

AI 驱动 软件研发 全面进入数字化时代

中国·北京 08.18-19

AI+
software
Development
Digital
summit



工业级知识语义框架及引擎架构

梁磊 蚂蚁集团

科技生态圈峰会 + 深度研习 ——1000+ 技术团队的选择



2023K+
全球软件研发行业创新峰会
上海站

会议时间 | 06.09-10



2023K+
全球软件研发行业创新峰会
北京站

会议时间 | 07.21-22



2024K+
全球软件研发行业创新峰会
深圳站

会议时间 | 05.17-18



K+峰会详情



会议时间 | 08.18-19
AiDD AI+软件研发数字峰会
北京站



会议时间 | 11.17-18
AiDD AI+软件研发数字峰会
深圳站



AiDD峰会详情

▶ 演讲嘉宾



梁磊

蚂蚁集团技术总监/资深专家

蚂蚁知识引擎负责人，个人主要技术方向为知识图谱、搜索推荐引擎及AI工程等，于2018年开始主导蚂蚁知识图谱的建设，基于蚂蚁多样性的金融业务场景构建了企业级知识图谱引擎架构，平台累计提报140+件专利，10余项软件著作权，主导支撑孵化的项目先后获得BU总裁特别奖、数据科学奖、优秀成果奖等，平台能力通过了CESI测评认证，目前也在主导IEEE 2807.2金融知识图谱标准化、SPG知识图谱语义标准化等工作。

目录

CONTENTS

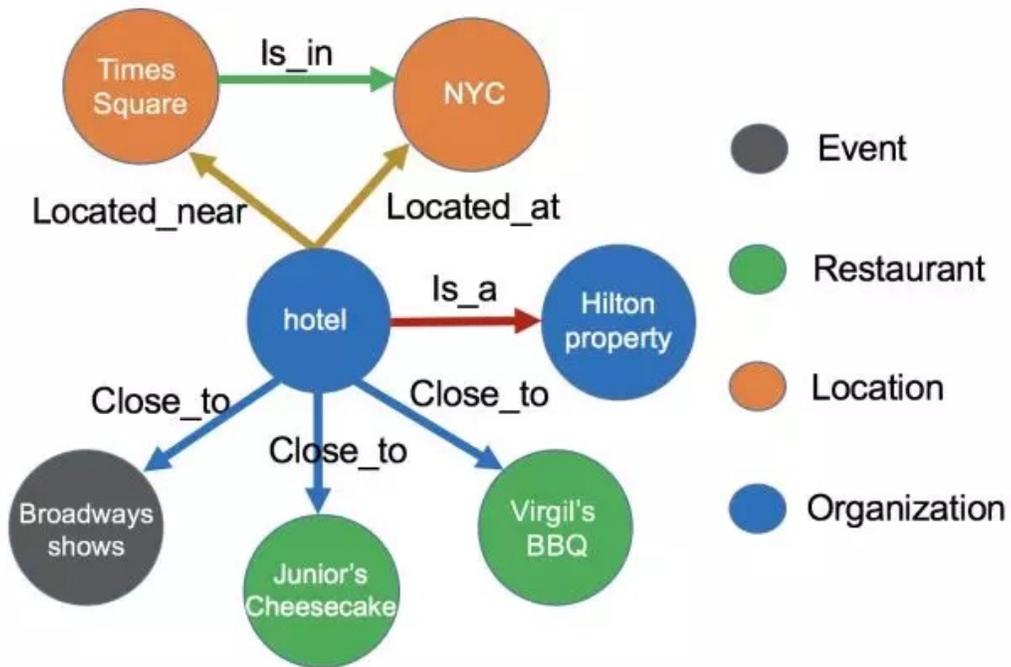
1. 什么是知识图谱
2. 图谱技术发展的机遇与挑战
3. SPG(Semantic-enhanced Programmable Graph)
4. 蚂蚁知识图谱引擎

PART 01

1. 什么是知识图谱

知识图谱的主要形态 - 文档要素知识化

This hotel is my favorite Hilton property in NYC! It is located right on 42nd street near Times Square in New York, it is close to all subways, Broadways shows and next to great restaurants like Junior's Cheesecake, Virgil's BBQ



材料来自公开资料

通用知识图谱的构建方式，通过理解文档中的核心要素(NER 命名实体识别)、关系谓词抽取(SPO三元组)来构建要素之间语义关联

知识图谱的主要形态 – 搜索引擎体验提效

窦靖童的爸爸的前妻的前夫是谁？

知识推理

结合性别推理
结合时间推理
结合路径推理

窦唯 / 前夫

李亚鹏

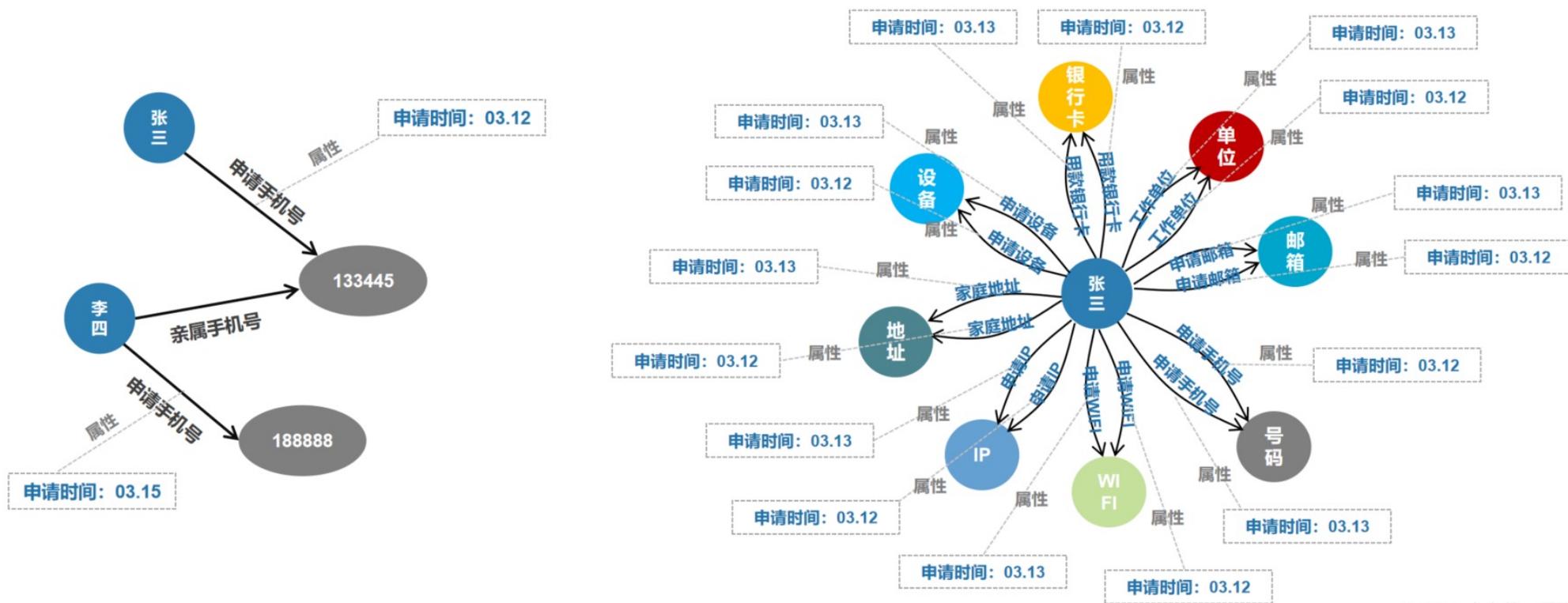
知乎 @三和厂妹



Person(x1) ^ Person(x2) ^ 前妻(x1, x2) -> 前夫(x2, x1)

Person(x1) ^ Person(x2) ^ 女儿(x1, x2) ^ 性别(x2, Male) -> 爸爸(x2, x1)

