

知识计算即服务：赋能企业知识化转型



郑毅 | 华为云-知识图谱负责人



目录

-  从数字化转型到知识化转型
-  知识图谱存储：图数据库
-  知识图谱构建与应用：华为云知识图谱云服务
-  企业知识计算案例
-  未来展望

AI是一种新的生产力，改变将涉及各个行业

公共

- 平安城市
- 智慧交通
- 灾害预警

教育

- 个性内容
- 注意力提升
- 机器助教

健康

- 早期预防
- 辅助诊断
- 精准治疗

媒体

- 舆情分析
- 内容摘要
- 内容审核

制药

- 缩短周期
- 精准试验
- 精准药物

物流

- 路径规划
- 货物监视
- 自动分拣

金融

- 文档处理
- 实时防欺
- 智能投顾

保险

- 高效鉴定
- 欺诈预防
- 智能回访

零售

- 无人超市
- 实时库存
- 精准推荐

制造

- 品质检测
- 工业物联
- 预测维护

电信

- 客户服务
- 网络维护
- 网络优化

油气

- 精准勘探
- 远程维护
- 运营优化

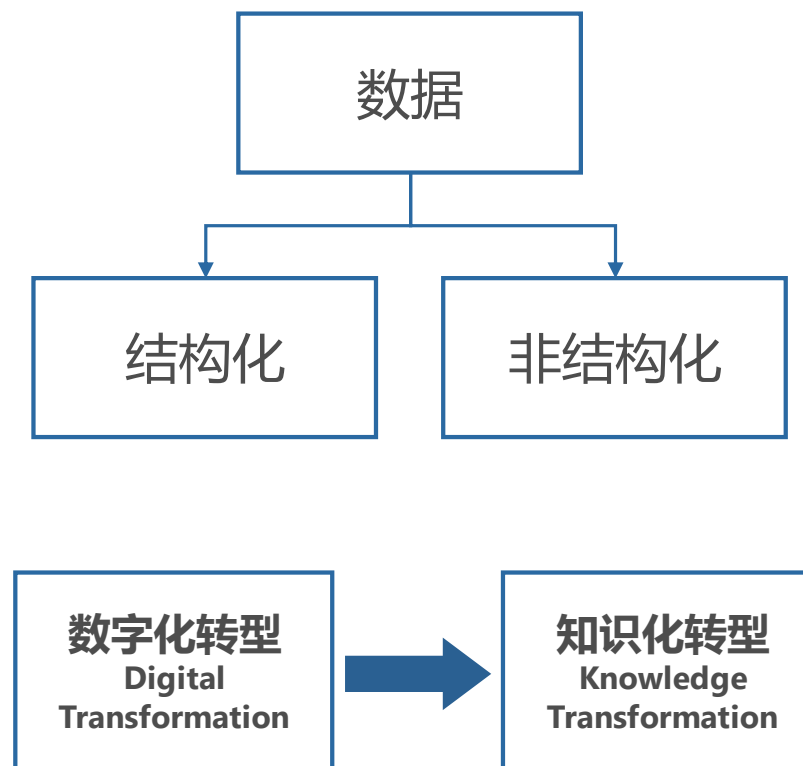
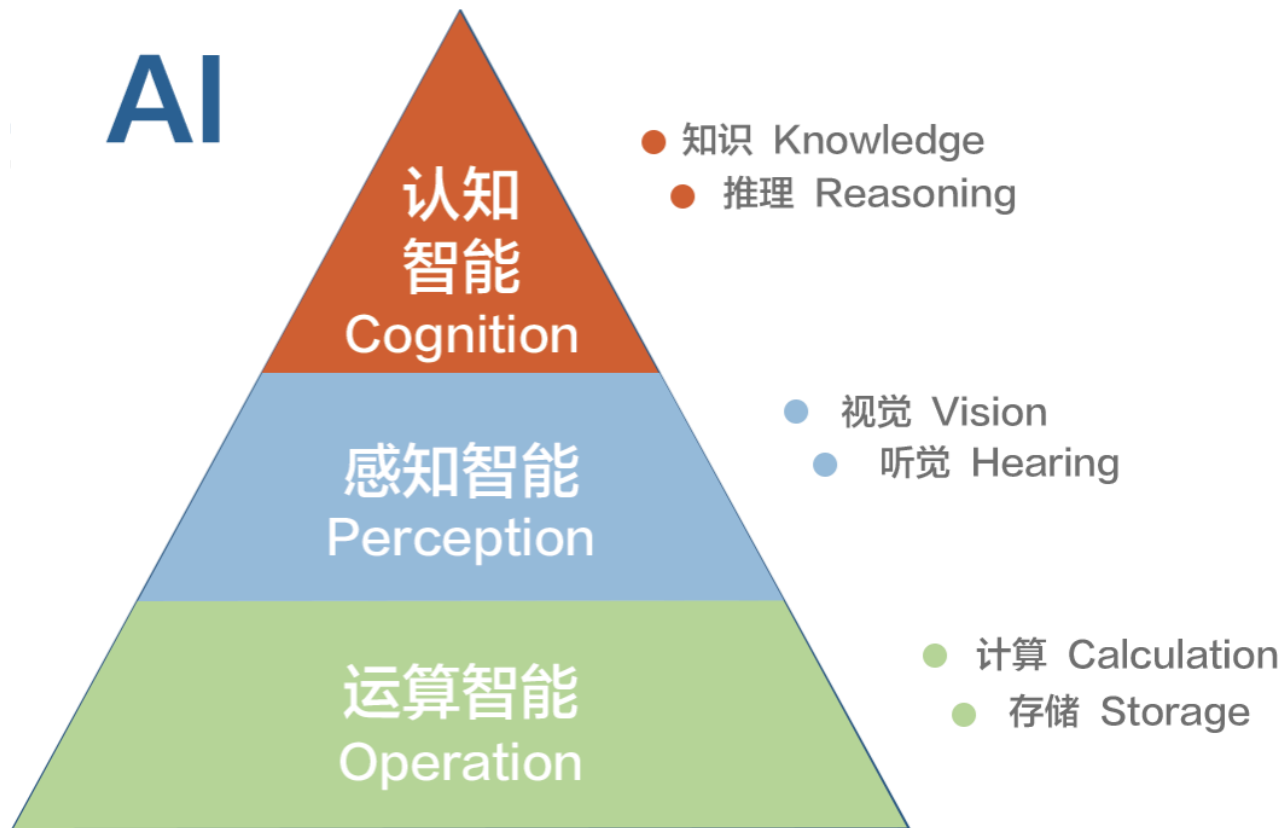
农业

- 施肥优化
- 远程作业
- 高效育种

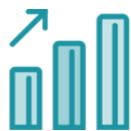
实践是探索AI落地的唯一方式：ABCD

$$\left[\begin{array}{cc} \text{数据} & \text{行业智慧} \\ \text{Big Data} & \text{Domain} \end{array} \right] \times \left[\begin{array}{c} \text{算法 Algorithm} \\ \text{算力 Computing} \end{array} \right] = 10+ \text{行业探索} \quad 200+ \text{项目实践}$$

AI v.s. 数字化转型 & 知识化转型



为什么要知识化转型?



效率提升

Efficiency improvement

重复工作
海量数据检索



专业传承

Expertise transfer

知识沉淀
专家经验传承



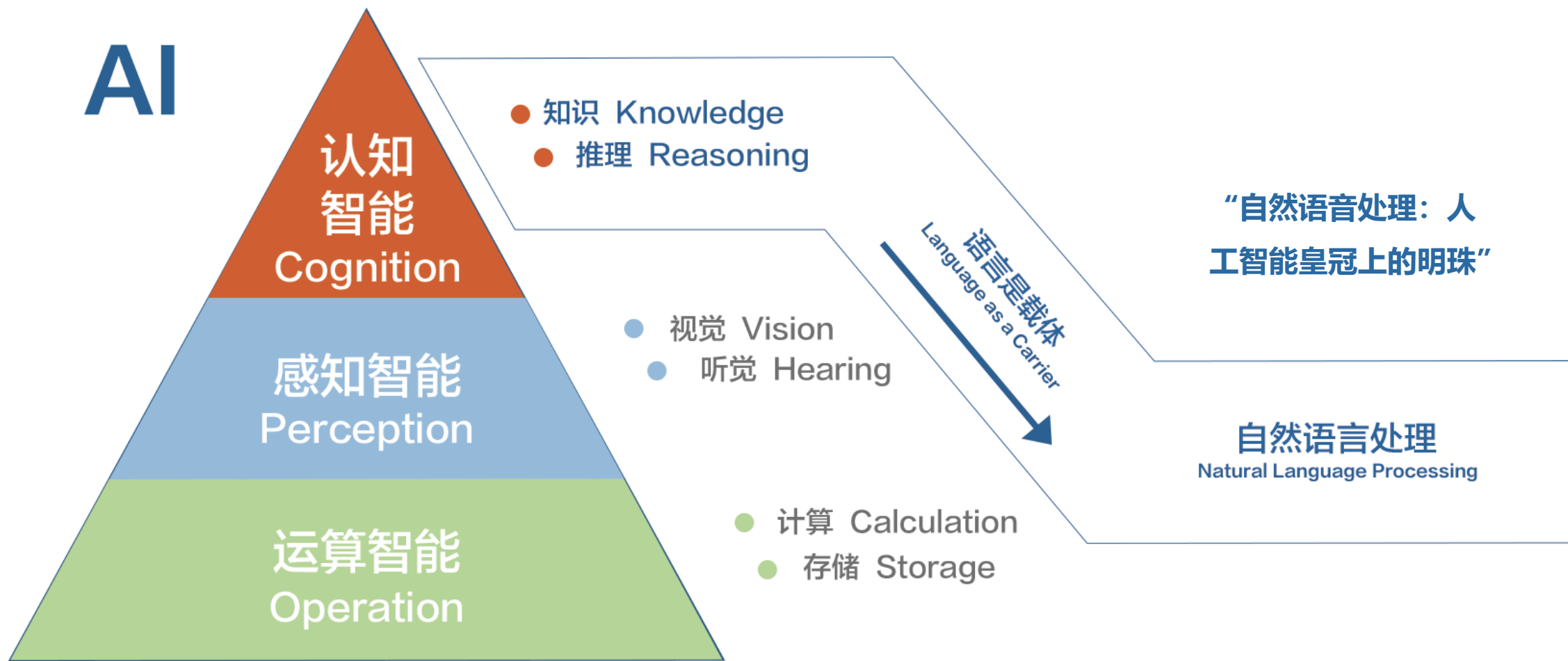
管理优化

Management optimization

组织精简
多部门协同

知识来自于组织、行业、客户、员工.....

人工智能和自然语言处理AI and NLP



自然语言处理的困难

歧义

冬天能穿多少穿多少，夏天能穿多少穿多少
乒乓球拍卖完了

语序

问题是没有钱 没钱是有问题
有钱是没问题 钱是没有问题

间接理解/暗示/上下文关联

甲：“晚上去KTV吗”？

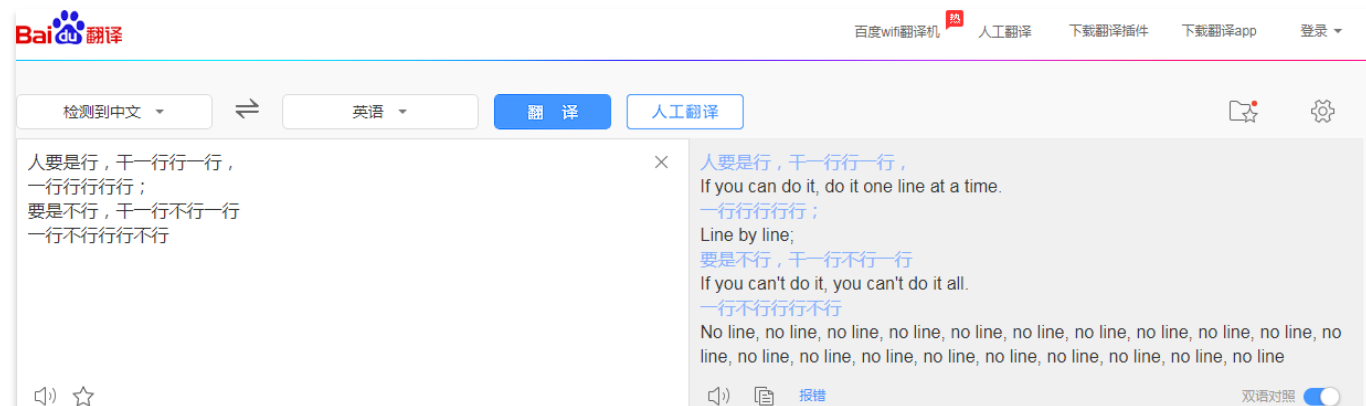
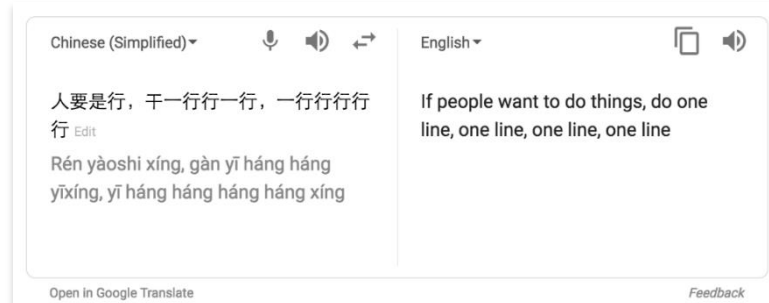
乙：“我弟从乌鲁木齐回来了”。

范冰冰&李晨：“我们不再是我们，我们依然是我们”

