ago 143 Commits

.github	fix(builder): bug fix (#336)	2 weeks ago
builder	feat(schema): support maintenance of simplified DSL (#3...	4 days ago
cloudext	feat(service): add concept instance query api (#144)	5 months ago
common/util	feat(reasoner): support remove duplicate (#177)	5 months ago
dev	feat(docker): update version (#337)	2 weeks ago
python	feat(schema): support maintenance of simplified DSL (#3...	4 days ago
reasoner	fix(reasoner): support triple in thinker context (#341)	3 days ago
server	feat(schema): support maintenance of simplified DSL (#3...	4 days ago
.gitignore	feat(reasoner): local runner support callable wrapper (#21...	4 months ago
.licenserc.yaml	feat(nn4k): add huggingface decode only model local sft ...	6 months ago
.scalafmt.conf	fix(reasoner): bugfix in AggregationPlanner (#80)	7 months ago
.scanignore	fix(docker): fix docker release (#70)	7 months ago

About

OpenSPG is a Knowledge Graph Engine developed by Ant Group in collaboration with OpenKG, based on the SPG (Semantic-enhanced Programmable Graph) framework. Core Capabilities: 1) domain model constrained knowledge modeling, 2) facts and logic fused representation, 3) kNext SDK(python): LLM-enhanced knowledge construction, reasoning and generation

spg.openkg.cn/en-US

knowledge-graph spg knowledge-reasoning llm-based-reasoning kg-semantic-framework

Thank you

OpenKG.CN
开放的中文知识图谱

OpenSPG Github

https://github.com/OpenSPG/openspg