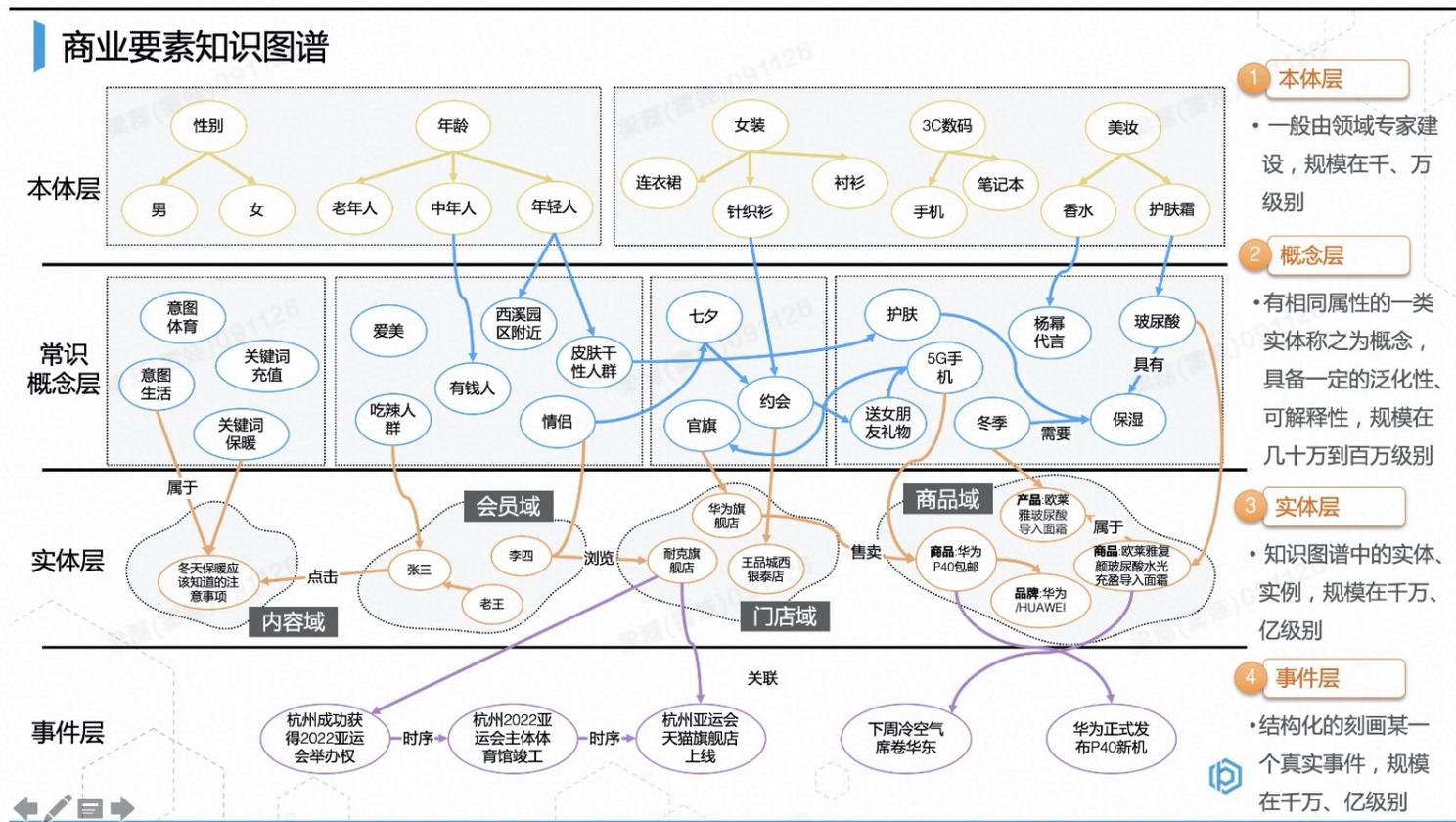
知识图谱的主要形态 – 应用于安全风险洞察



CSDN @白白的一团团

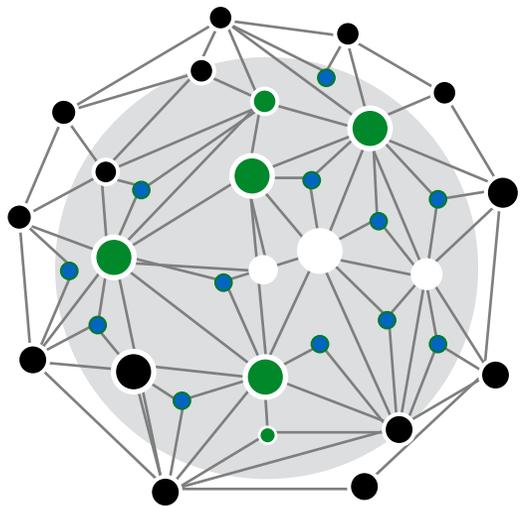
知识图谱在风控中的应用，构建以“人”为中心的上下文介质网络，辅助发现风险线索

知识图谱的主要形态 – 应用于企业知识管理



基于知识图谱实现知识管理，构建事件、实体、常识概念的分层

▶ 知识图谱：语义、高阶、多元、链式知识化数据管理方案

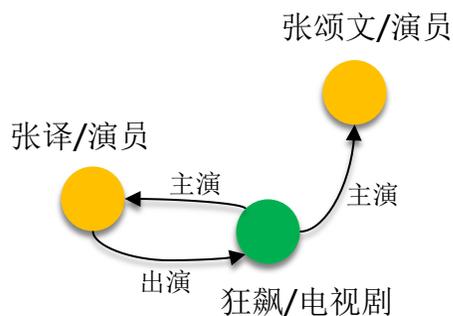


定义

建模世界万物的关联关系

它是一种建模世界的方法

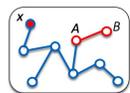
实现数据的知识标准化、语义互连



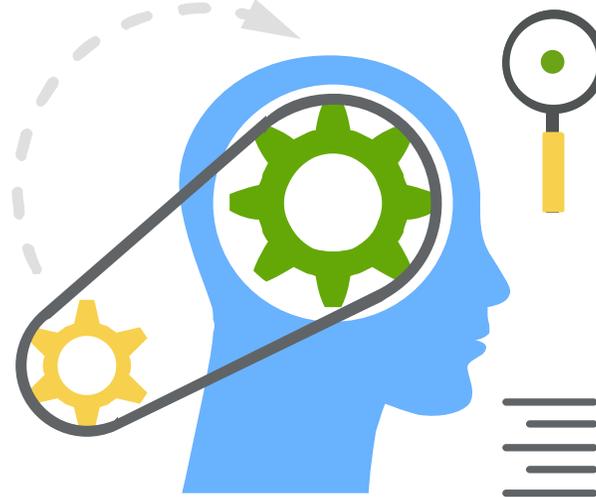
思考



文本：一维



图谱：二维、多元...

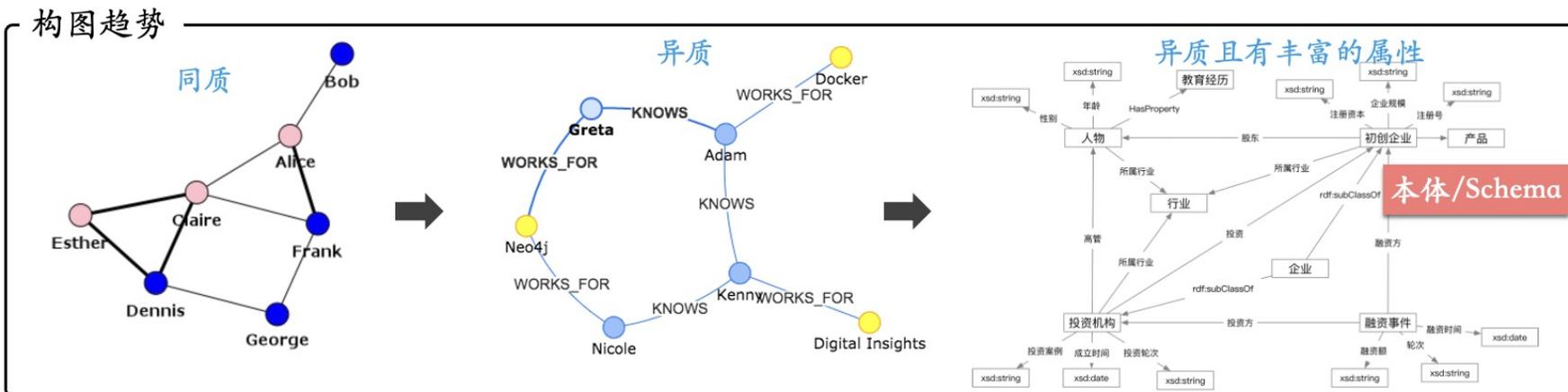


终态

人工智能的大脑

材料来自公开资料

知识图谱：语义、高阶、多元、链式知识化数据管理方案



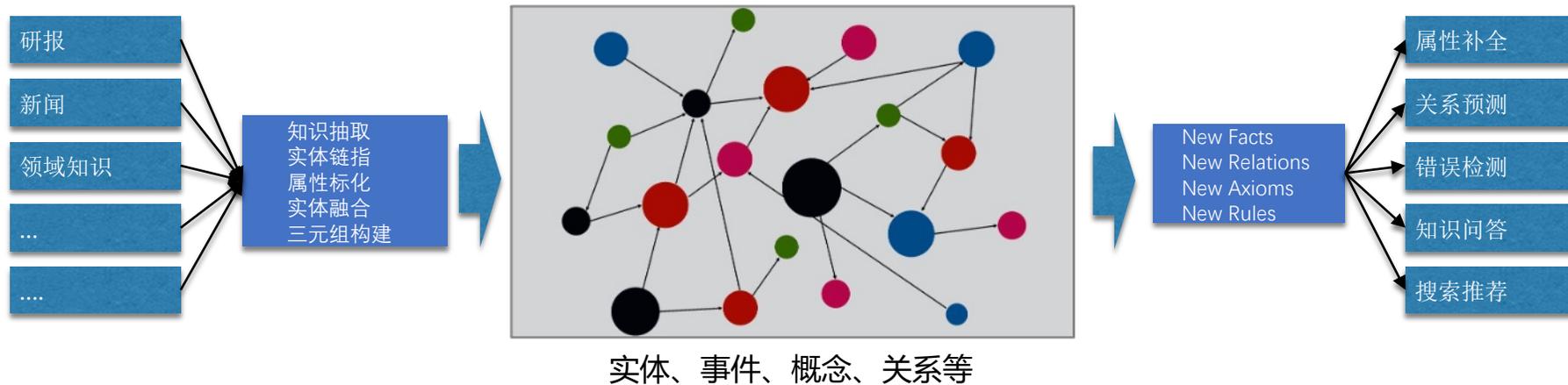
图谱算法特点

- Knowledge Graph = Graph + Knowledge Base
 - Graph: 学习 **Structure**
 - Knowledge Base: 学习 **Semantics**
- 异质** 信息融合学习
 - 存在不同类型的节点/边
 - 节点/边可以有丰富的属性
- 依赖 **NLP** 及 **图** 算法能力
- 依赖 **专家**, 更体系化
 - 知识建模、抽取、表示、推理等

领域知识图谱生命周期

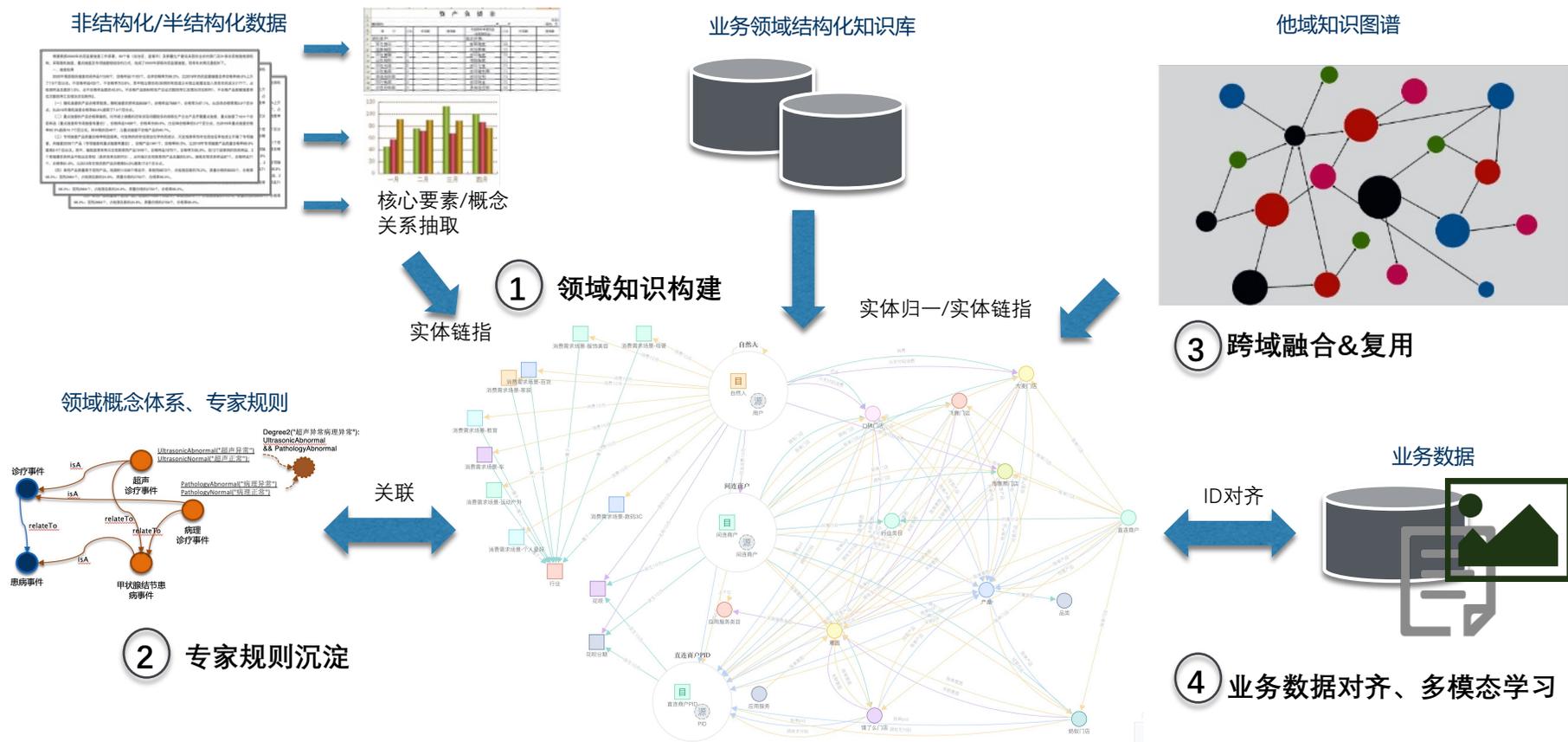


知识图谱：整体链路和构建方式



- **领域知识沉淀**：基于图谱知识语义、图结构实现知识表示和互联，是实现领域知识积累的有效方法
- **知识标准化**：利用知识图谱相关技术不断提升实体、概念、关系、事件等的标准化和归一化水平
- **知识融合&复用**：构建领域知识图谱，通过融合、推理等服务多下游任务，为业务降本提效
- **知识推理发现**：基于图谱推理洞察发现更多稀薄知识，服务风控、信贷、理赔、商家运营、营销推荐等场景