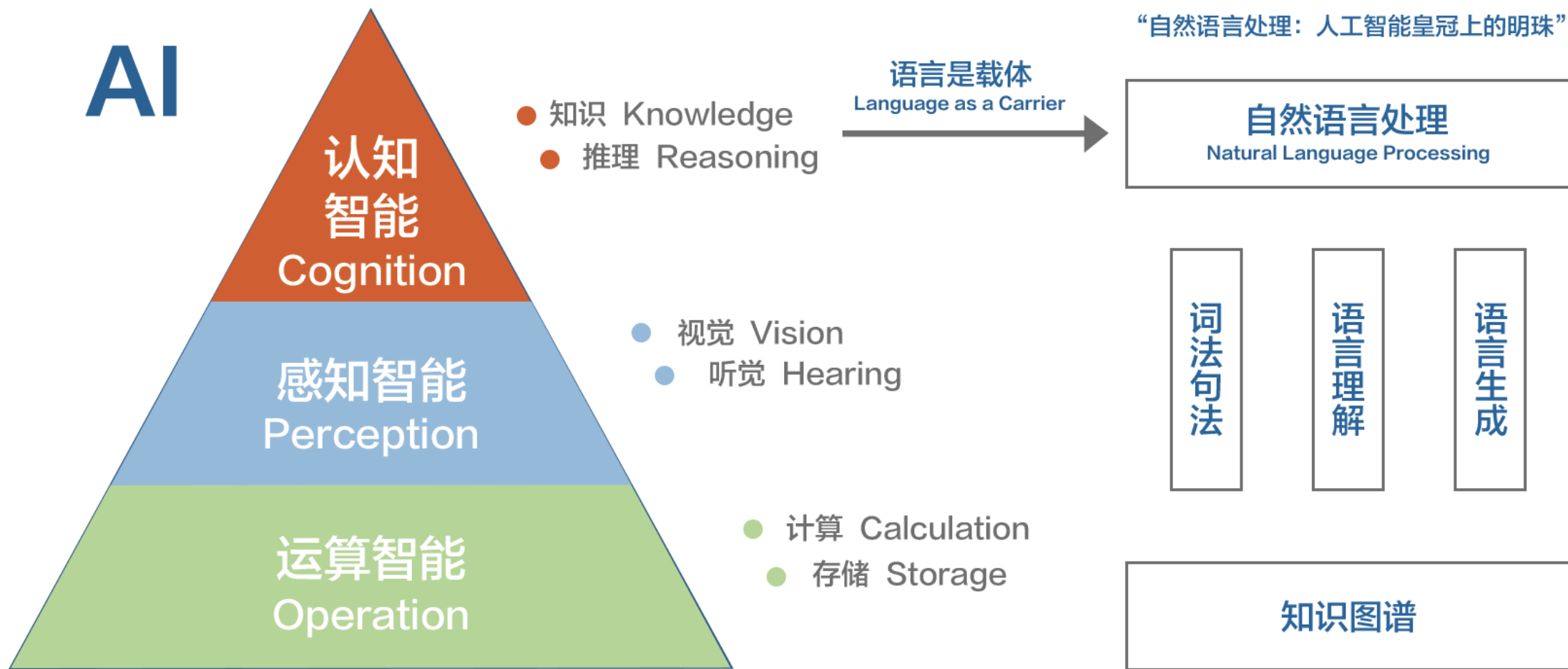
分词断句/多义词

骑车差点摔倒，好在我一把把把把住了。
过几天天天天气不好
小龙女动情地说：“我也想过过过儿过过的生活”

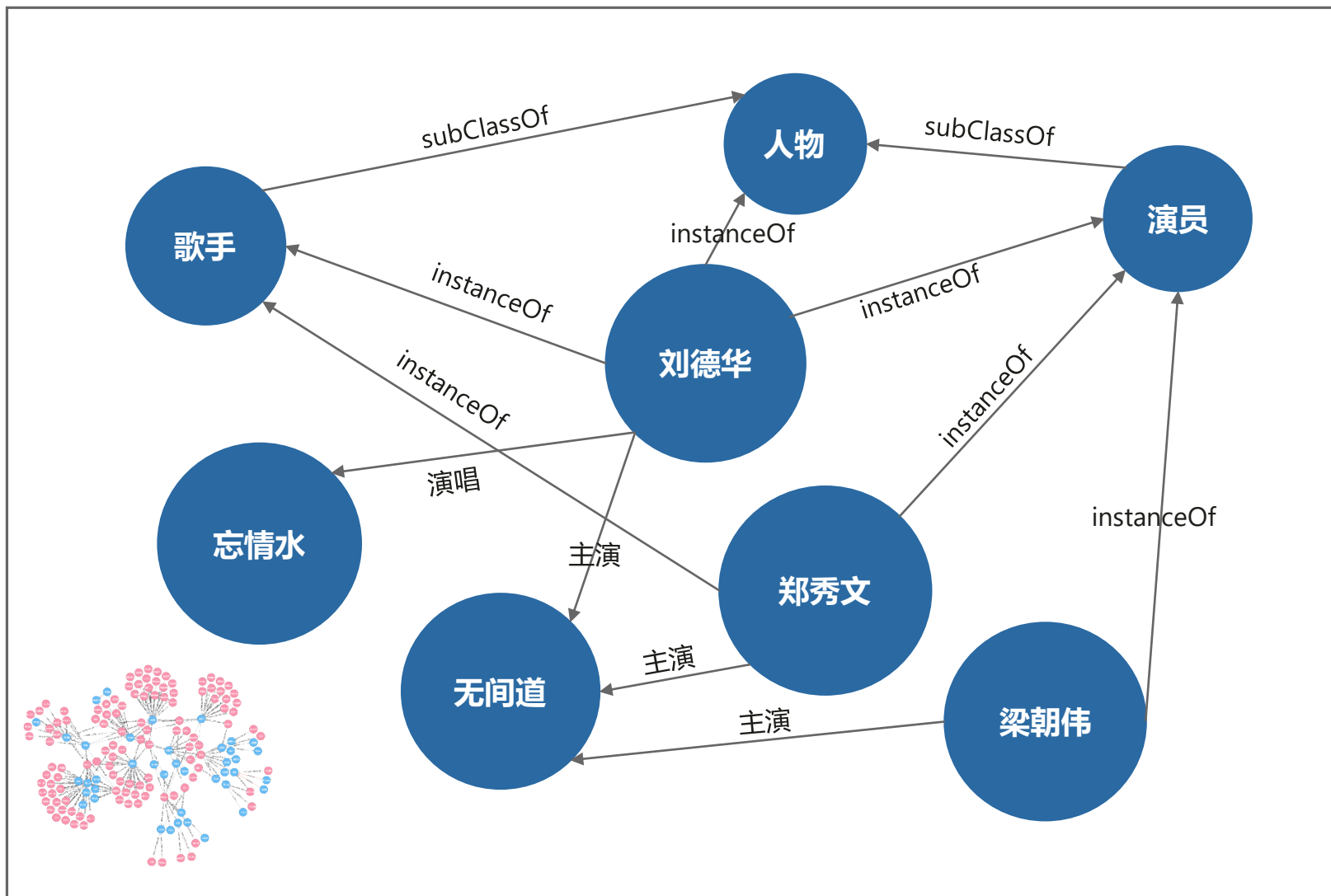
人要是行，干一行行一行，
一行行行行行；
要是不行，干一行不行一行
一行不行行行不行



人工智能和自然语言处理AI and NLP

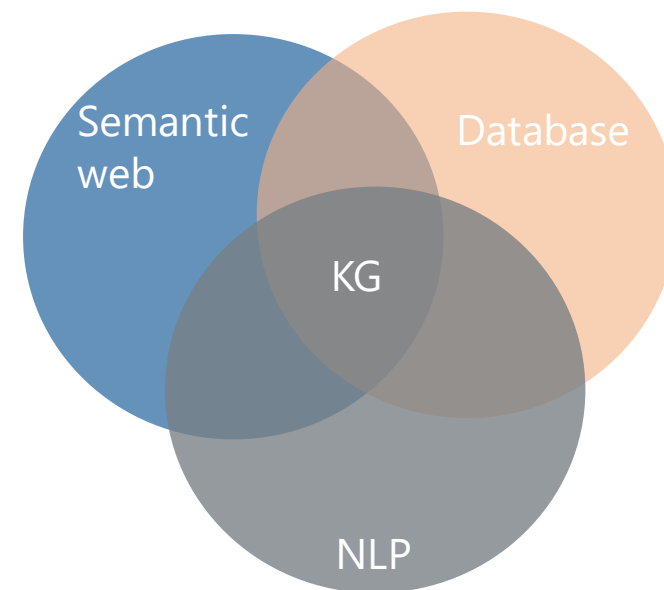


知识图谱是什么?



Entity 实体

Relationship 关系



知识图谱分类

通用知识图谱	行业知识图谱
本体/知识来源于 众包的百科网站	本体/知识来源于 专家经验、领域专业文档
构建数据多为 (半)结构化数据 , 构建 过程自动化程度较高	构建数据 形式多样 , 既有 结构化数据 , 也有 非结构化数据 , 构建 过程自动化程度较低
强调知识的 广度 , 错误 容忍度相对较高	强调知识的 深度 , 错误 容忍度低
面向 普通用户	面向领域 专业用户

通用知识图谱

国内版 国际版

oscar 2019 winners

All Images Videos 翻译成中文 关闭歌词

SUNDAY, FEB 24 AT 8 PM ET / 5 PM PT*

Oscars 2019 WINNERS 2019 RED CARPET 2018 WINNERS ALL FILMS

Academy Awards - Winners in 2019

Lady Gaga Academy Award for Best Original Music Score

Rami Malek Academy Award for Best Actor

Mahershala Ali Academy Award for Best Supporting Actor

Alfonso Cuarón Academy Award for Best Director

Black Panther 2018 · Sci-Fi ★★★★★

Green Book Academy Award for Best Picture

Roma 2018 · Drama ★★★★★

News about Oscar 2019 Winners

Academy Awards

The Academy Awards are a set of awards for the film industry, given annually by the Academy of Motion Picture Arts and Sciences to recognize excellence in cinematic achievements.

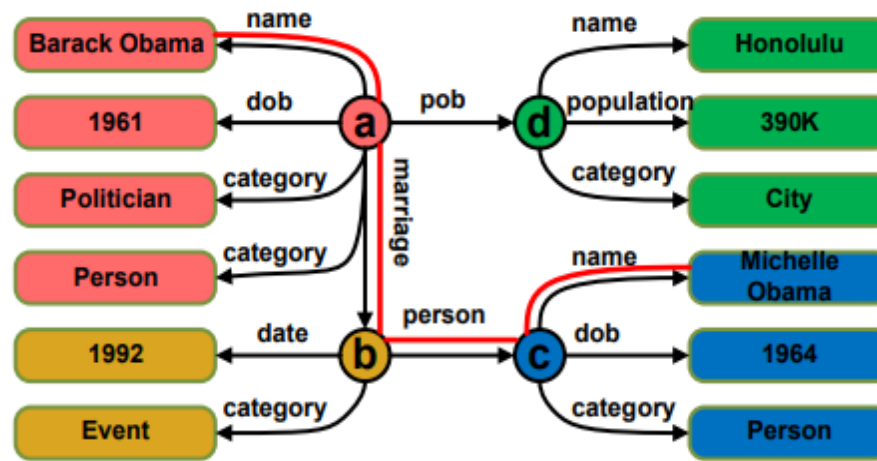
Wikipedia YouTube Official site

Best picture: Green Book (Mahershala Ali)

Best director: Alfonso Cuarón (Roma)

Best lead actor/actress: Rami Malek / Mahershala Ali

Best supporting actor/actress: Mahershala Ali / Rami Malek



Google

美国上一任总统出生在哪里?

All News Images Maps Videos More Settings Tools

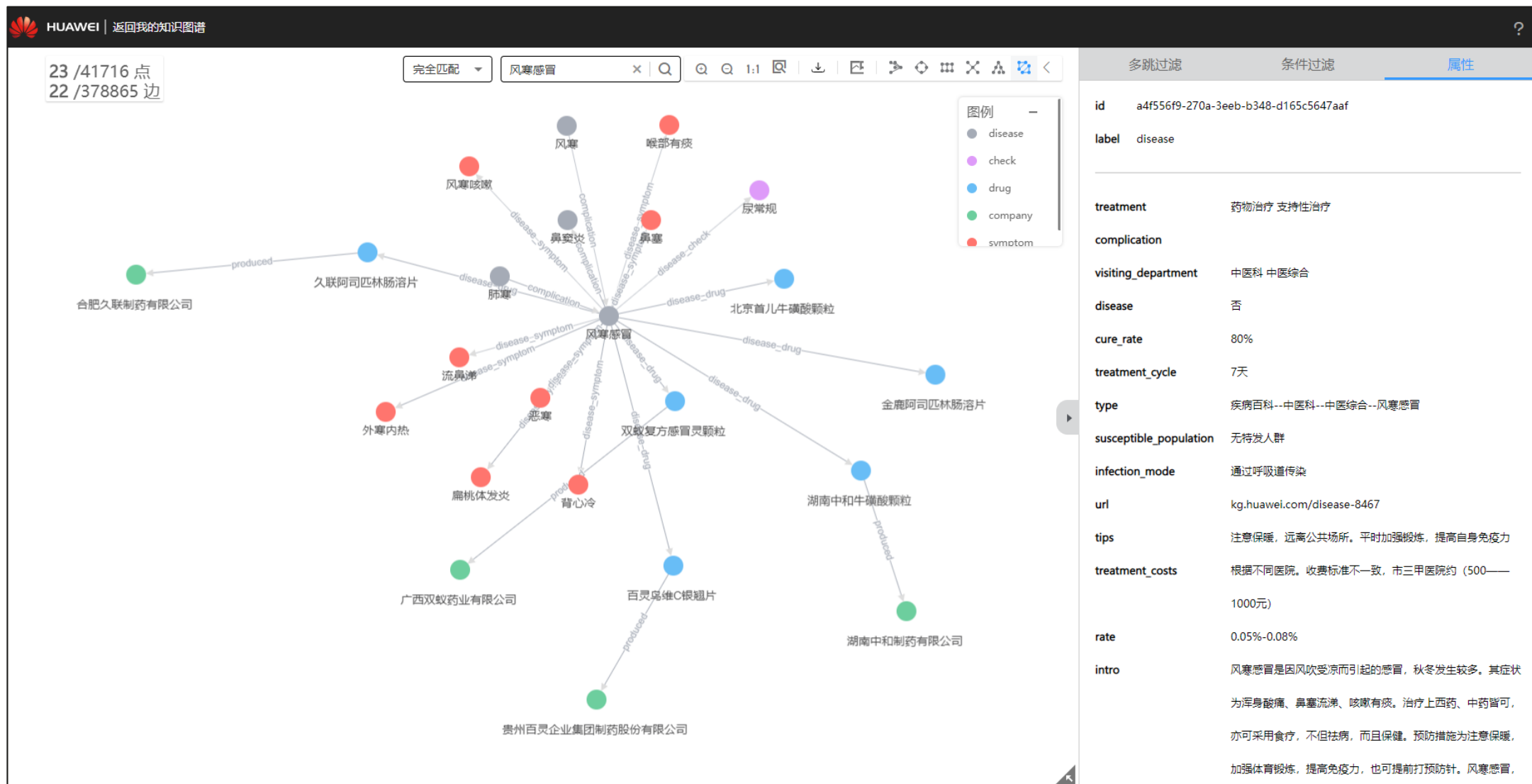
About 1,780,000 results (0.84 seconds)

Barack Obama / Place of birth

See photos

Kapi'olani Medical Center for Women and Children, Honolulu, Hawaii, United States

行业知识图谱



业界首个全生命周期知识计算解决方案

赋能行业构建属于企业自己的知识计算平台



知识获取

行业数据解析器 | 网页解析器
语音识别 | 专家标注
多模态信息抽取 | 文字识别



知识建模

知识表征 | 图谱构建
知识融合 | 多模态表征
图嵌入 | 机理建模



知识管理

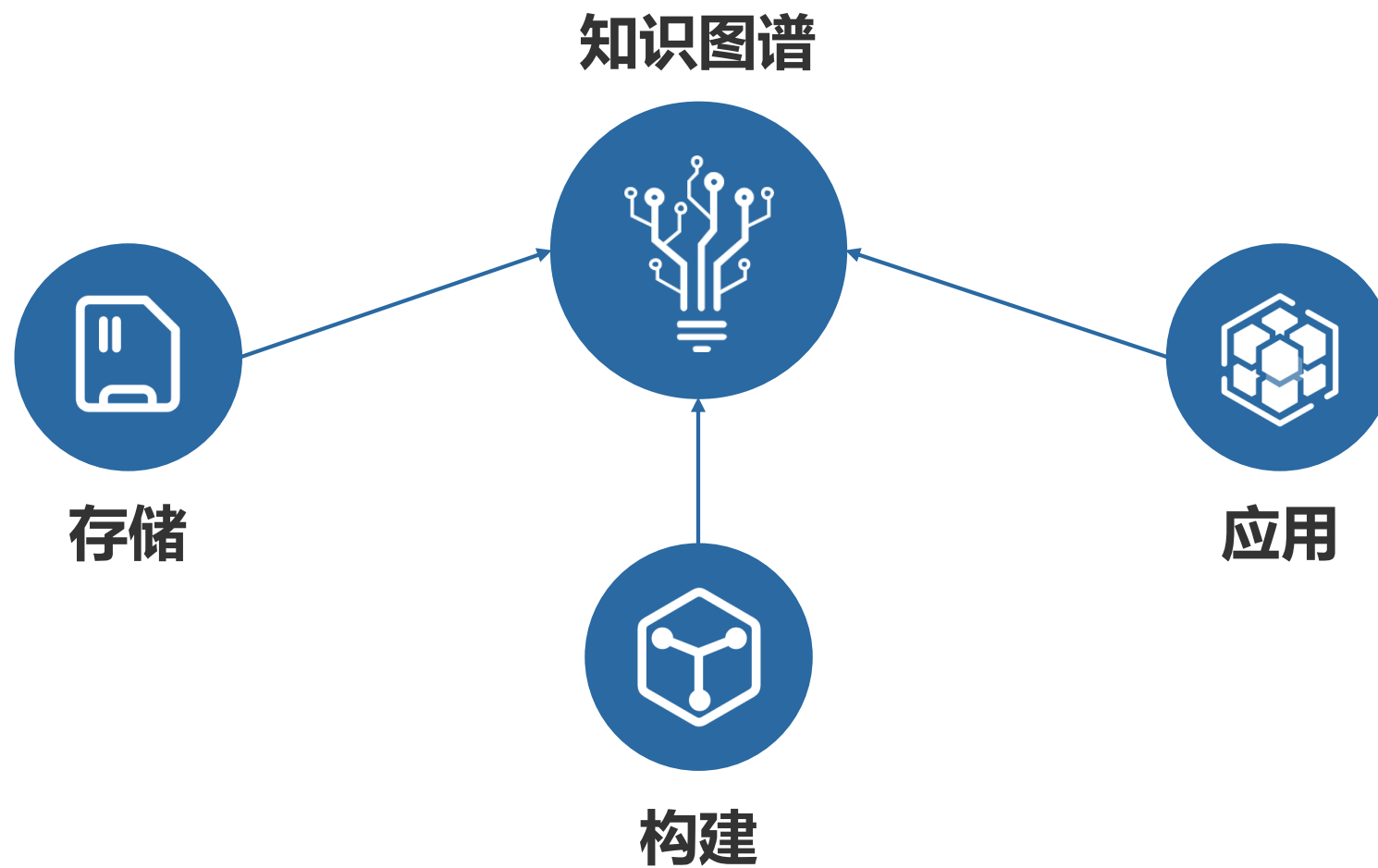
生命周期管理
质量管理
知识补全



知识应用

基础 **高级**
知识搜索 | 智能对话
知识推荐 | 预测分析
软测量 | 知识推理

知识图谱三大问题

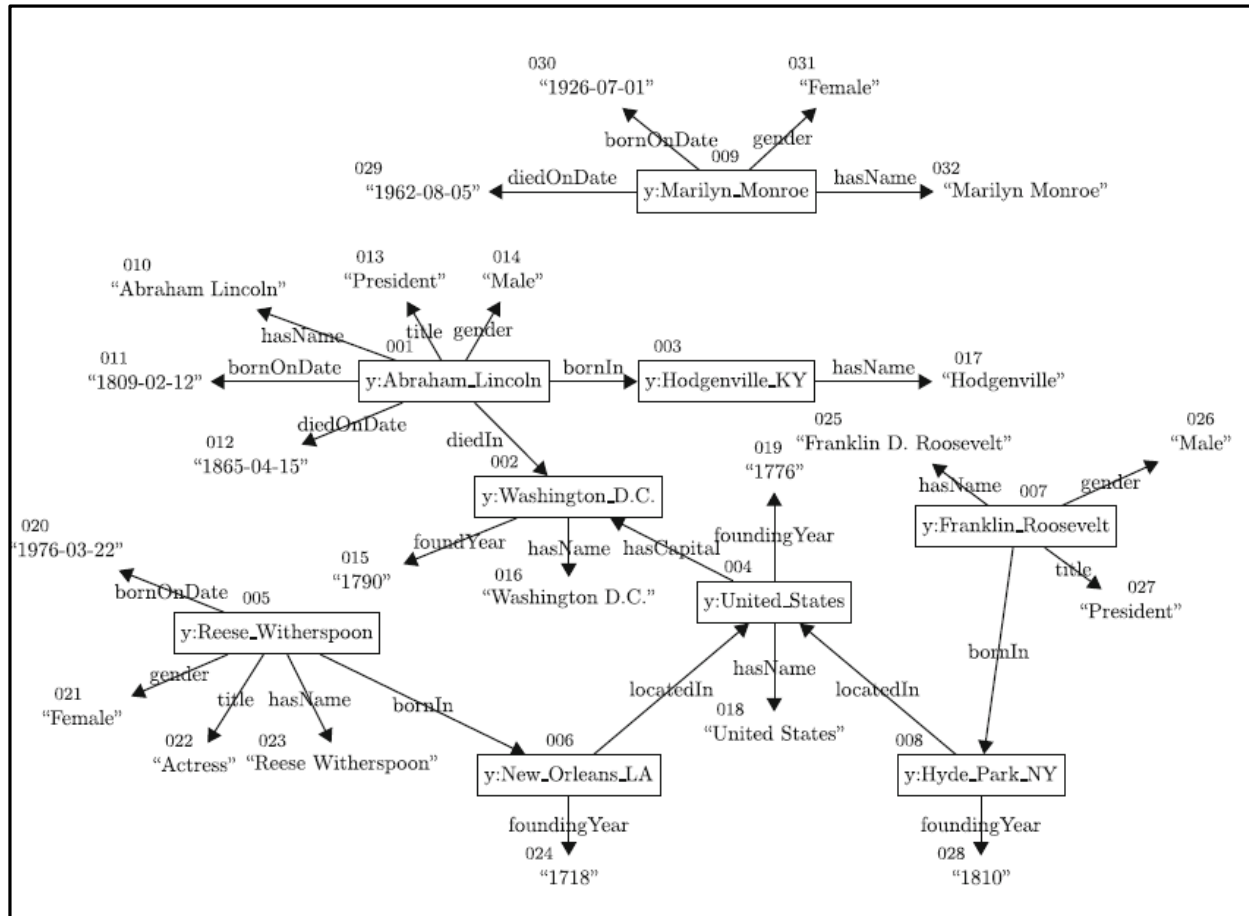


知识图谱存储

知识图谱的存储——两种表现形式

Subject	Property	Object
y:Abraham_Lincoln	hasName	"Abraham Lincoln"
y:Abraham_Lincoln	bornOnDate	"1809-02-12"
y:Abraham_Lincoln	diedOnDate	"1865-04-15"
y:Abraham_Lincoln	bornIn	y:Hodgenville_KY
y:Abraham_Lincoln	diedIn	y:Washington_DC
y:Abraham_Lincoln	title	"President"
y:Abraham_Lincoln	gender	"Male"
y:Washington_DC	hasName	"Washington D.C."
y:Washington_DC	foundingYear	"1790"
y:Hodgenville_KY	hasName	"Hodgenville"
y:United_States	hasName	"United States"
y:United_States	hasCapital	y:Washington_DC
y:United_States	foundingYear	"1776"
y:Reese-Witherspoon	bornOnDate	"1976-03-22"
y:Reese-Witherspoon	bornIn	y:New_Orleans_LA
y:Reese-Witherspoon	hasName	"Reese Witherspoon"
y:Reese-Witherspoon	gender	"Female"
y:Reese-Witherspoon	title	"Actress"
y:New_Orleans_LA	foundingYear	"1718"
y:New_Orleans_LA	locatedIn	y:United_States
y:Franklin_Roosevelt	hasName	"Franklin D. Roosevelt"
y:Franklin_Roosevelt	bornIn	y:Hyde_Park_NY
y:Franklin_Roosevelt	title	"President"
y:Franklin_Roosevelt	gender	"Male"
y:Hyde_Park_NY	foundingYear	"1810"
y:Hyde_Park_NY	locatedIn	y:United_States
y:Marilyn_Monroe	gender	"Female"
y:Marilyn_Monroe	hasName	"Marilyn Monroe"
y:Marilyn_Monroe	bornOnDate	"1926-07-01"
y:Marilyn_Monroe	diedOnDate	"1962-08-05"

三元组形式



图形式

逻辑视图：RDF or Property Graph

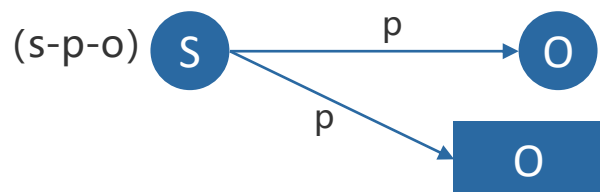
RDF Graph

结点

实体
literal (字面量)

边

关系 (实体->实体)
属性 (实体->属性值)



结点和边都没有内部结构

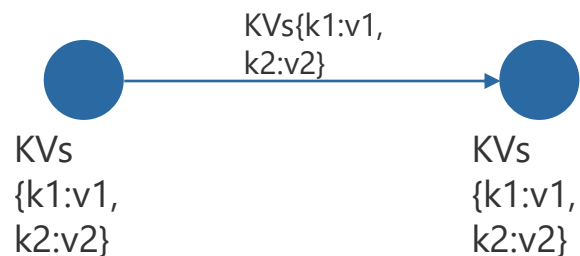
Property Graph

结点

实体
key-value pairs

边

关系 (实体->实体)
key-value pairs



结点和边都可以有内部结构，用以表达实体/关系的属性

图数据库的发展

Graph 1.0: 小规模原生图存储 (2007年-2010年)

以Neo4j为代表, 采用原生 (Native) 的方式实现了图存储, 获得了比关系型数据库快得多的复杂关联数据查询性能。然而, 当时在软件架构设计上只支持单机部署, 性能受制于单机的配置, 业务扩展能力也有限。

Graph 2.0: 分布式大规模图存储 (2010年-2016年)

扩展性成为业界的痛点, 诞生了诸如OrientDB, Titan 这样的项目。在这一发展阶段, 支持分布式大规模图存储是关注的重点, 图存储是否以原生的方式实现, 变成了不是那么首要的问题 (OrientDB选择了支持原生图, 而JanusGraph则是在底层数据库之上封装了一层, 实现了图的语义)。

Graph 3.0: 大规模“存储+运算”一体化 (2016年-至今)

随着人工智能的飞速发展, 对运算的需求变得越来越多, 图数据库也不再仅仅满足于图存储, 分布式大规模“存储+运算”一体化成为趋势。在这一阶段, 我国科技迎头赶上并实现弯道超车, 诞生了许多集原生、分布式、存储+运算于一体的优质图数据库项目。

图数据库 VS 关系型数据库 VS NoSQL数据库

图数据库与关系型数据库对比

分类	关系型数据库	图数据库
模型	表结构	图结构
存储信息	高度结构化数据	结构化/半结构化数据库
2度查询	低效	高效
3度查询	低效/不支持	高效
空间占用	中	高

其他NoSQL数据库与图数据库对比

分类	图数据库	键值数据库	列存储数据库	文档型数据库
数据模型	图	哈希表	列式数据存储	键值对扩展
优势	利用图结构相关算法(最短路径、节点度关系查找等)	查找速度快	查找速度快; 支持分布横向扩展; 数据压缩率高	数据结构要求不严格; 表结构可变; 不需要预先定义表结构
劣势	分布式存储比较困难	数据无结构化, 通常只被当作字符串或者二进制数据	功能相对受限	查询性能不高, 缺乏统一的查询语法
举例	Neo4j、JanusGraph	Redis	HBase	MongoDB