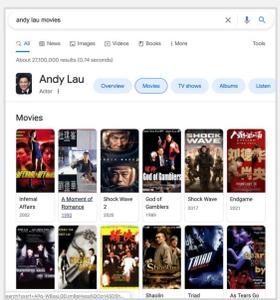
知识图谱：整体链路和构建方式



PART 02

2. 图谱技术发展的机遇与挑战

知识图谱技术发展的机遇与挑战



通用知识图谱(2012)

以google、百度为代表

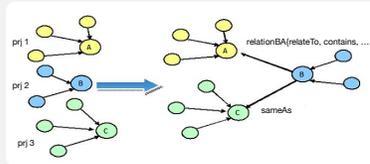


领域知识图谱(2018)

企业数字化升级

艾瑞咨询《2022中国知识图谱市场行业报告》，市场空间：2021年/107亿元，2026年/290亿元

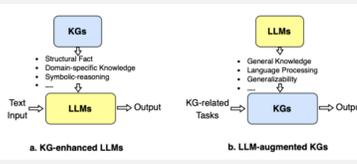
金融、公安等为主要拉动



基于图谱的知识互联

Data Fabric

企业级知识图谱



LLMs + KG

《Unifying Large Language Models and Knowledge Graphs》

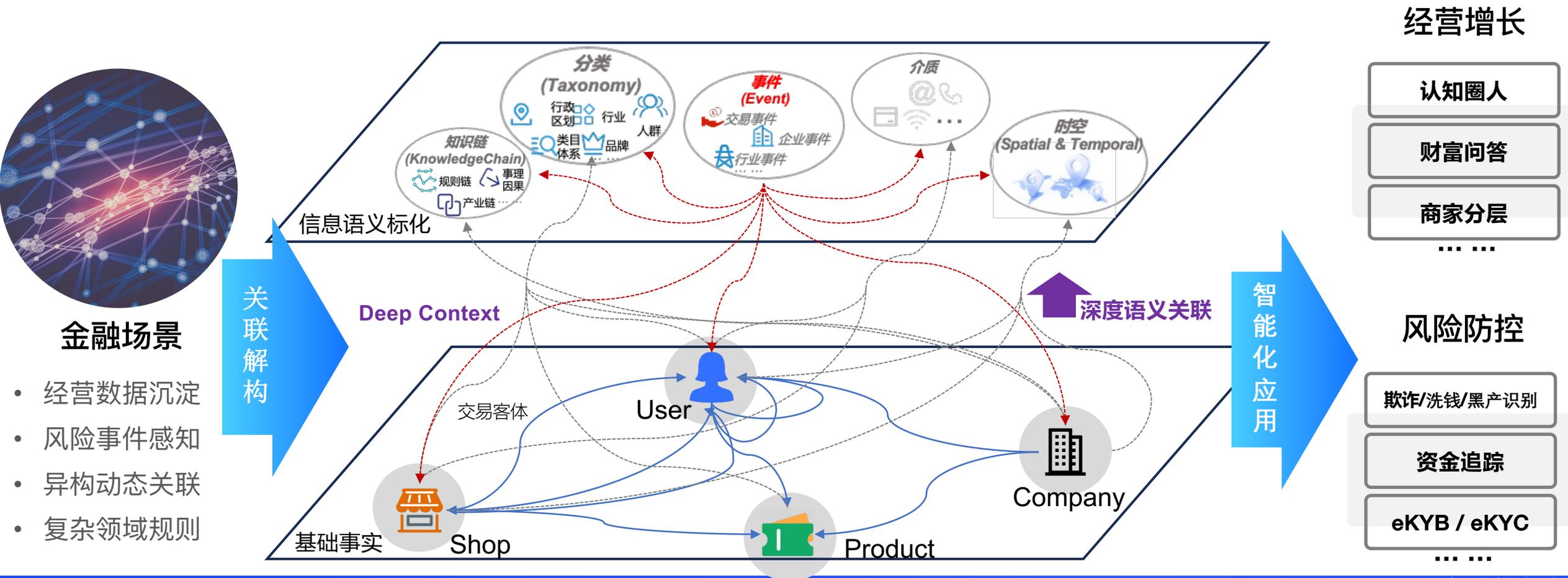
大模型 + 知识图谱双驱动

超越通用知识图谱，推动知识驱动的企业数字化升级

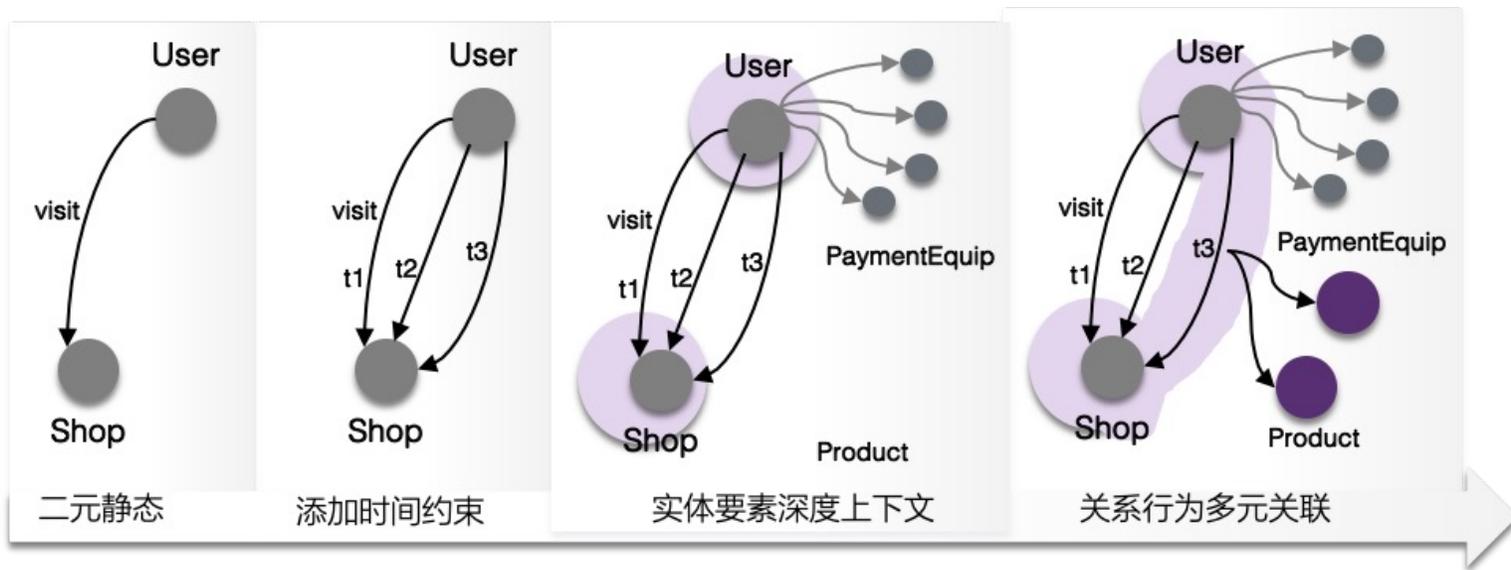
广全、静态常识、正确率容忍、头部覆盖

深精、动态时空、可解释要求、薄客洞察

以蚂蚁金融场景图谱应用为例 - 从静态常识到Deep Context语义关联的需求跃迁



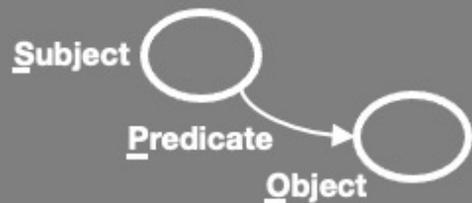
知识管理模式从二元静态到时空多元的模式跃迁



- 常识知识图谱仅使用概念层归纳，无法感知个体差异，无法实现面向个体推理和判断
- 企业级知识管理需要具备较强上下文感知能力，以实现稀薄客群的经营理解和风险洞察
- 知识增强的大模型也要求知识图谱有更深的领域常识和事实实体、事件的覆盖