主流产品

	阿里云GDB	阿里云Graph Compute	百度云BGraph	蚂蚁GeaBase	Neo4j	Janus Graph	华为云GES
定位	图数据库	查询分析一体化平台	图数据库	图数据库	图数据库	图数据库	查询分析一体化平台
商用情况	公测	公测	技术合作咨询	公测	线下销售	开源	商用
最大规模	取决于数据盘大小	千亿边	数亿节点千亿边	万亿边	取决于单机数据盘大小 (百亿~千亿)	千亿边	万亿边
支持语言	Gremlin	Gremlin	Gremlin	自实现语言	Cypher	Gremlin	Gremlin
性能	点边查询2w TPS 多跳查询与Neo4j相当	三跳查询毫秒级响应	数万TPS	单机点边查询80w TPS 三跳查询毫秒级响应	三跳查询毫秒级响应	三跳查询响应超时	点边查询2w TPS 多跳查询1X倍优于Neo4J
算法能力	无	通过Gremlin集成了cc, pagerank等算法	对Gremlin进行扩展, 支持特色的高性能图算法	无	部分简单算法, 最新版本对接算法库获取算法能力	无	集成了丰富的内置算法 (20+) 且性能X倍优于Powergraph
事务能力	支持	无	无	无	支持	无	无
优势	完整事务能力支持 扩容方便	深度优化的Gremlin 弹性伸缩	扩展Gremlin, 特色图算法	适合金融领域, TPS性能优秀	商业成功, 软件成熟	开源	丰富且性能优秀的内置算法, 高性能原生查询接口

华为云-图存储计算引擎

典型场景

<p>政企 大中小企业单位</p> <ul style="list-style-type: none"> IT资产管理运维 供应链管理 职员权限关系网管理 	<p>金融 银行、保险</p> <ul style="list-style-type: none"> 洗钱套现挖掘 异常交易挖掘 贷前风险预判 诈骗团伙发现 	<p>安平 公安</p> <ul style="list-style-type: none"> 人员全息档案构建 嫌疑人多维关系审查 同行同住 轨迹分析追踪 犯罪团伙挖掘 	<p>互联网 社交、电商</p> <ul style="list-style-type: none"> 社交网络推荐 电商商品个性化推荐 用户360画像 反欺诈
---	---	---	---

大规模

百亿节点，万亿边规模大图

高性能

单实例2W+QPS，秒级响应

查询分析一体化

- 一份数据做两件事：查询和分析一体化
- 25+高性能算法，进行多场景数据挖掘

简单易用

- 向导式分析+标准Gremlin查询语言
- 原生REST API
- 可视化编辑和图分析探索



多次斩获国际大奖



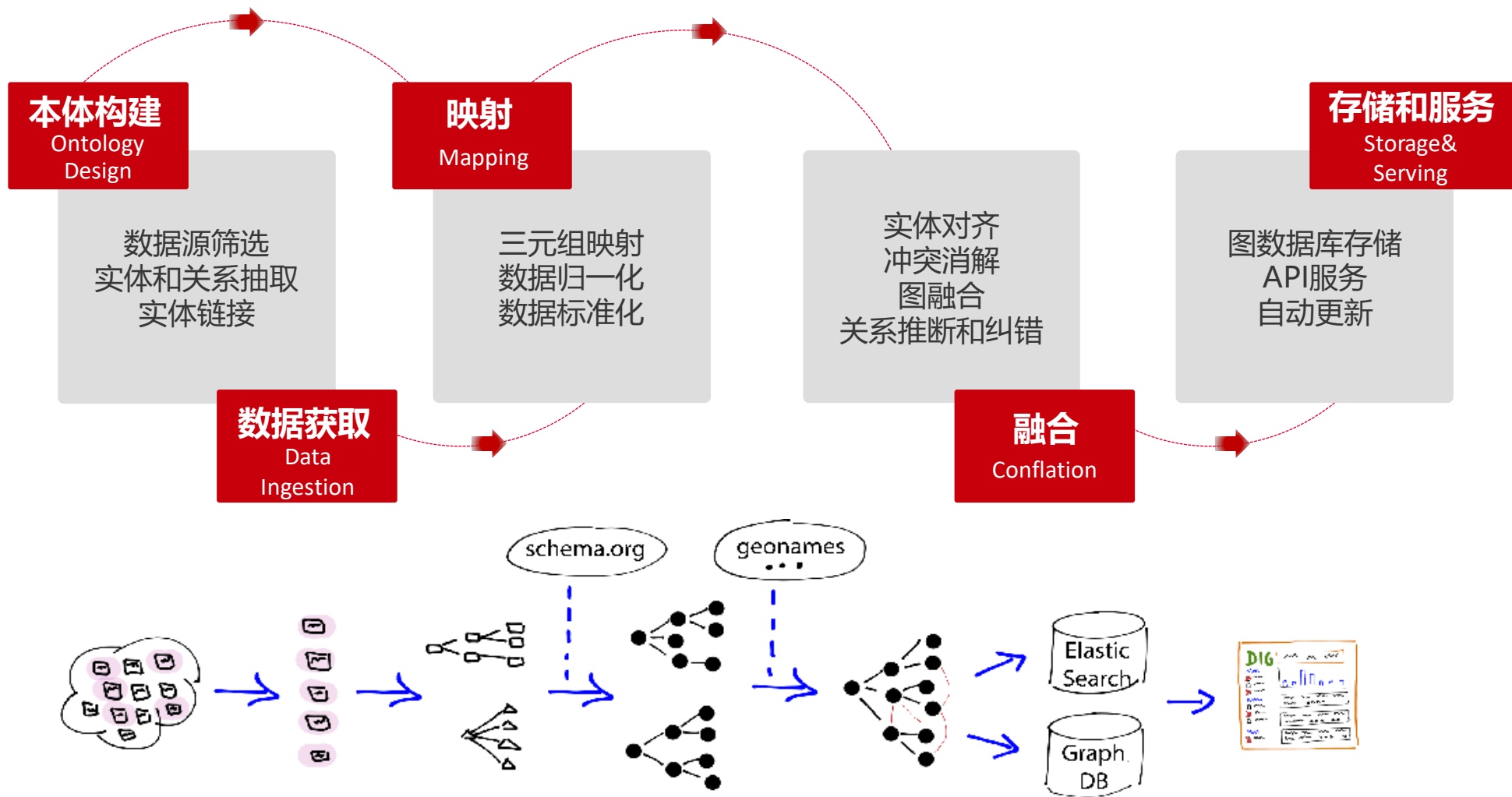
2018数博会双料大奖



2019人工智能峰会最高奖

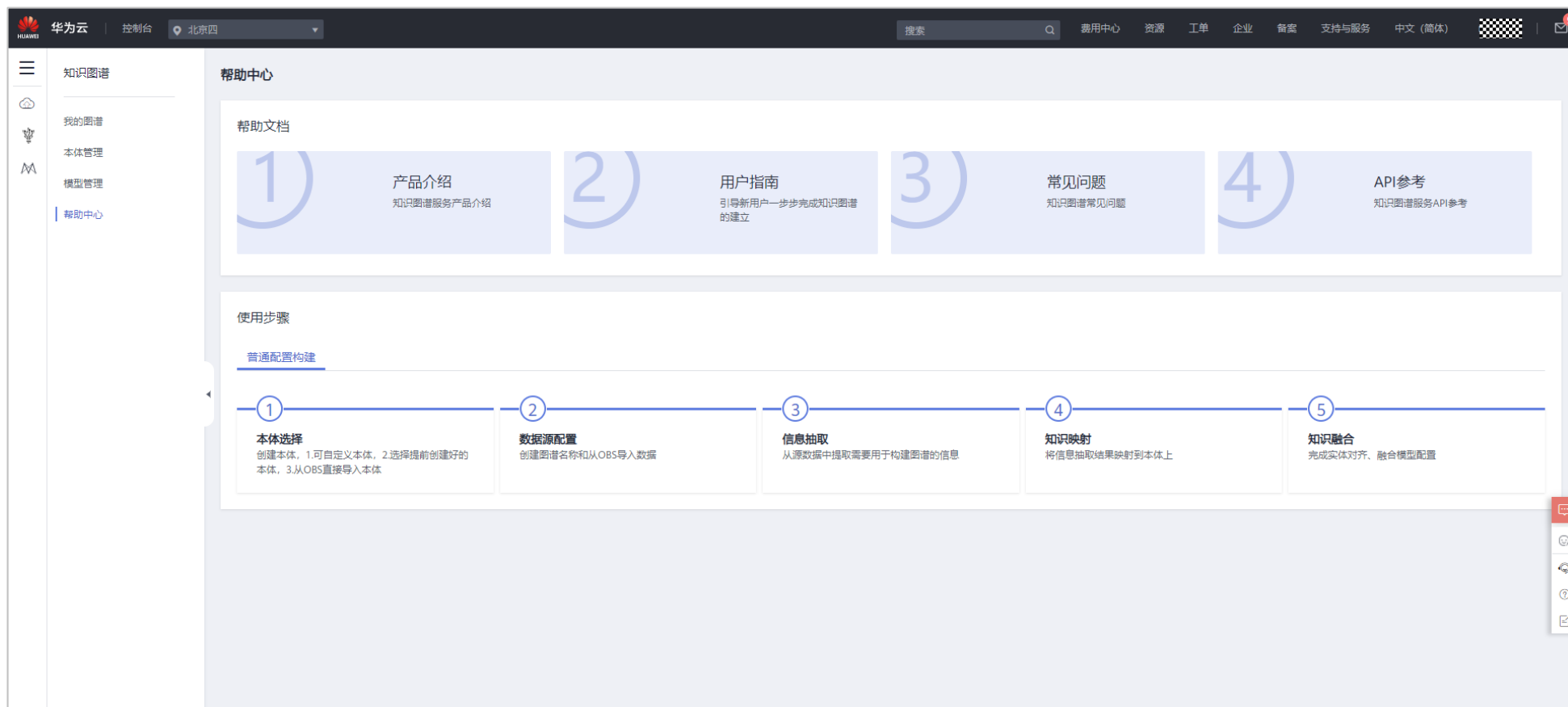
知识图谱构建与应用

知识图谱构建流程

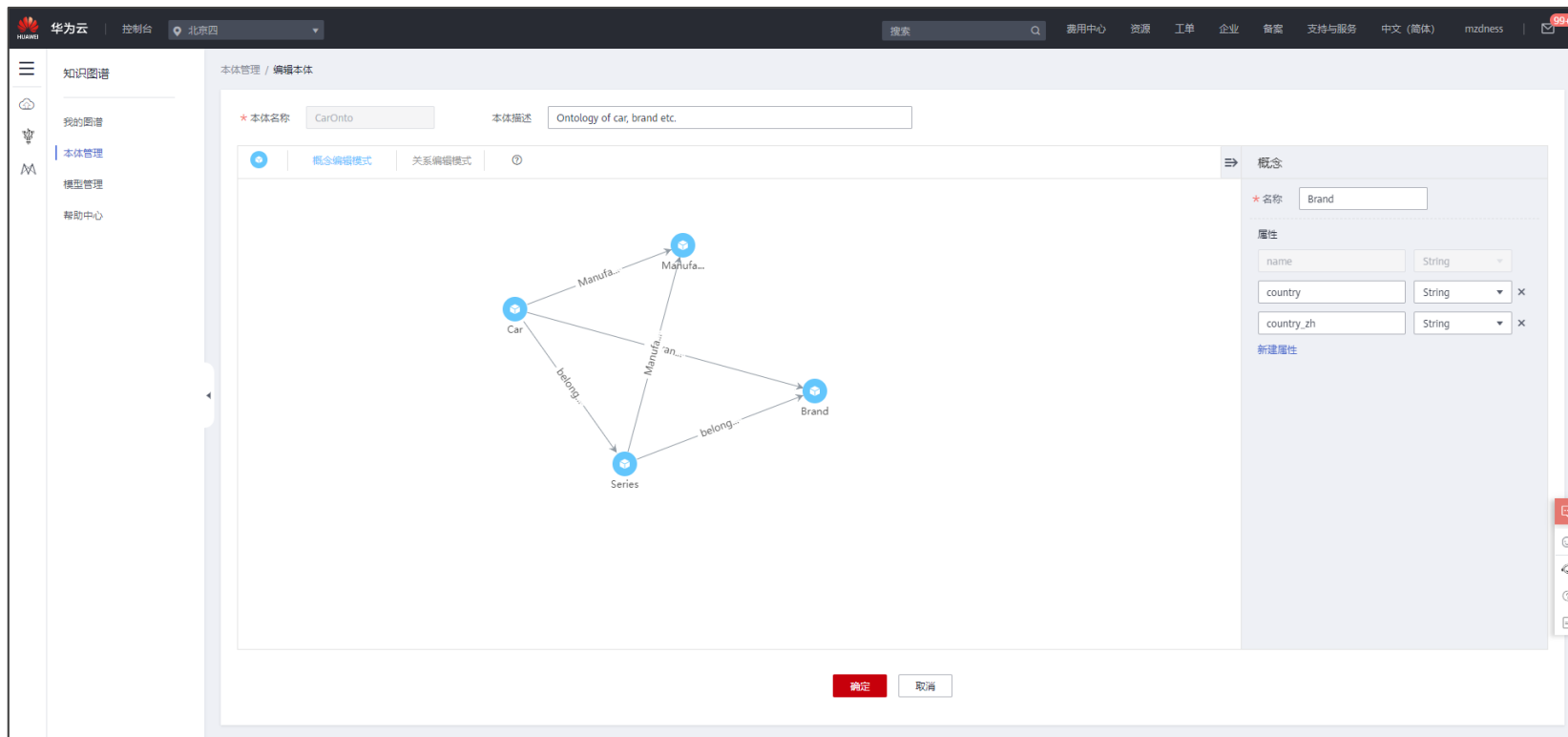


华为云知识图谱云服务

基于华为云图引擎数据库（GES）、华为云AI平台（ModelArts）以及华为云大数据平台等，提供企业知识图谱平台服务，可以快速开发、定制属于企业客户的知识图谱及应用。



本体/Schema设计



1

图形化编辑

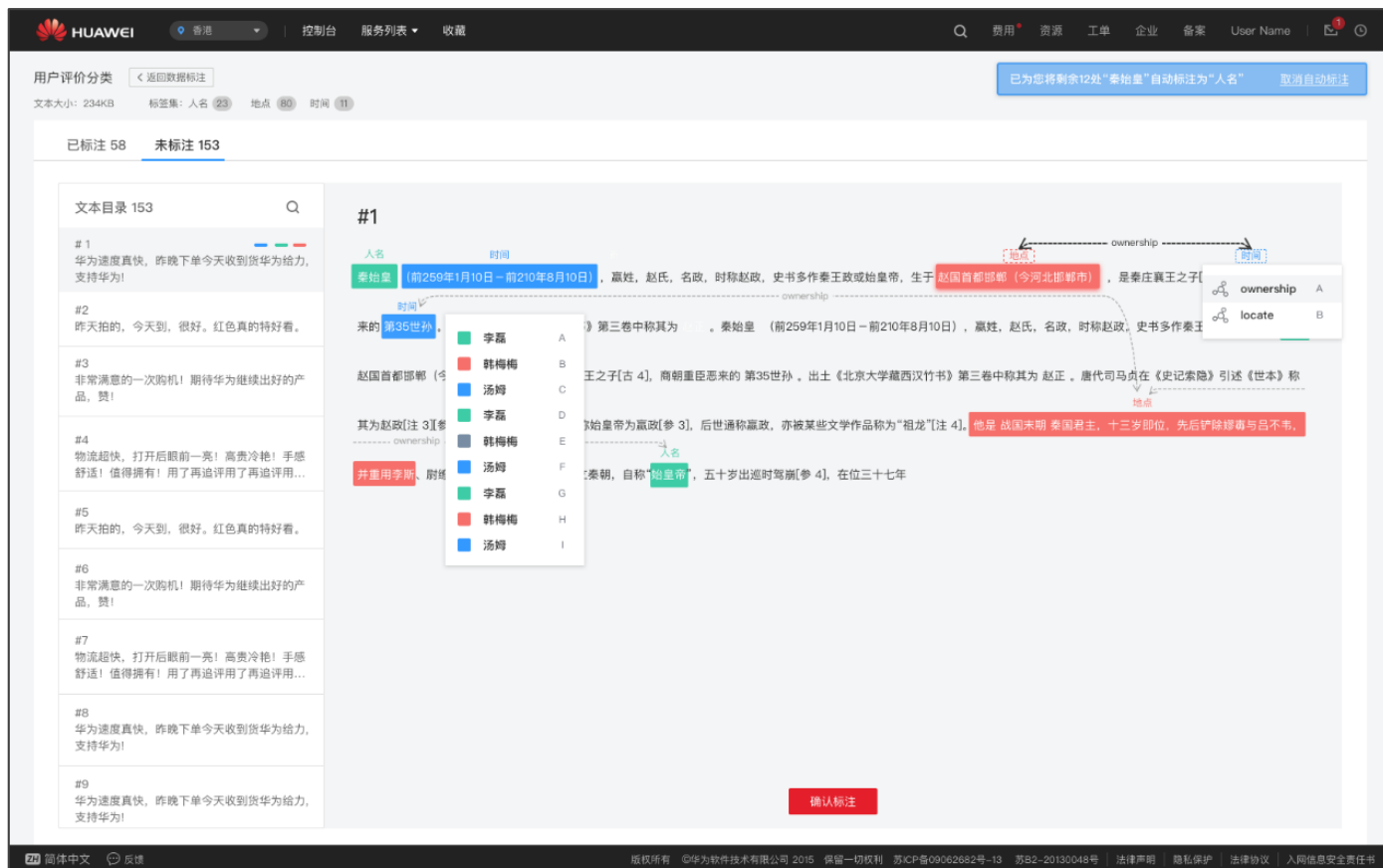
托拉拽即可完成本体设计，所见即所得，高效、直观

2

本体导入导出

支持本体配置文件导出
支持通过配置文件创建本体

多种文本标注工具



实体标注

框选标注实体, 高效、直观, 简单易用



三元组标注

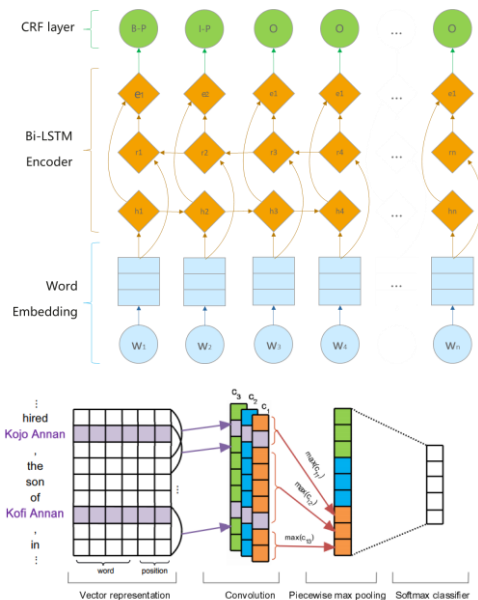
框选标注实体, 拖拽标注关系, 高效、直观, 简单易用



团队协同标注

支持团队多人协同标注, 提升协同标注效率

信息抽取算法

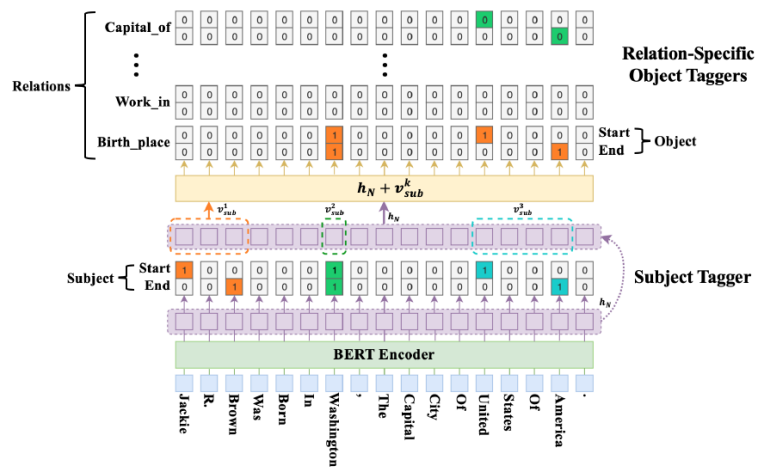


传统流水线式抽取 NER+PCNN

1

丰富的抽取算法

传统流水线式抽取
联合学习抽取
小样本抽取

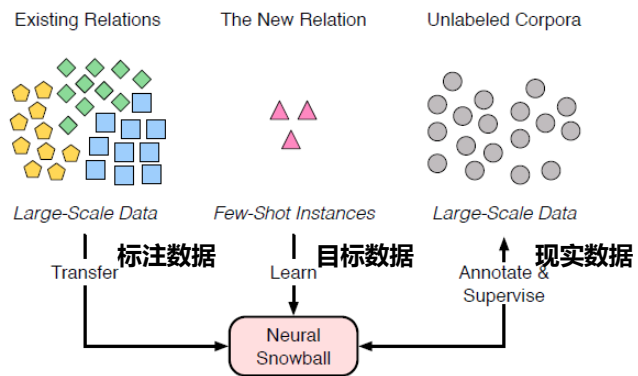


联合式抽取 HBT框架

2

适用不同场景

联合式抽取效果好，标注多
小样本抽取标注少，效果略低



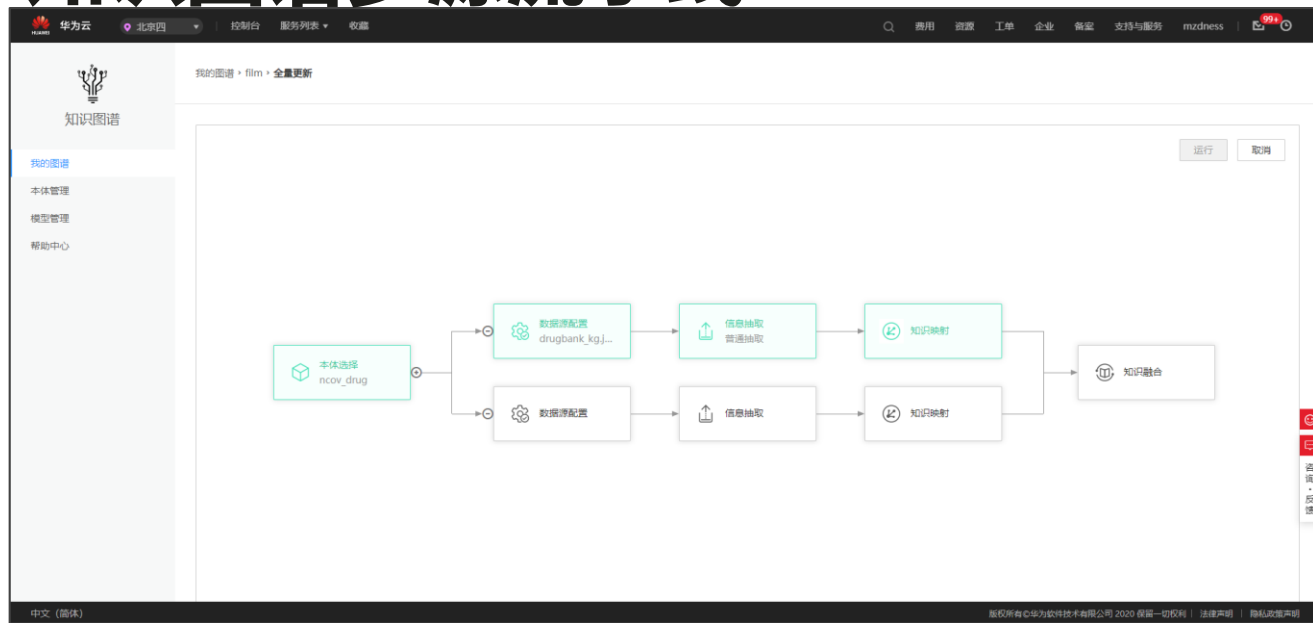
小样本抽取

3

算法模型效果

降低人工60%标注量
效果接近最佳模型

知识图谱多源流水线



The screenshot displays the '我的图谱' (My Knowledge Graph) interface for a graph named 'gene_demo'. It features a dashboard with the following components:

- 图谱ID:** 95c5619b-146d-46da-ae80-cf193e807117
- 创建时间:** 2019-10-10 14:20:40
- 最后更新时间:** 2020-02-26 16:26:17
- 图谱预览:** A visual representation of the knowledge graph.
- 流水线预览:** A visual representation of the data pipeline.
- 图谱统计数据:**

实体个数	关系个数
217070	449770
- 版本统计:**

上线/未上线
1/4
- Job统计数据:**

发布次数	全量更新次数	增量更新次数
3	14	-

Below the dashboard is an '更新日志' (Update Log) section with a table of updates:

运行编号	事件类型	本体	状态	操作时间	操作
z1fee8d9-88fe-4be2-857a-da87a28bf938	全量更新	changqing_v1	已完成	2020-02-26 16:26:17	查看
00ddcdf5-9033-40c-9045-24bc924f9859	图谱发布	changqing_v1	已完成	2020-02-18 09:29:06	查看
0df440db-4da5-47d3-888a-b490db8ce31d	全量更新	changqing_v1	已完成	2020-02-18 09:22:39	查看
bd4fcadb-01ed-44c5-a634-ed6813c2362c	图谱发布	changqing_v1	已完成	2020-02-18 01:31:01	查看
ef42da5d-ced4-4ebb-9e80-e67bc8343f90	全量更新	changqing_v1	已完成	2020-02-18 01:26:41	查看
137d7bbc-3970-47ac-8e6f-a8f36ebfa6f8	全量更新	changqing_v1	已完成	2020-02-18 01:10:40	查看