▶ 知识图谱技术发展的机遇与挑战

缺少统一的知识建模方法



RDF: 用于知识交换, 强语义, 高门槛, 工业落地少

LPG: 用于图存储与查询, 弱语义, 低门槛, 工业应用多

知识构建/获取成本



知识抽取缺少统一范式

专家经验复杂且缺乏复用

缺少统一的技术框架



建模、构建、存储、推理链路长, 工具多

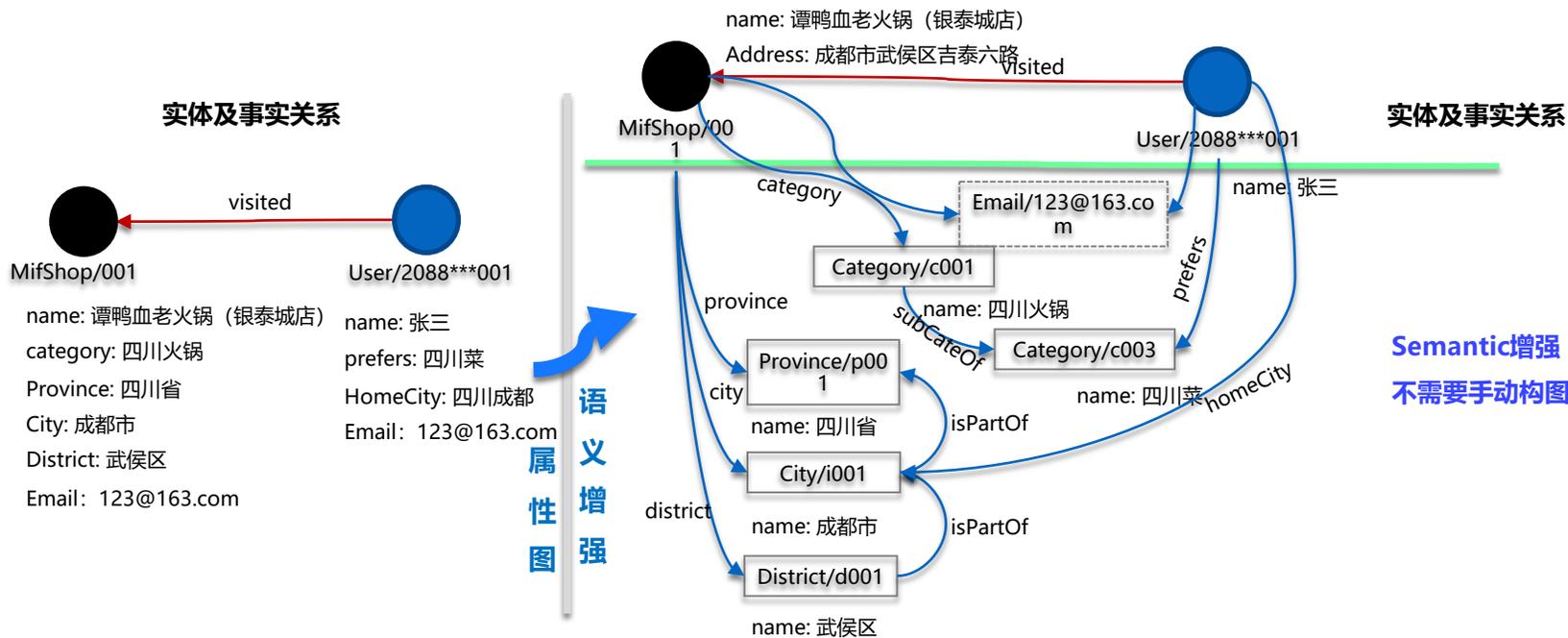
跨场景迁移成本高

知识图谱自身技术体系发展需与时俱进

PART 03

3.SP(G(Semantic-enhanced Programmable Graph))

SPG: Semantic-enhanced Programmable Graph(语义增强示意)



Things, not Strings

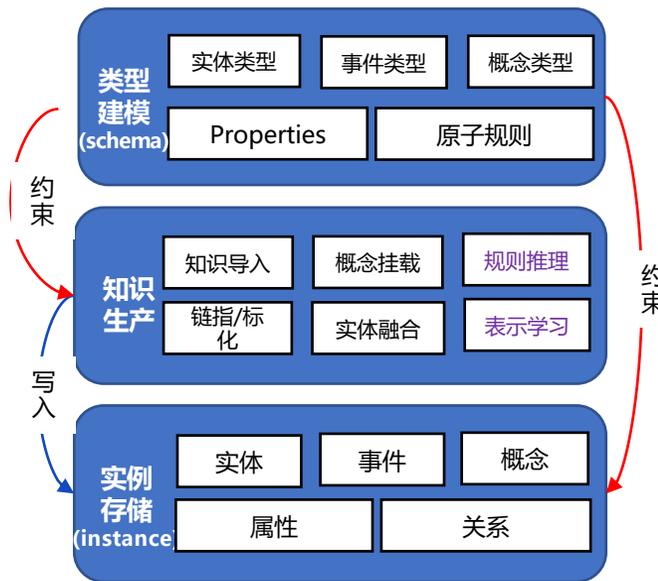
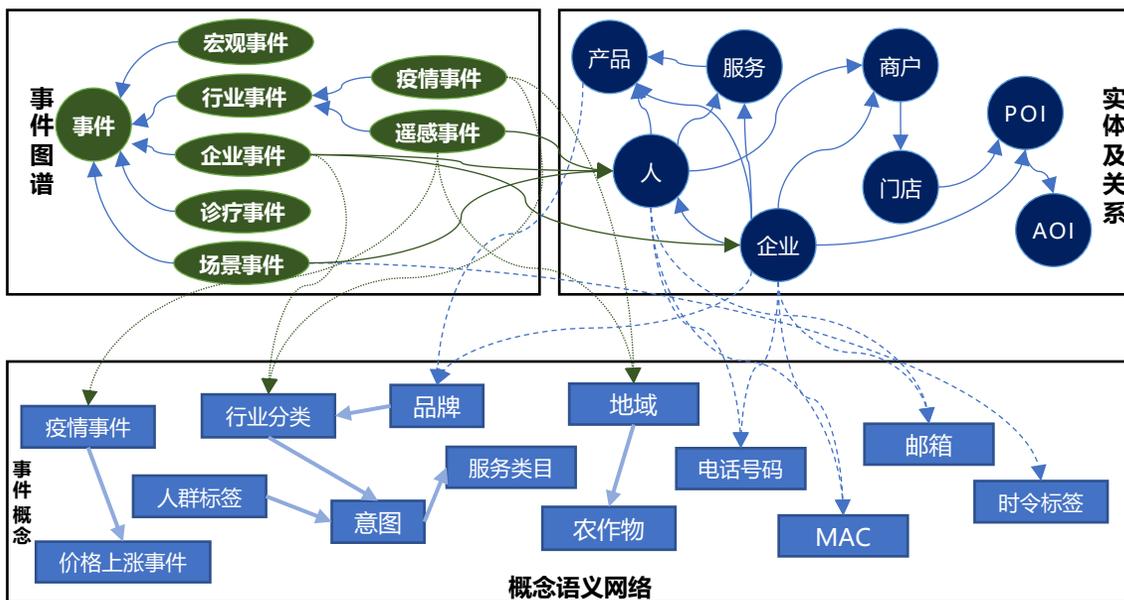
知识的三个显著特点:

- 1、必须有明确的领域类型(every Thing has a Class)
- 2、每个实例类型内必唯一(each instance is unique within an Entity Class)
- 3、语义明确的谓词修饰(nothing exists in isolation)

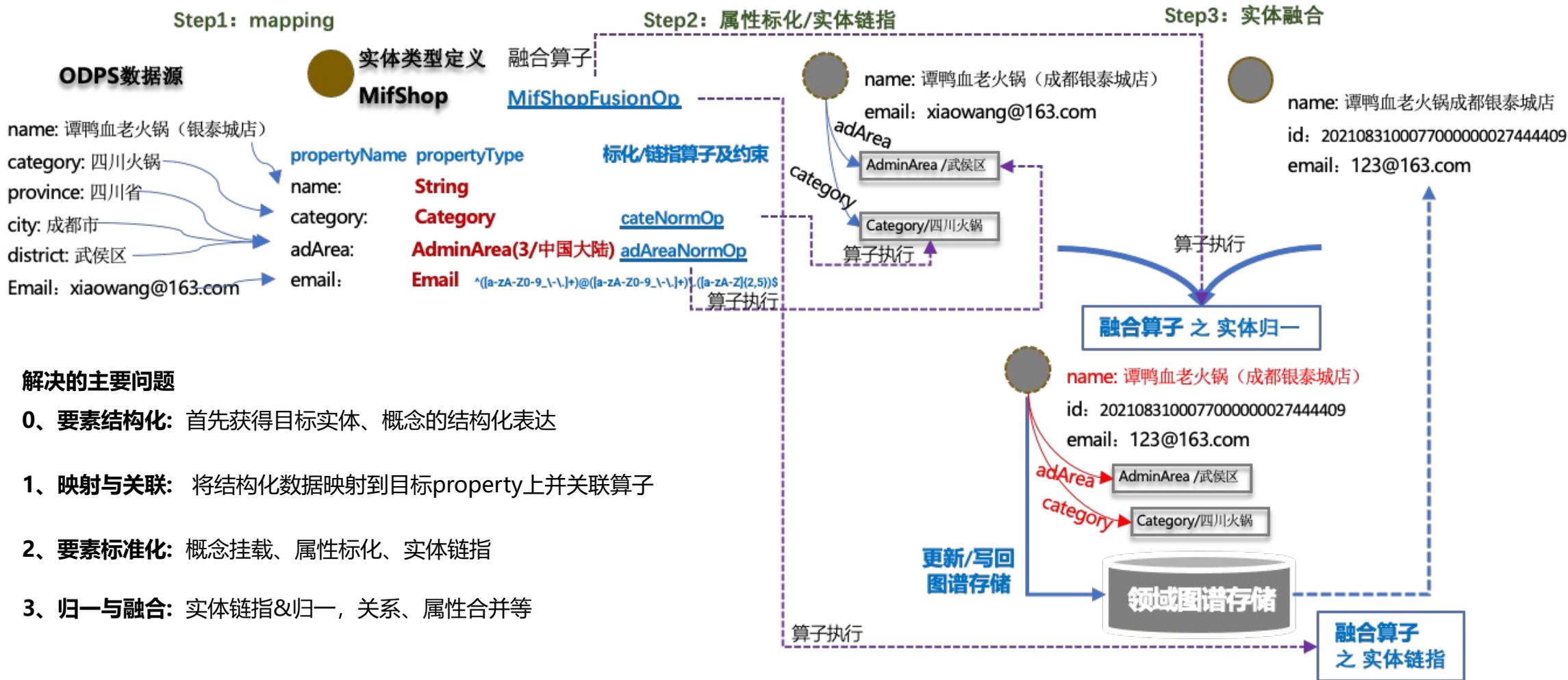
SPG: 主体知识分类模型(Class-Instance Paradigm)

业界主流的划分为实体、概念，但使用阶段并无清晰界限，我们对知识类型的定义：

- **实体**：业务相关性比较强的客观实例，通过实体Properties(属性、关系)刻画个体画像，如用户、企业、商户等
- **概念**：实体从具体到一般的抽象，表述的是一组实体集合。相对静态、具有较强复用性，如人群标签、领域标准类型、语义词汇(如HowNet)等
- **事件**：加入时间、空间、标的等约束的实体类型，如通过NLP、CV等抽取出来的行业事件、企业事件、诊疗事件等