多源流水线

支持多源数据以流水线式自动构建图谱



知识融合

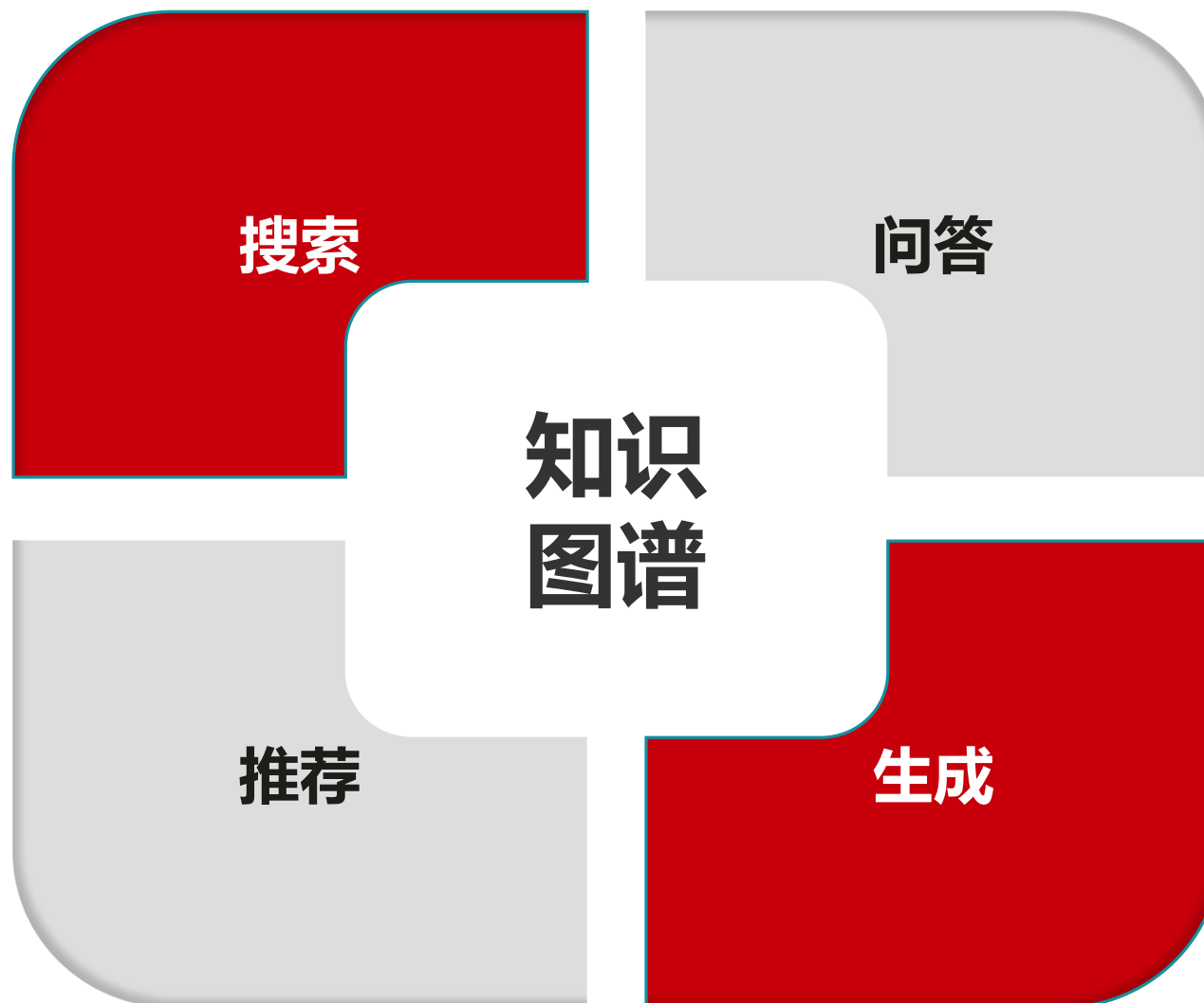
支持灵活的知识融合策略配置，可面向不同场景实现多源融合



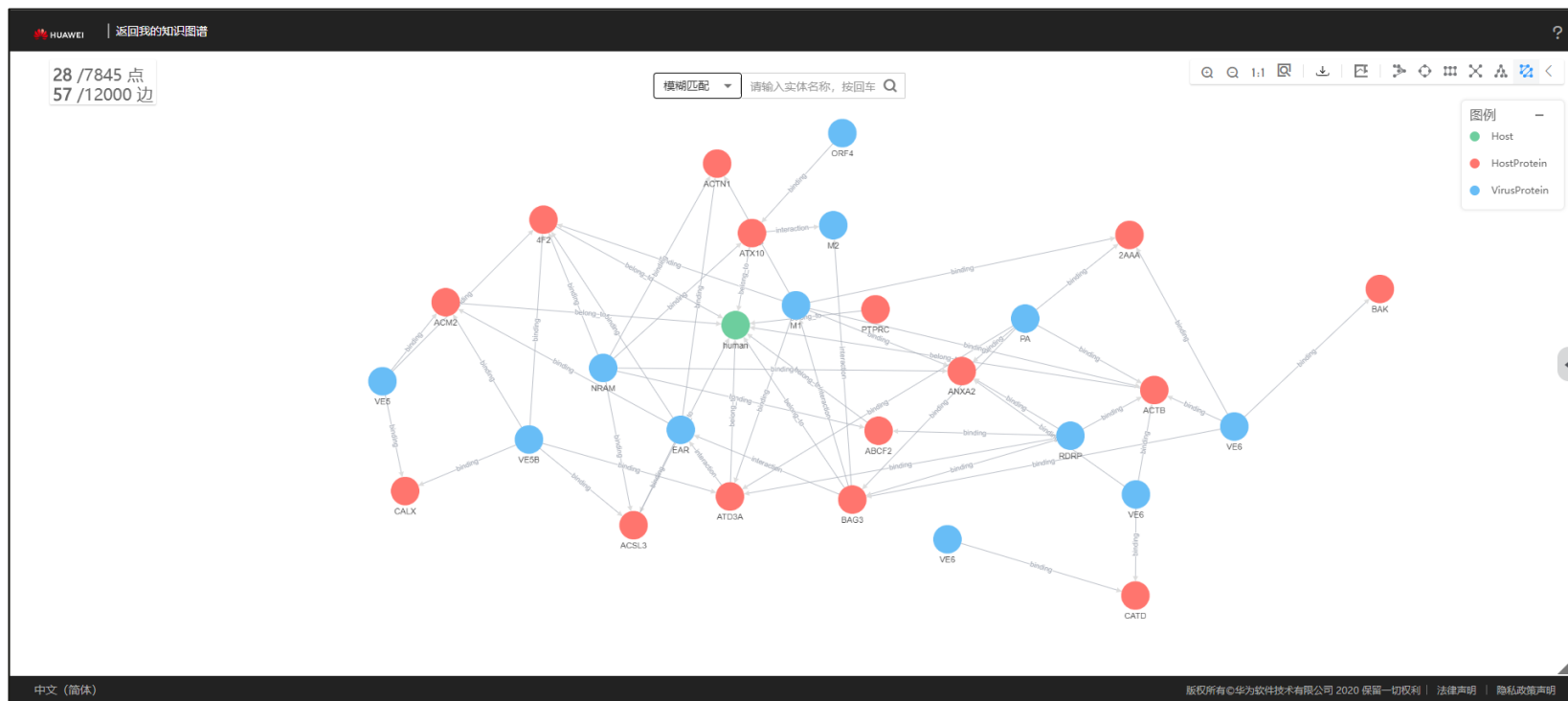
增量更新

支持自动化图谱增量更新

知识图谱应用



知识应用-查询可视化



1

多种知识应用

图谱查询及可视化
知识图谱问答
知识推理
实体链接

2

开箱即用

每个图谱都有自己的知识应用
图谱构建完成, 一键开启应用

3

支持接口调用

提供Rest APIs调用, 支持二次嵌入
业务系统, 快速对接集成

知识应用-KBQA

The screenshot displays the Knowledge Graph KBQA interface. On the left, there is a sidebar with navigation options: 知识图谱, 我的图谱, 本体管理, 模型管理, and 帮助中心. The main area shows a knowledge graph preview and a list of services, including '知识图谱问答KBQA' and '实体链接Entity Link'. A central '问答体验' (Q&A Experience) window is open, showing two example questions and answers. The first question is '奥迪A4L的颜色有哪些' (What are the colors of the Audi A4L?), and the answer is '奥迪A4L的颜色是朱鹭白,白金色,探戈红,花剑银,季风灰,探索蓝,传奇黑。' (The colors of the Audi A4L are Zhumu White, Platinum, Tango Red, Flower Sword Silver, Monsoon Grey, Exploration Blue, and Legend Black.) The second question is '奥迪A4L的制造商是一汽-大众奥迪。' (The manufacturer of the Audi A4L is FAW-Volkswagen Audi.), and the answer is '实体:' (Entity:). Below the Q&A window, there is a text input field and a '发送' (Send) button. To the right, there are two panels showing API examples. The top panel, titled '示例' (Example), shows a request example with a POST request to a specific URL and a JSON body containing a query. The bottom panel shows a response example with a JSON body containing an answer and a frame.

我的图谱 > 汽车图谱_v1

知识图谱

我的图谱

本体管理

模型管理

帮助中心

图谱ID: bda47db8-2c73-4926-a...

图谱预览

图谱版本 更新日志 图谱应用

服务

知识图谱问答KBQA
基于知识图谱中的知识提供问答服务

状态 已开通

实体链接Entity Link
识别句子中出现的知识图谱实体

状态 已开通

问答体验

奥迪A4L的颜色有哪些

奥迪A4L的颜色是朱鹭白,白金色,探戈红,花剑银,季风灰,探索蓝,传奇黑。
实体: 奥迪A4L

查看json

制造商呢

奥迪A4L的制造商是一汽-大众奥迪。
实体:

查看json

请输入内容

发送

示例

请求示例

```
POST https://nlp-ext.cn-north-4.myhuaweicloud.com/v1/1604d8170cd84440a81281e5a3a04f8b/kg/kg-:
{
  "query": "小明的体重"
}
```

Request Header:
Content-Type: application/json
X-Auth-Token: MIINRwYJKoZIHvcNAQcCoIIN0CCDTQCAQExDtALBg1ghkgBZQMEAgEwggUVBgkqhkiG...

响应示例

成功响应示例

```
{
  "answer": "小明的体重是73。",
  "frame": [
    {
      "query": "g.V('6939670486880563336').valueMap('体重')",
      "results": [
```

✓ 开箱即用的KBQA, 提供Rest APIs, 支持二次封装

企业知识计算案例

案例：新冠科研图谱

截至 2020-11-11 09:50:26 数据统计

较上日+200197

14210538

现有确诊

较上日+438135

51587390

累计确诊

较上日+230859

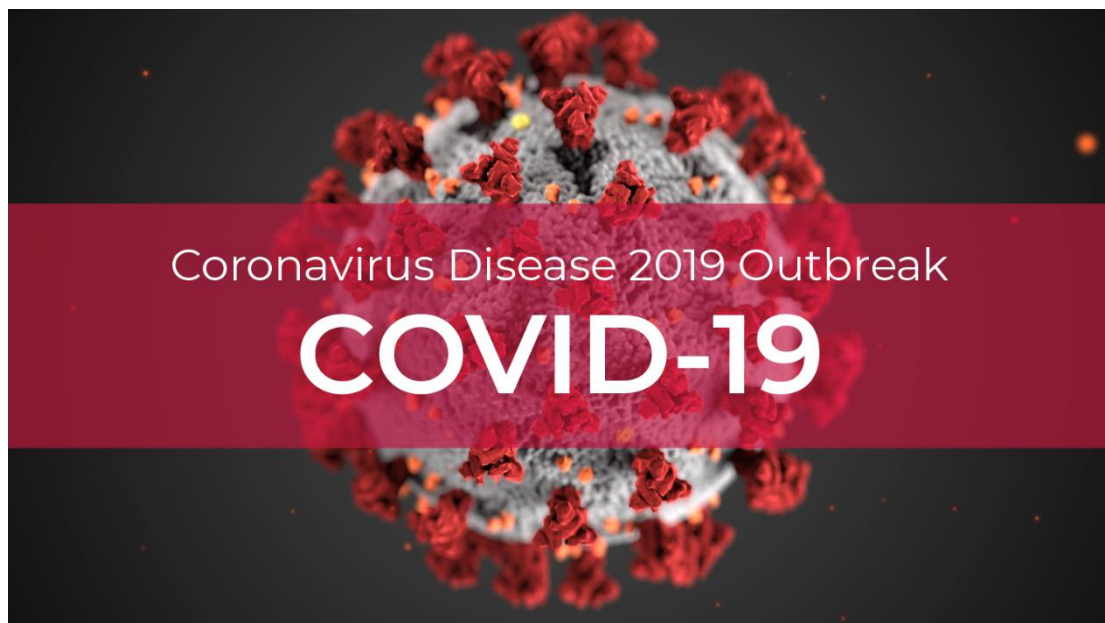
36106080

累计治愈

较上日+7079

1270772

累计死亡

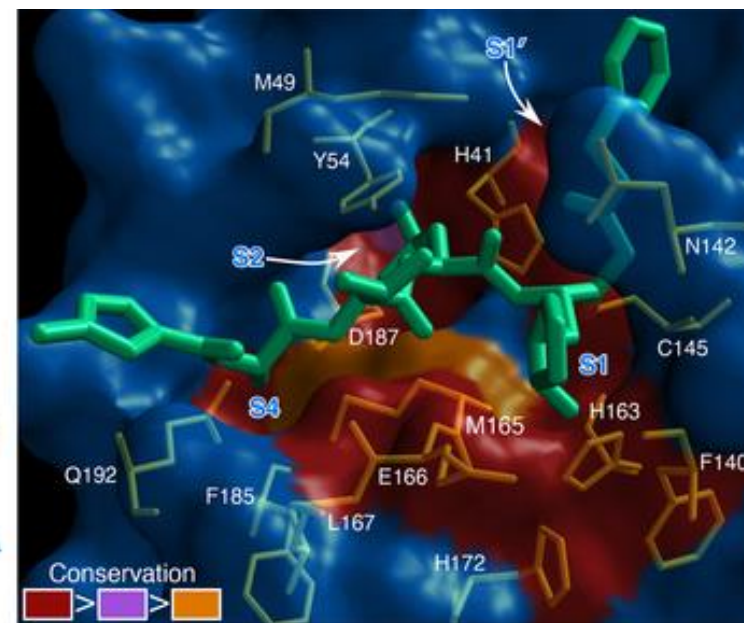
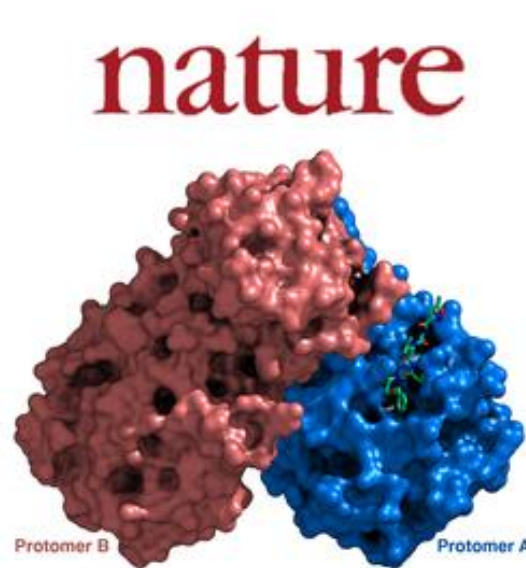


新型冠状病毒爆发

图片来源:

<https://www.diagnosticimaging.com/coronavirus/disinfecting-imaging-equipment-during-covid19-outbreak>

<http://www.shanghaitech.edu.cn/eng/2020/0410/c1260a51039/page.htm>



科学研究、药物研发

案例背景

新冠疫情日益严重，研究人员迫切需要能够系统梳理和整合新型冠状病毒的相关知识，加速对新冠病毒的机制研究和抗病毒药物研发。

信息来源多、更新快

10+ 医疗领域专业数据库，无有效关联
6000+ 篇新冠/冠状病毒论文，每日新增XX+
人工整合、知识检索及分析效率低

知识抽取融合难

已有算法模型效果不佳
未结合生物医学领域语料知识
无有效工具提供全流程图谱构建平台，标注成本高

缺少知识化服务

高效知识检索能力
交互式知识关联分析能力
基于AI+知识图谱的病毒-药物-蛋白关联预测等

整体框架



Ontology Design



Application

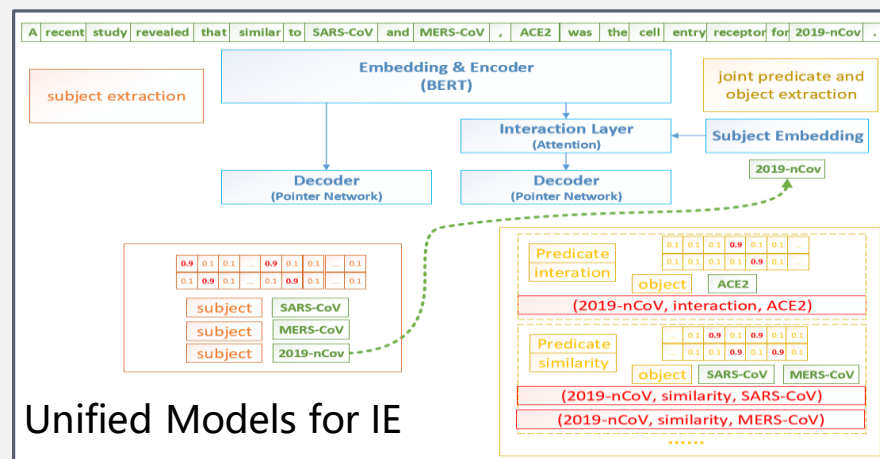
Open Scientific Dataset



Research Literature



Knowledge Extraction for Structure Data



Manual Labeling

Manual Check

Drug Discovery



Genetic Diagnosis Report Generation



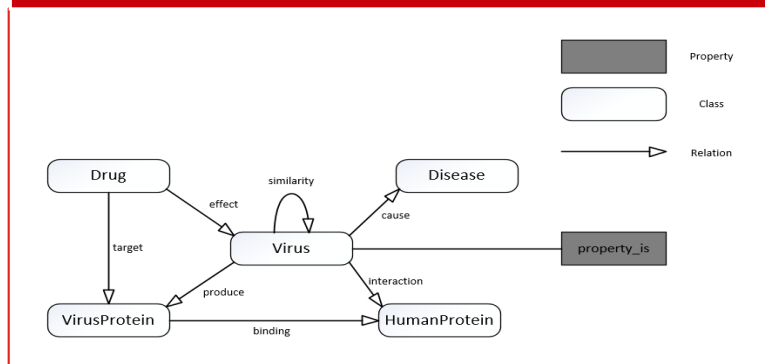
Junior Doctor Training



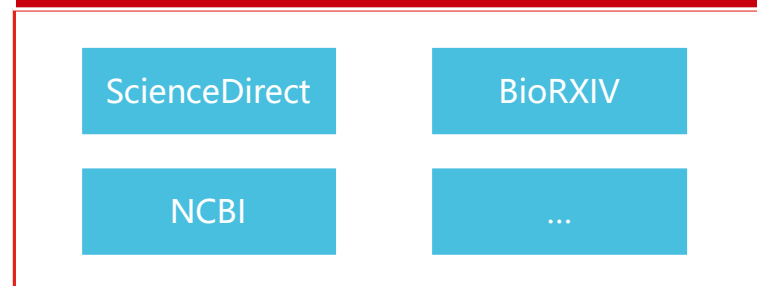
2020-02-10, 通过OpenKG首次发布病毒、蛋白、药物知识图谱, 迭代更新多个版本, <http://www.openkg.cn/dataset/covid-19-research>

准备工作

Schema/本体



数据源



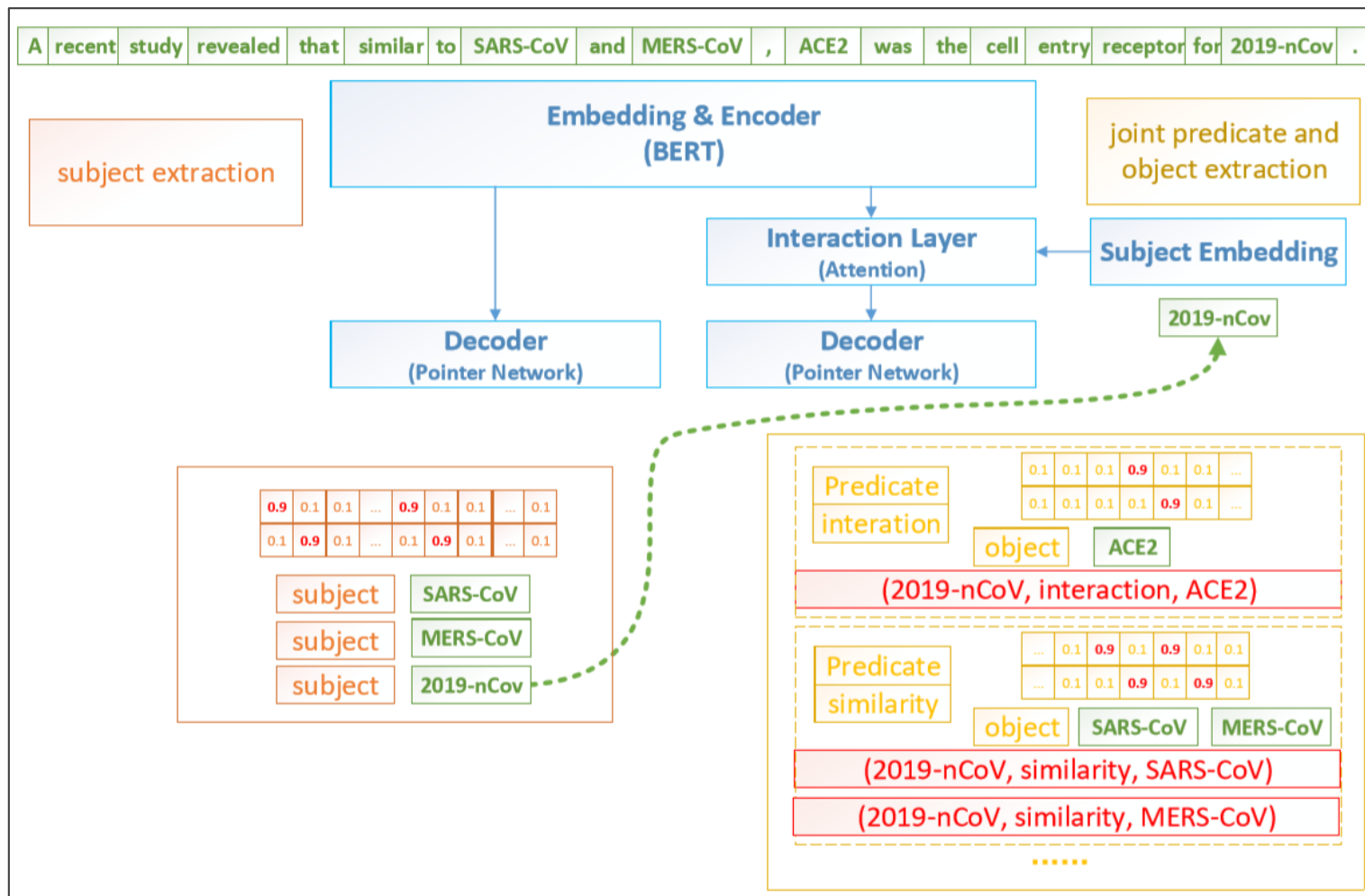
数据处理



样本标注

- ✓ (SARS-CoV-specific monoclonal antibodies, effect, 2019-nCoV)
- ✓ (lopinavir, target, 2019-nCoV M Pro)
- ✓ (2019-nCoV, similar, SARS-CoV M Pro)

三元组联合式抽取

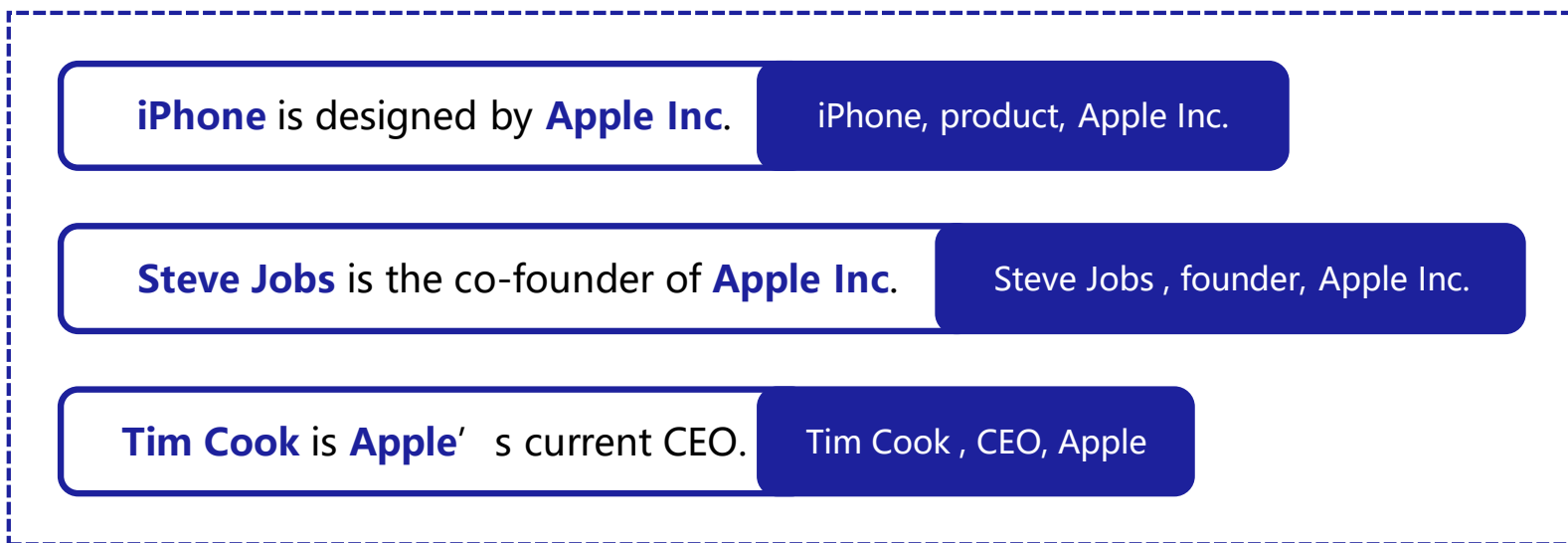


- ✓ **医疗文献中抽取实体关系的模型**
- **主语抽取部分(subject extraction)**
- **谓词宾语联合抽取部分(joint predicate and object extraction)**
- **多轮问答模板**
- **机器阅读理解MRC框架**

Li X, Yin F, Sun Z, et al. Entity-Relation Extraction as Multi-Turn Question Answering[C]. ACL 2019.

小样本学习优化

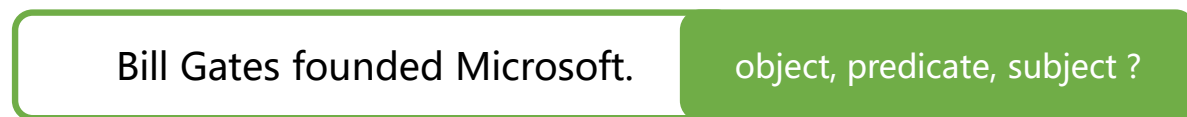
Supporting Set: 3 way 1 shot



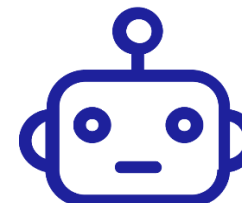
- ✓ 标注周期长, 效率低
- ✓ 模型预标注+人工矫正
- ✓ 小样本学习

Training

Query Set:



Querying

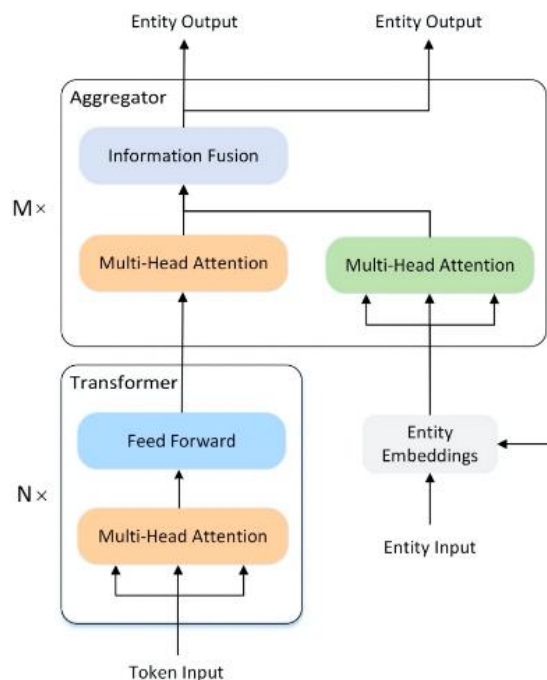


Decoding

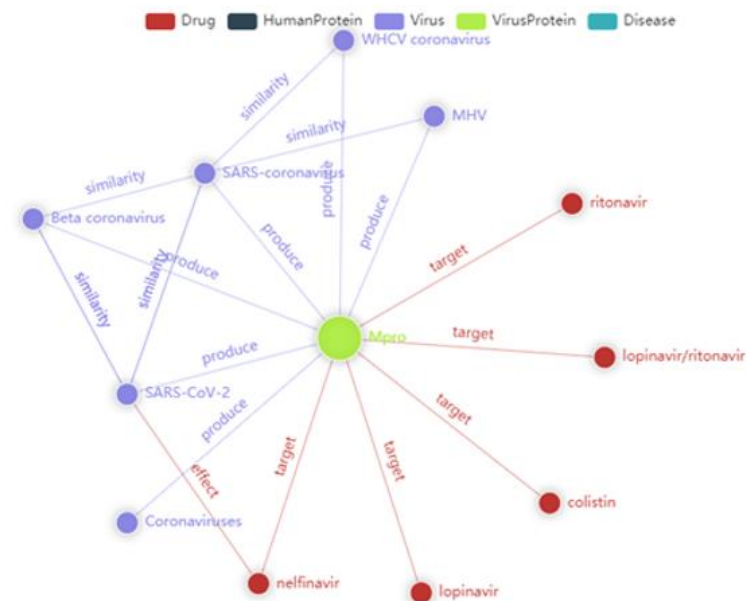
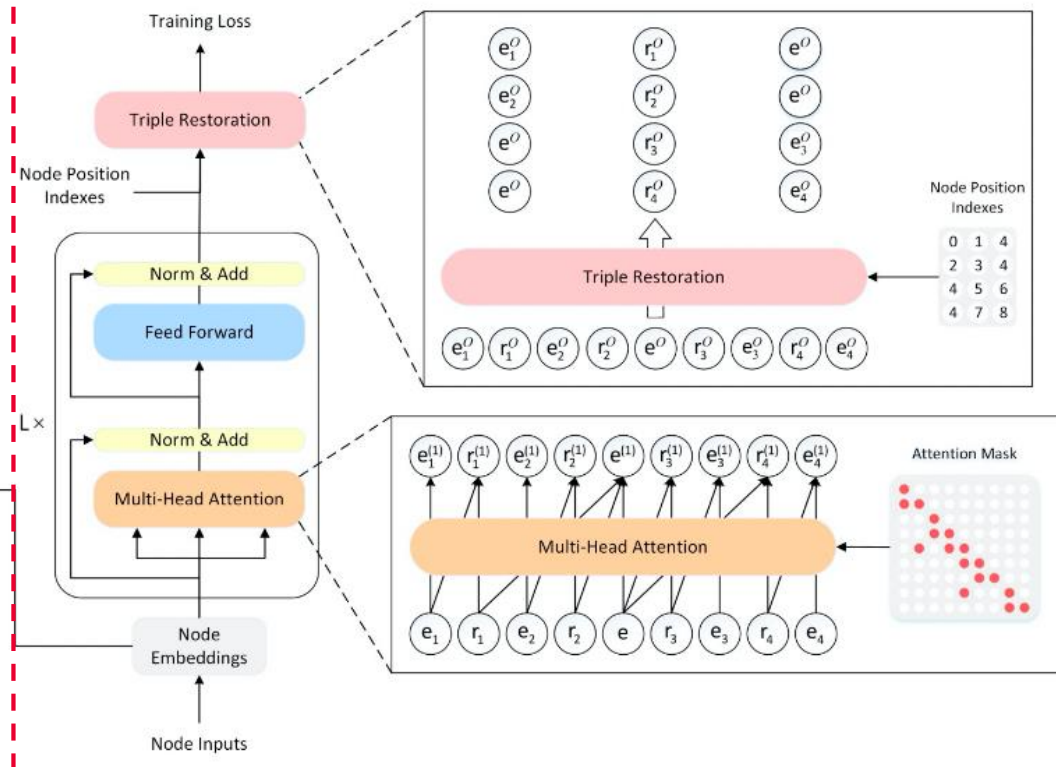
预训练模型优化