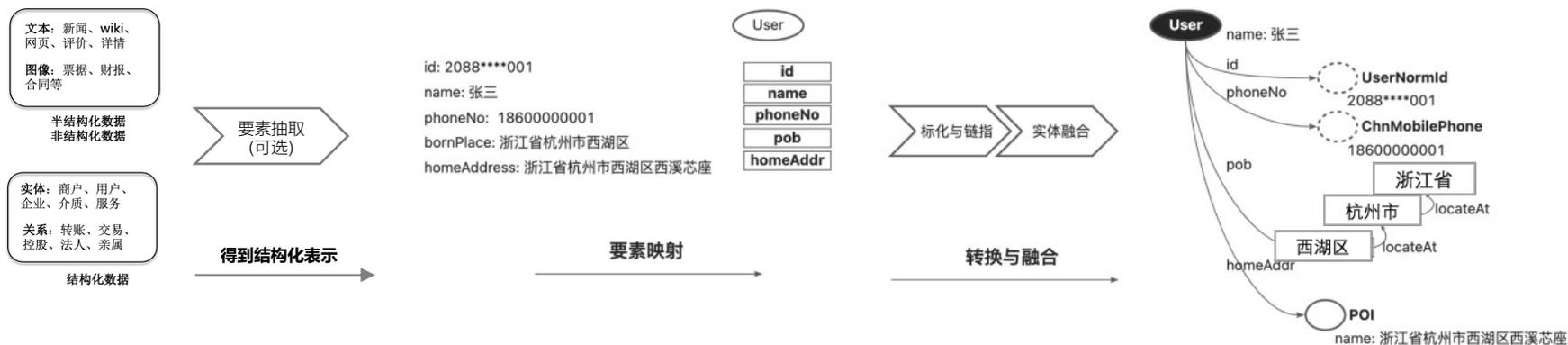
SPG: 非完备数据集下图谱的构建/融合(Programmable)



解决的主要问题

- 0、要素结构化: 首先获得目标实体、概念的结构化表达
- 1、映射与关联: 将结构化数据映射到目标property上并关联算子
- 2、要素标准化: 概念挂载、属性标准化、实体链指
- 3、归一与融合: 实体链指&归一, 关系、属性合并等

SPG: 知识生产算子框架kNext

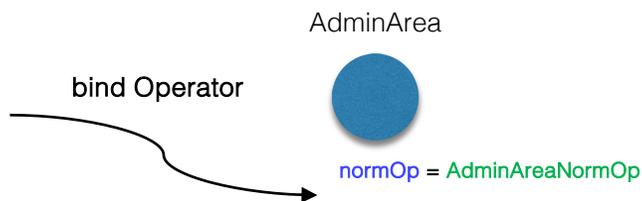


实体类型定义

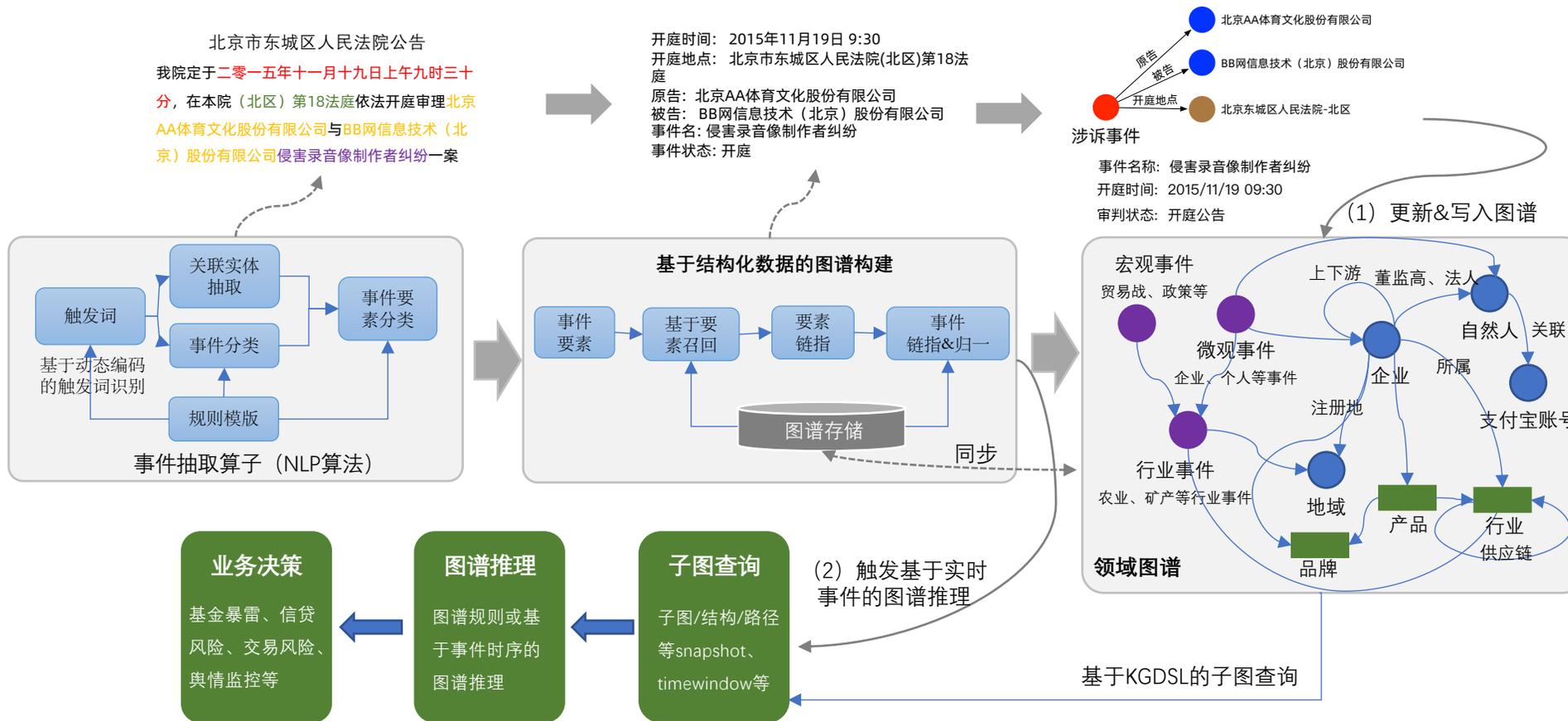
算子开发与绑定

```
EntityClass User {
    id
    name
    phoneNo
    bornPlace
    homeAddress
    UserNormId
    String
    ChnMobilePhone
    AdminArea
    POI(高德POI)
}
```

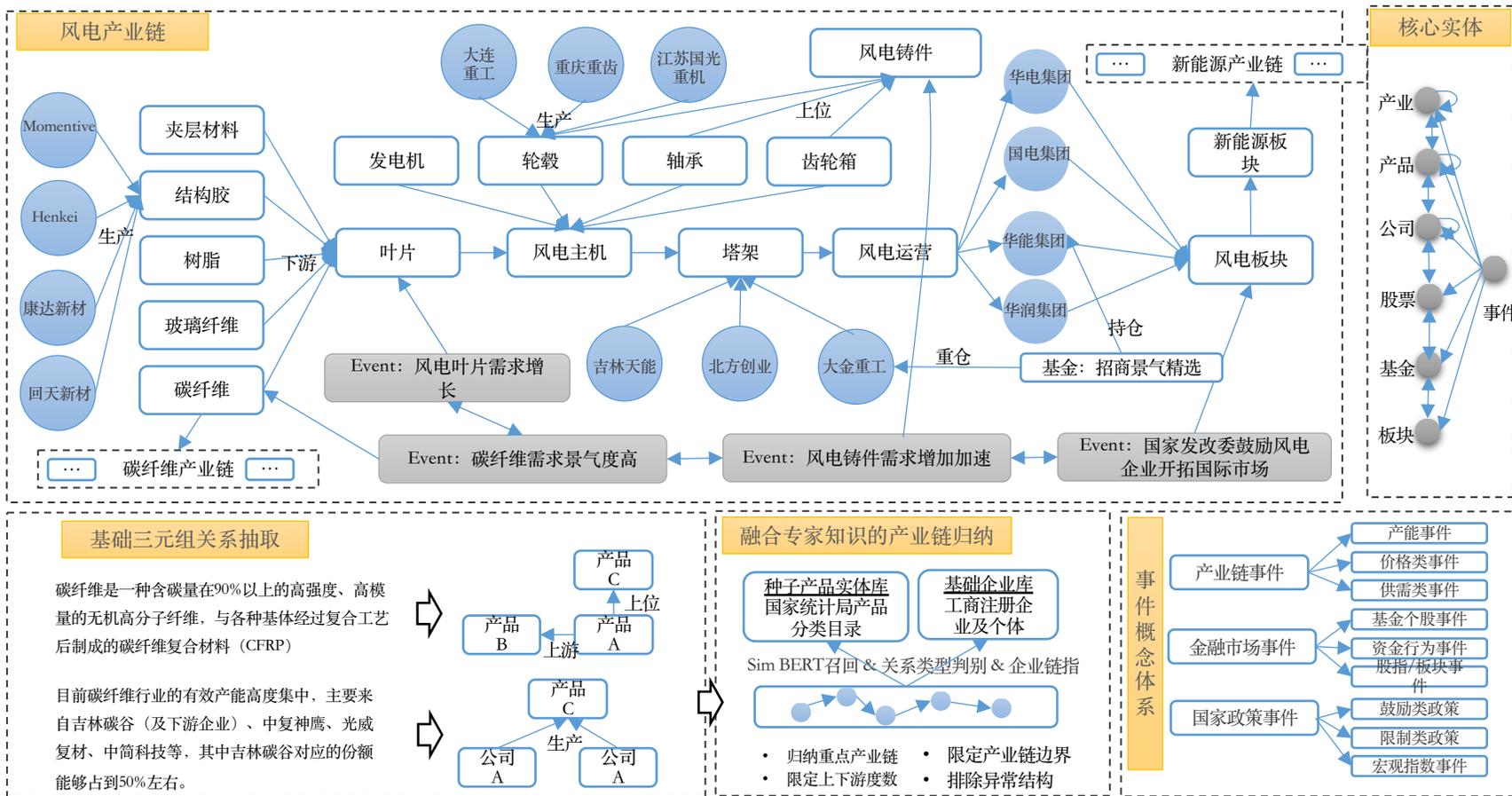
```
@BaseOp.register("AdminAreaNormOp", bind_to="AdminArea", is_api_iface=True)
class AdminAreaNormOp(PropertyNormalizeOp):
    def eval(self, property: str, record: Vertex = None) -> Union[str, Trace]:
        # property = "中国成都市", 需要标化到成都
        # 简单模式
        if "成都" in property:
            return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



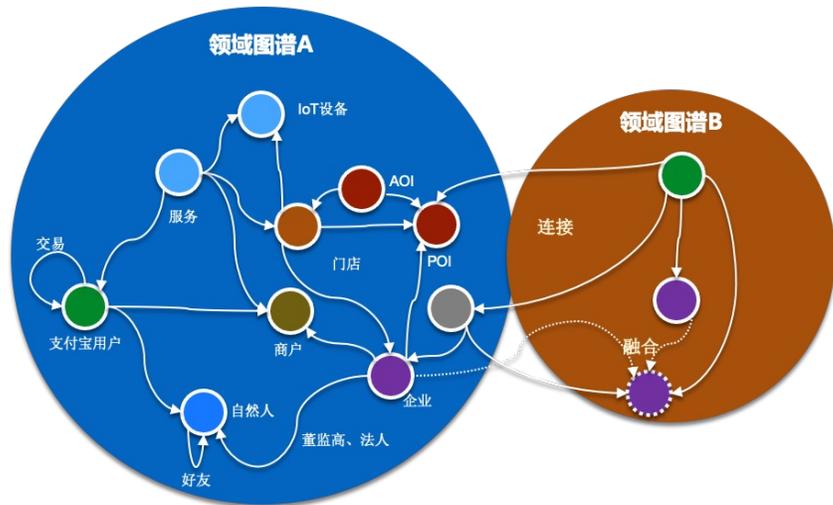
SPG: 基于知识生产框架的事理图谱事件抽取



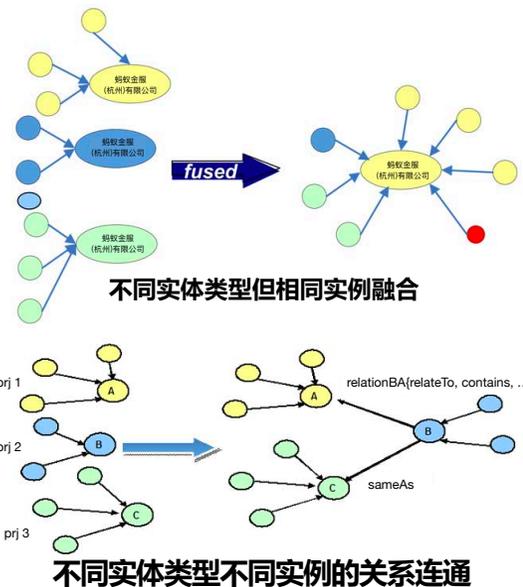
SPG: 基于产业链的事理图谱风险传导与信息协同



SPG: 通过知识图谱实现连接即可用的跨域知识复用

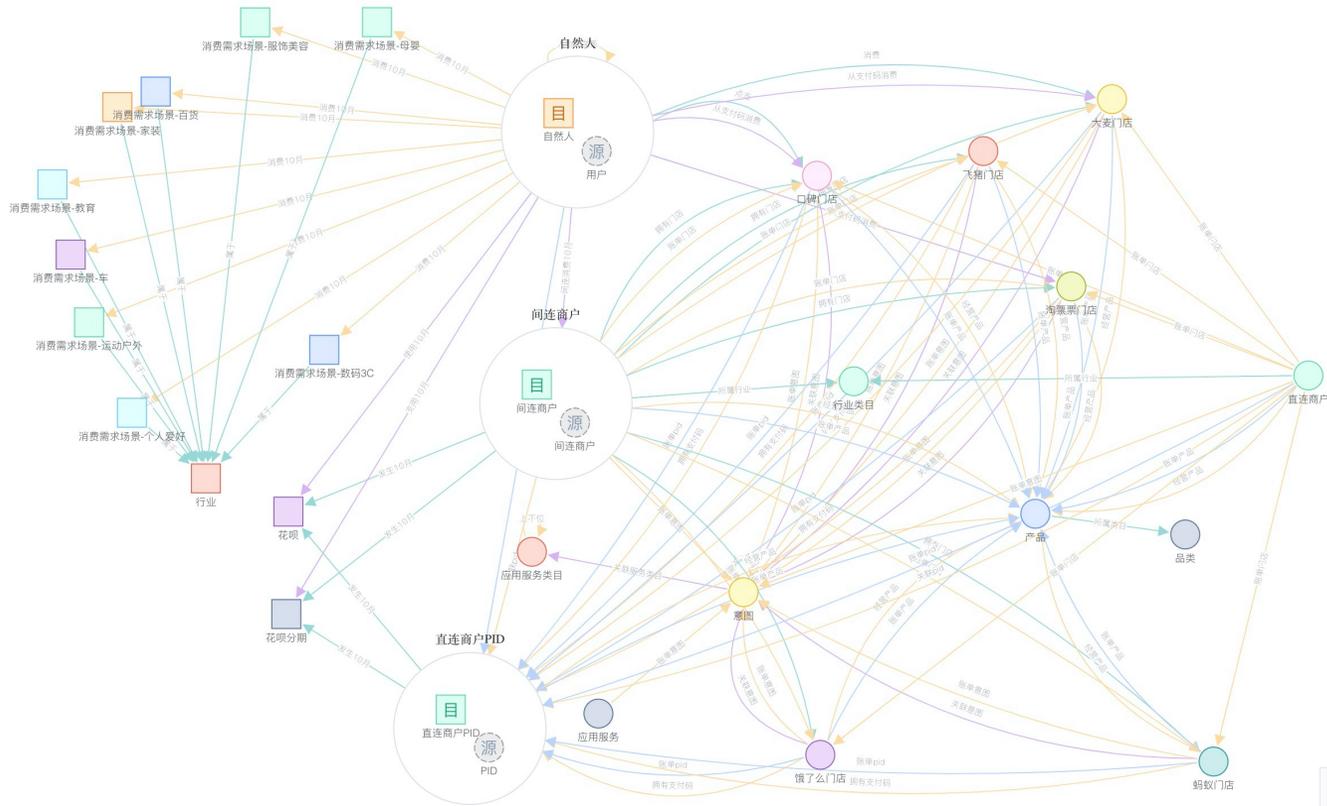


示意：领域图谱B 融合/复用 领域图谱A



- **跨业务的知识复用：**基于图谱本体模型（面向对象），实现跨业务的知识连接、复用
- **减少无效数据拷贝：**减少无效的数据拷贝，连接即可应用，标准化知识服务链路
- **业务快速价值落地：**减少业务找数据的成本，通过知识复用带来更大业务价值，降本提效

SPG: 商家图谱&消金图谱融合支持商家运营

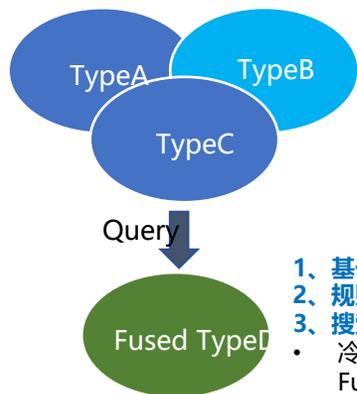


业务目标: 通过融合商家图谱用户、商户等数据, 进一步提升商户行业分类的准确率, 提升B端作业效率及C端转化, 应用图谱推理算法(KARI)

业务收益:

- 融合后的商户平均度数从1.73提升至2.37, 对全量的直连/间连商户表征的覆盖率提升至80%。
- 融合后的图谱表征覆盖率提升至94.3%, 覆盖了近一年的有交易商户
- 在多下游应用 (竞对识别、商户行业标识、花呗分期预测等) 中有较明显效果
- 通过融合增加15类实体、53类关系覆盖, 自动融合, 减少无效数据拷贝

SPG: 跨图谱知识融合的代表与编程范式



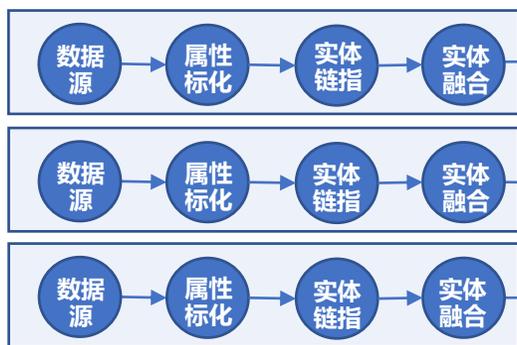
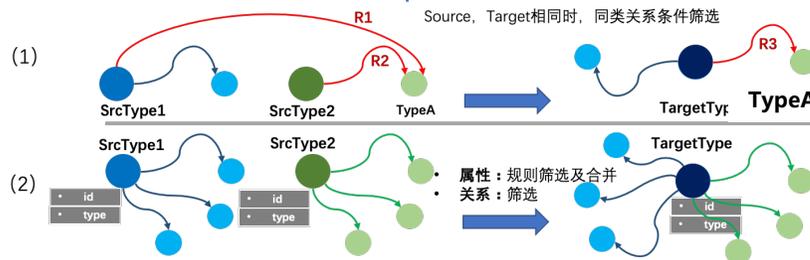
```
Define FusedPOI equivalentClass fuseOf(AmapPOI, AlipayPOI) {
  Rule {
    linkFunc = samePoiSimilarityFunc
    fuseStrategy = poiFuseRuleStrategy
  }
}
```

可自助开发的多链指策略

- 1、**基于配置**: 链指计算在外部系统完成, 输入源、目标 ID列表
- 2、**规则链指**: 如 SrcType1.id == SrcType2.id
- 3、**搜索引擎**: items = Search(query); rerank(query, items)
 - 冷启动阶段固定选择一类型(如TypeB)构建索引, 完成一轮融合得到 Fused TypeD
 - 基于Fused TypeD更新或重建索引, 实现A-C消歧融合 (关系生成、概念挂载不需要)
- 4、**向量检索**: cosine (emb(A),emb(B)) 通过离线批量或在线调用得到 pair list

```
(1) SrcType1->TypeA[R1] as R1;
     SrcType2->TypeA[R2] as R2;
     TargetType->TypeA[R3] as R3;
     R3 = isExist(R1) ? R1 : R2;
(2) TargetType.type = isNotNull(SrcType1.type) ? SrcType1.type : SrcType2.type
... ..
```

可溯源的归一规则定义



单领域图谱构建



跨图谱融合

Fused Namespace

- 1、领域图谱生产与跨域融合相互独立
- 2、保障各领域图谱的独立迭代
- 3、保障知识引用价值可跟踪

SPG: 谓词语义与逻辑符号(Logical Symbols)



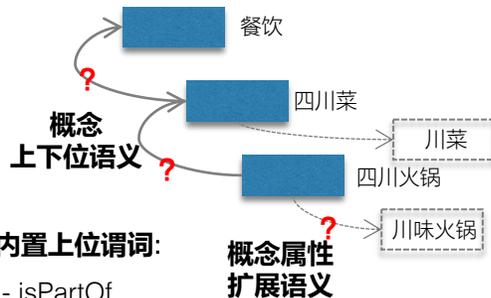
内置谓词:

- equivalentClass
- belongTo
- sameAs

伪码示例:

```

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)
    ... ..
  }
)
    
```



内置上位谓词:

- isPartOf
- subCategoryOf
- isA
- ...

概念属性扩展语义

内置属性谓词:

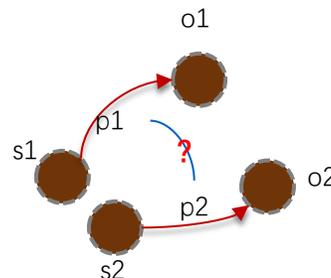
- hasAlias
- synonym
- ...

提示词联想示例:

getPrompts(四川火锅)

上位词: 四川菜, 川菜

同义词: 川味火锅, 川式火锅, 重庆火锅, 巴蜀火锅



内置谓词:

- 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/爱成都火锅人群)
{
  GraphStructure{}
  Rule {
    s.prefers contains(川式火锅)
  }
}
```

```
GraphStructure {
  (s1:Crowd/爱成都火锅人群) -[p:visited]->(o:MifShop)
  (s2:Product) -[p:availableOn]->(o:MifShop)
}
Rule {
  s2.category contains(四川菜)
}
Action {
  get(s2.name)
}
```

SPG: 基于语义逻辑实现知识衍生

```
Define (s:User)-[p:belongTo]->(o:TuringCrowd/大龄未婚青年) {  
  GraphStructure {  
  }  
  Rule {  
    R1: s.age > 25 && s.age < 50  
    R2: s.marriageStatus = '未婚'  
    p: R1 && R2  
  }  
}
```



```
Define (s:User)-[p:belongTo]->(o:TuringCrowd/高收入大龄未婚青年) {  
  GraphStructure {  
    (s)-[p1:belongTo]->(o1:TuringCrowd/大龄未婚青年)  
  }  
  Rule {  
    R1: s.incomeLevel = '大于等于4万'  
    p : R1 and p1  
  }  
}
```

一阶谓词表达式



$User(s) \wedge age(s, x) \wedge x > 25 \wedge x < 50 \wedge marriageStatus(s, 未婚)$
->
 $belongTo(s, TuringCrowd/大龄未婚青年)$

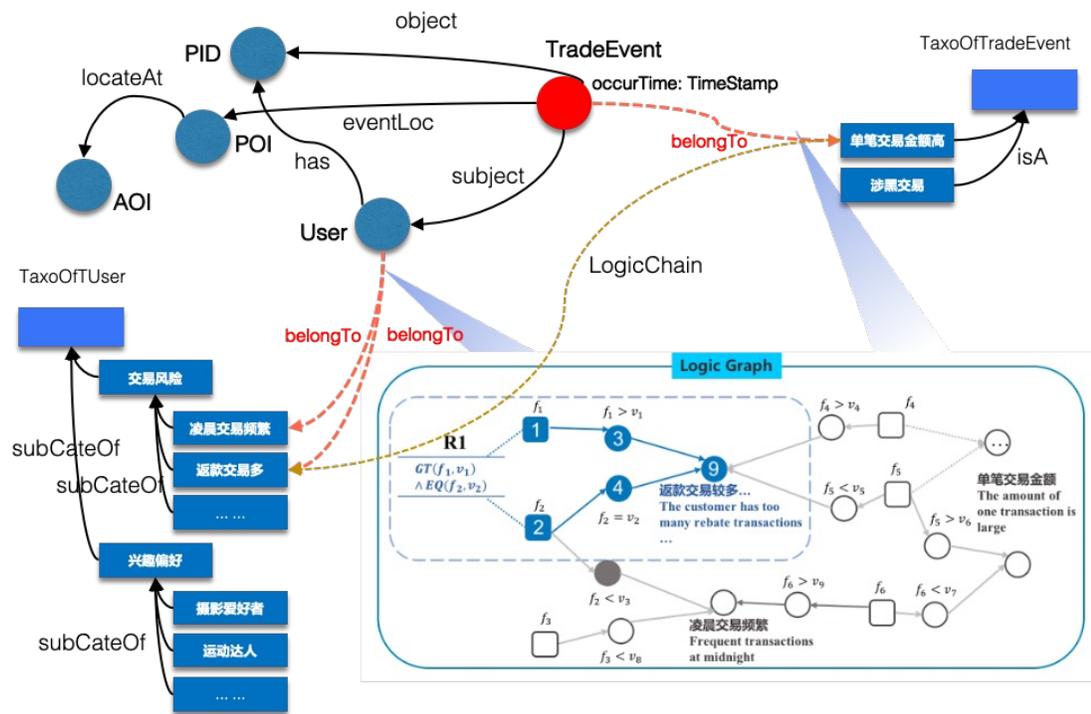
一阶谓词表达式



$User(s) \wedge (s, TuringCrowd/大龄未婚青年) \wedge incomeLevel(s, 大于等于4万)$
->
 $belongTo(s, TuringCrowd/高收入大龄未婚青年)$



SPG: 基于语义逻辑实现LogicChain



涉黄用户: 凌晨交易频繁 & 单笔交易金额高

大额套现用户: 返款交易多

Define (e:TradeEvent)-[p:belongTo]->(o:TaxoOfTradeEvent/单笔交易金额高)

```

{
  GraphStructure{
  Rule {
    e.amount > 500
  }
}

```

Define (s:User)-[p:belongTo]->(o:TaxoOfUser/交易风险/返款交易多)

```

{
  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/单笔交易金额高)
  }
  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/交易风险/凌晨交易频繁)

```

{
  GraphStructure{
    (e1:TradeEvent)-[ps1:subject]->(su1:User)
  }
  Rule {
    s.id == su1.id
    hour(e1.occureTime) between(0, 4)
  }
}

```

SPG: 基于专家规则的图谱推理

KGDSL: 支持专家规则定义及推理

业界常见做法: 基于图谱语义的规则符号推理

(TBox and ABox Reasoning in Expressive Description Logics)

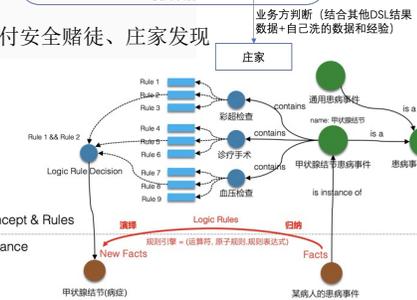
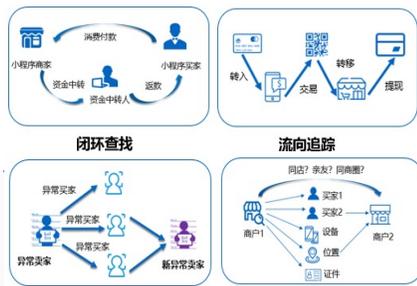
类型约束	$Man \cap Woman \subseteq \emptyset$ $Man(Alice) \oplus Woman(Alice)$	=>
类型推导	$Mother \subseteq Woman$ $Woman \subseteq Person$ $Mother \subseteq Person$	=>
关系推导	$has_son(Alice, Bob)$ $has_son \subseteq has_child$ $has_child(Alice, Bob)$	=>

TBox
Undergraduate \sqsubseteq Teach
Bachelor \equiv Student \sqcap Undergraduate
Master \equiv Student \sqcap Undergraduate
PhD \equiv Master \sqcap Research
Assistant \equiv PhD \sqcap Teach

ABox
Master(Chen)
PhD(Enzo)
Assistant(Rui)

```

Define (s:`FinancialEventTaxonomy`/`铁矿石价格上涨事件`)-[p:leadTo]-{
nEventTaxonomy/`钢铁净利润下降事件`) {
  GraphStructure {
    (s)-[s:subject]->(c)->[r1:product]->(e:Product)->[r2:downs.
  )
  Action {
    createNodeInstance(
      type = o,
      value = {
        subject = d.name
        index = "净利润"
        trend = "下降"
      }
    )
  }
}
    
```

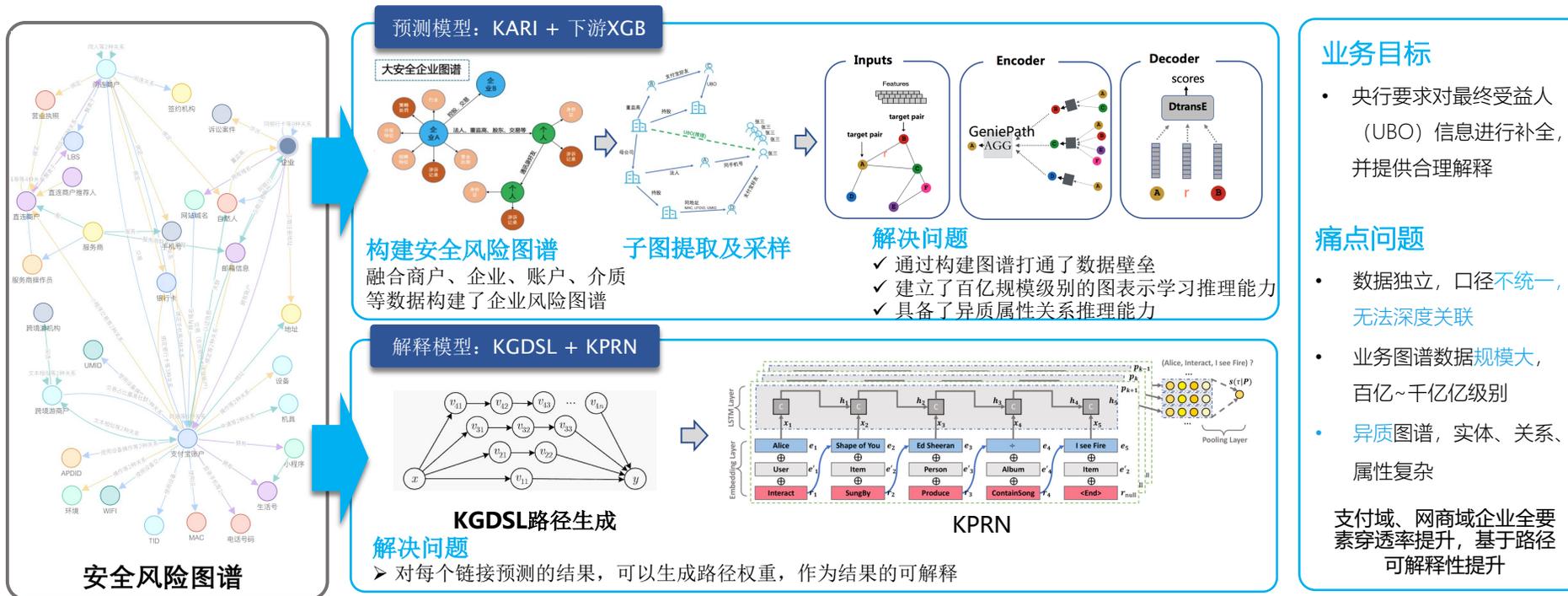


- 1、符号推理常用表示, 严格语义图为基础
- 2、面向业务应用有较高门槛
- 3、缺少逻辑规则、阈值判断等表达

结合蚂蚁应用特性的表示

- 基于Geaflow VC/Traversal构建的KGDSL, 支持离线批量、近线分钟级规则推理
- 支持graph pattern、logic rule推理表达, 支持决策、更新图谱、聚合判定等多种Action表达
- 离线规则推理规模>300亿, 近线规则推理时效性1m - 30m时效性
- 适用面: 安全、理赔、资金、信贷等

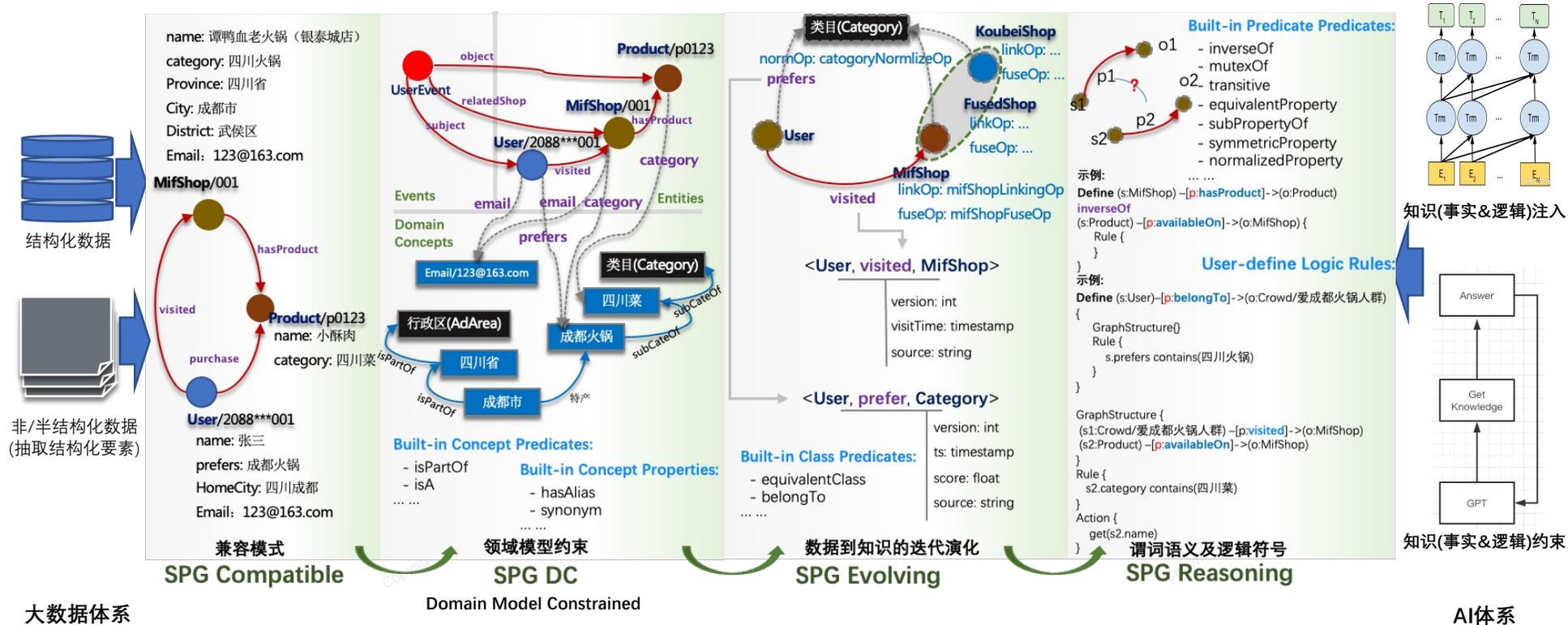
SPG: 探索逻辑规则与模型结合的可解释推理



- KARI: 支持异质图学习的encoder-decoder框架, 基于图谱KGDSL输出多维度子图特征, 度(degree)、pagerank、邻居统计、规则等
- KGDSL + KPRN: 基于图谱关联路径及规则学习输出推理结果的图谱可解释路径

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

衔接大数据与AI技术体系，帮助机器更好的理解世界

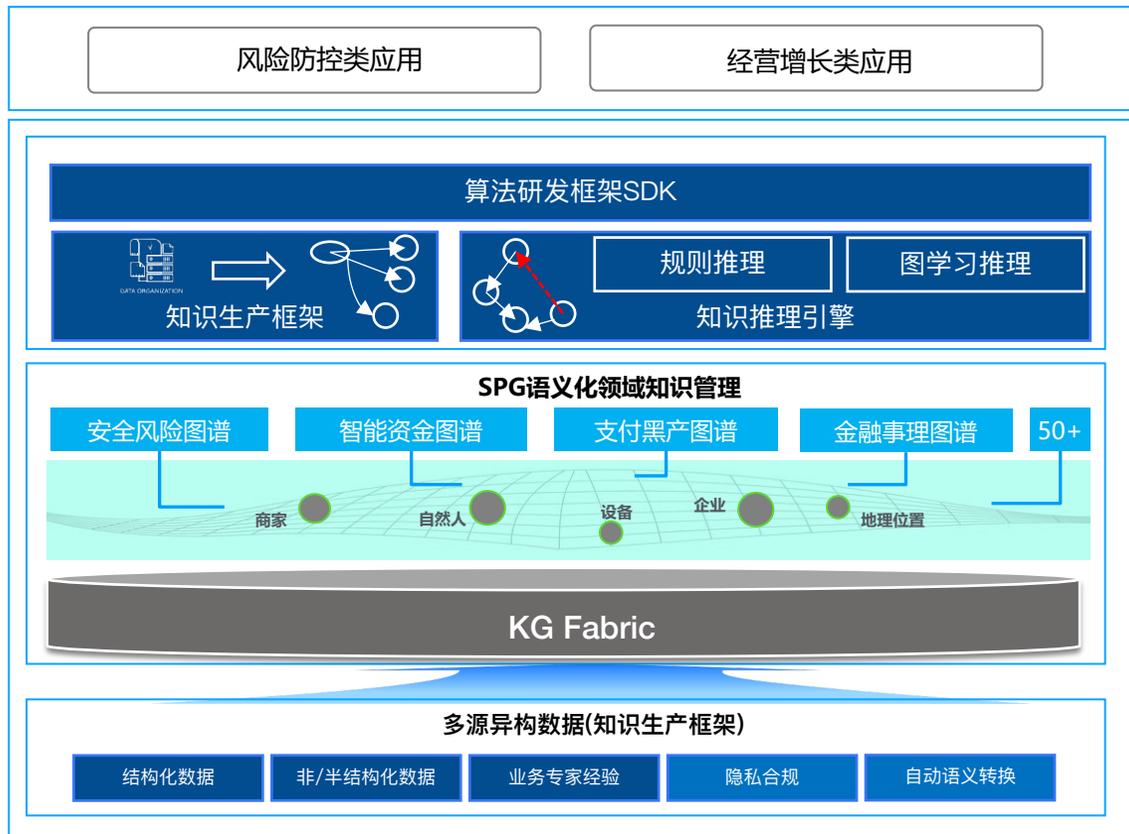


PART 04

4. 蚂蚁知识图谱引擎



蚂蚁知识图谱 – 基于SPG的知识图谱引擎



首个SPG知识引擎
跨场景可迁移、新领域图谱快速孵化

KG Fabric
分钟级、零拷贝，1.3万亿级线性扩展

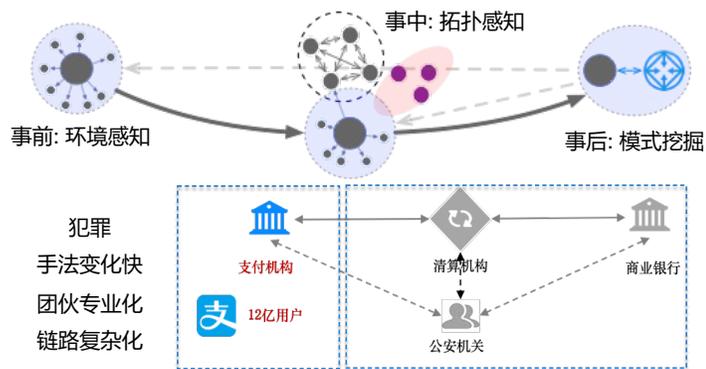
知识推理引擎
千亿规模、规则引导学习、可解释推理

算法框架及SDK
建模、构建、推理(python + KGDSL/GQL)

授权专利35项， 提报专利140+项， 参编标准10项， 获浙江科技进步二等奖、数博会优秀成果奖等

蚂蚁知识图谱 – 新技术范式的应用效益

以‘资金链’为核心构建风险防控体系



人民银行: 推进“资金链”治理, 支付行业常态化治理格局持续完善, 组织商业银行、支付机构、清算机构协助公安机关阻断大量涉诈资金转移, 挽回大量人民群众损失。

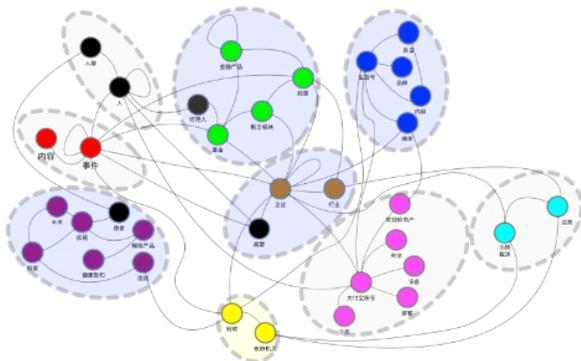
打击新型网络犯罪, 守护用户资金安全

- 语义重构资金链, 超大规模, 跨62天的1w-hop, avg 30s深度追金
- 面向海量交易实时风险防控, 克服深度实时追金大数据/图不可计算难题
- 公安的主动联动, 与黑产的主动对抗, 保护用户442万余人, 涉诈资金45亿余元
- 23省/87地市公安省厅, 405名受害群众返还涉诈资金1046万余元

社会价值

蚂蚁知识图谱 – 新技术范式的应用效益

KG Fabric 实现跨图谱互联网互通



Gartner: 基于知识图谱的下一代数据管理范式。链接数据孤岛，数据知识语义化

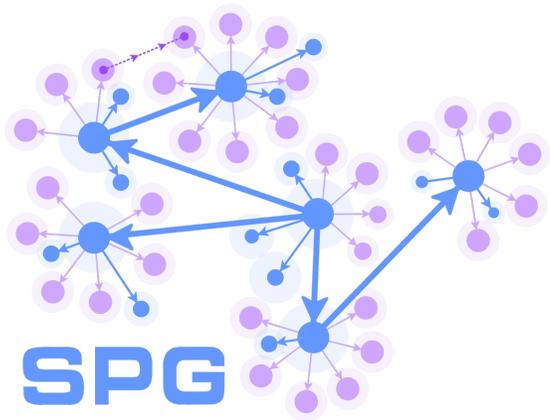
- 连接即可用、零拷贝、分钟级跨图谱融合
- 商家图谱 x 消费类图谱: 平均度数从1.73提升至2.37, 减少84%重复拷贝
- 资金图谱 x 黑产图谱 x 商户图谱, 节省计算/存储资源57.14%, 处理时效上提高4倍

经济价值

蚂蚁知识图谱 – 新技术范式的应用效益

OpenKG共建，工业级知识语义框架SPG

8月27日SPG白皮书发布，欢迎提前关注公众号，下载交流... ..



以属性图为基础，构建机器可理解的知识符号体系



语义增强可编程图谱框架

微信扫描二维码，关注我的公众号

- **主体模型**准确刻画主体语义，**谓词逻辑**跨主体的逻辑依赖传导
- 无缝兼容大数据架构，ER2SPG、大规模图学习、可解释推理等
- 联动LLMs作为领域知识库，事实校验、语义联想、逻辑知识链等
- **填补工业级知识语义框架的空缺**，OpenKG合作**8月发布SPG白皮书**，开源筹备中

技术价值

Q&A



感谢聆听