挑战：如何将医疗领域知识融入预训练模型？

集成Entity Embedding训练 BERT-MK



医疗知识图谱预训练Entity Embedding



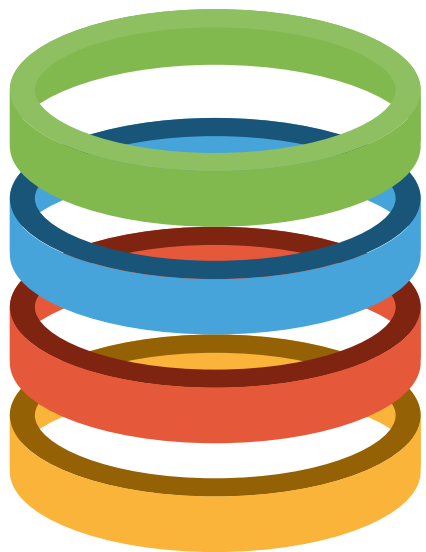
BERT-MK 医疗预训练模型

Integrating Graph Contextualized Knowledge into Pre-trained Language Models, EMNLP 2020

经验总结

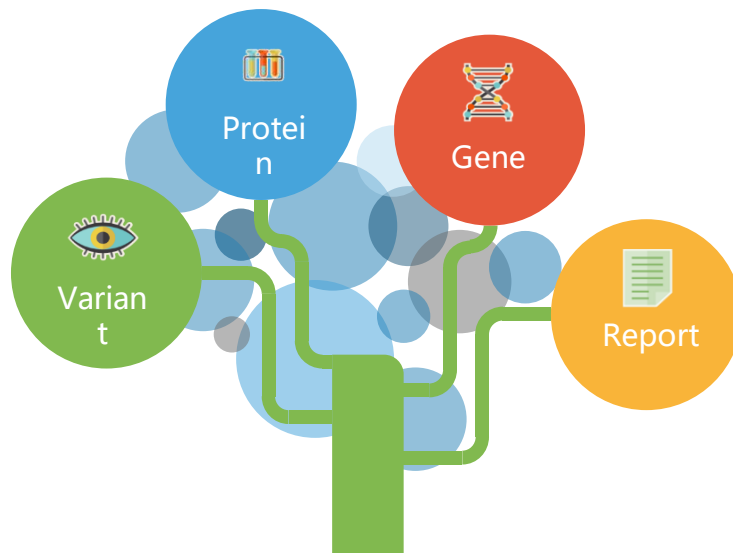
模型优化	1. 远程监督	远程监督训练实体抽取模型
	2. 数据增强	小样本学习, 模型预标注 + 人工矫正
	3. 预训练模型	BERT -> BERT-MK 医疗领域语料
迁移性	案例Pipeline适用于其他医疗图谱、甚至其他行业图谱构建	
工具便利性	Schema设计、数据标注都依赖领域专家、领域知识 为领域专家提供更高效率的工具	

案例：基因检测知识计算



数据孤立

数据散落在多个数据库或文件夹，查询费时费力，未有效整合



无知识体系

数据未以知识化形式有效组织和表达，未发挥出数据的价值



报告生成效率低

领域专家撰写检测报告，效率低，人员成本高

解决方案

方案描述

构建行业知识图谱，形成体系化知识，提供知识层面服务

多个数据库、数据文件的清洗与整合

实体关系查询、可视化、推理、全文搜索等服务

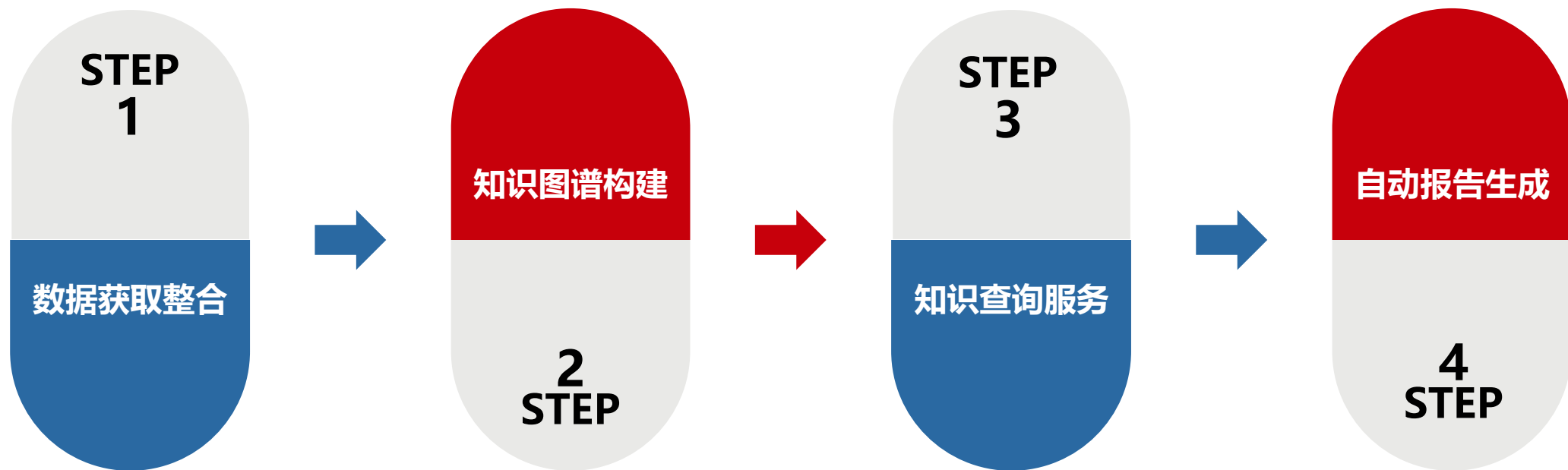
自动对基因及突变进行总结，形成100字以内的概括性解释报告

客户价值

基因解释报告生成效率提升XX%

数据打通，知识化整合，提升业务查询效率和效果

知识共享、知识沉淀



基因检测解释报告自动生成

Application_name		基因检测	
Document_name		基因检测结果	
Config	personality	neutral	
positiveGeneMutations	Gene	EGFR	TP53
	mutantFrequency	7.5	2
	Exons	21	8
	AAChangeRefgene	NM_005228.3:c.2573T>G(p.L858R)	NM_000546.5:c.817C>G(p.R273G)
	mutationSpot	L858R	splicing



肺癌热点突变和融合基因检测解释报告

- 1 本样本经病理评估肿瘤组织含量为5%;
- 2 本样本EGFR基因检测到突变,位于21外显子,突变命名为NM_005228.3:c.2573T>G(p.L858R),突变频率为7.5%;EGFR基因L858R突变提示晚期非小细胞肺癌(NSCLC)患者可能对EGFR-TKIs类药物(如吉非替尼,厄洛替尼,阿法替尼,埃克替尼,奥希替尼)靶向治疗敏感;
- 3 本样本TP53基因检测到突变,位于8外显子,突变命名为NM_000546.5:c.817C>G(p.R273G),突变频率为2%;根据TCGA的研究,TP53基因突变在肺腺癌中的发生率为46%,肺鳞癌中的发生率为90%;肿瘤组织发生TP53基因突变的患者预后可能较差;目前尚没有针对TP53基因突变的靶向药物获批用于肿瘤治疗;
- 4 本样本未检测到ALK,ROS1,RET及NTRK1基因融合;提示肺癌患者可能对克唑替尼、卡博替尼等药物治疗不敏感;\n5. 请结合其它临床信息及其它检测结果综合使用本结果。

案例：汽车领域知识计算



汽车领域知识计算解决方案



流水线式
自动构建知识图谱

本体构建 数据源配置 信息抽取 知识映射 知识融合



数据驱动
数据结构化效率提升30%

Text in

红旗 H7 的胎压监测系统是由 PEPS、胎压低频天线、胎压传感器、仪表及导航显示屏组成，用诊断仪读取故障码 B12EE

输入自然语言

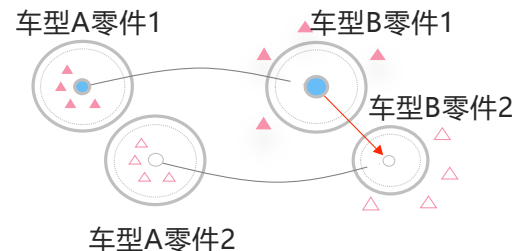
Data out

零部件 PEPS
零部件 胎压低频天线
零部件 胎压传感器
零部件 仪表及导航显示屏
故障码 B12EE

自动识别关键实体
形成结构化数据



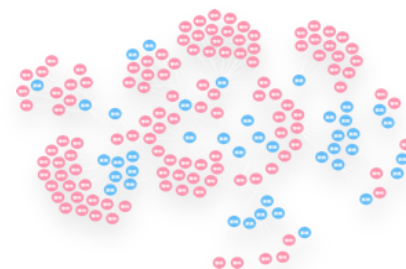
领域适配知识挖掘
智能知识推理



共件工况预测



自研图引擎
万亿级超高性能查询



客户省心

↑ 4% 一次修复率
↓ 23% 等待时间

厂家增效

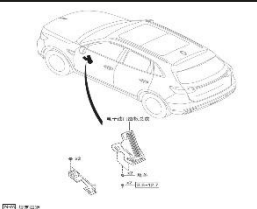
30% ↓ 厂家支持介入率
30% ↓ 专家培养周期

快速创建售后维修应用

知识图谱流水线, 快速构建知识图谱

故障维修案例报告

日期: 2019年7月11日					
设备名称	车型/品牌	车号	品牌	故障描述	故障类型
客户	姓名	性别	年龄	联系电话	联系地址
维修技师	姓名	工号	职称	所属部门	所属班组
故障描述: 客户反映车辆在行驶过程中出现异响, 经检查发现为发动机气门机构异响。					
维修建议: 建议更换气门机构相关部件, 并进行气门间隙调整。					
客户评价: 维修人员服务态度好, 技术专业, 维修效率高, 非常满意。					



维修手册

维修案例

销售公司	技术支持	服务支持
发件	技术支持	发布日期: 2019年7月25日
车型	编号	
联系人	联系电话	
主题		

培训课件

汽车发动机维修

1. 发动机总成修理工艺

2. 气缸体、气缸盖和曲柄连杆机构的修理






质量反馈信息

type: MananCase
name: fu f muse
客户反馈: 新车行驶里程达到1500公里以上, 发动机异响, 严重影响驾驶体验。建议厂家尽快提供技术支持, 解决质量问题。

车辆维修记录

客户名称	车牌号	车型	品牌	故障描述	维修内容
张三	京A1234	大众	朗逸	发动机异响	更换气门机构
李四	京B5678	本田	雅阁	变速箱异响	更换变速箱油

多源异构数据, 10+种

1 数据源配置

创建图谱名称和从OBS导入数据

2 本体选择

创建本体, 1.可自定义本体, 2.选择提前创建好的本体, 3.从OBS直接导入本体

3 信息抽取

从源数据中提取需要用于构建图谱的信息

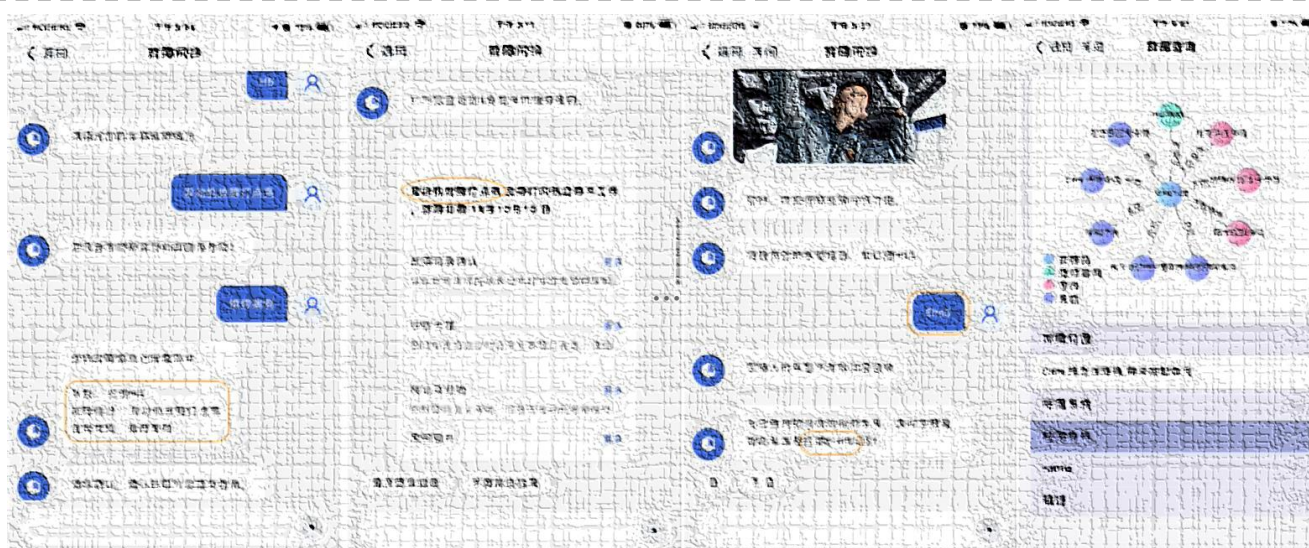
4 知识映射

将信息抽取结果映射到本体上

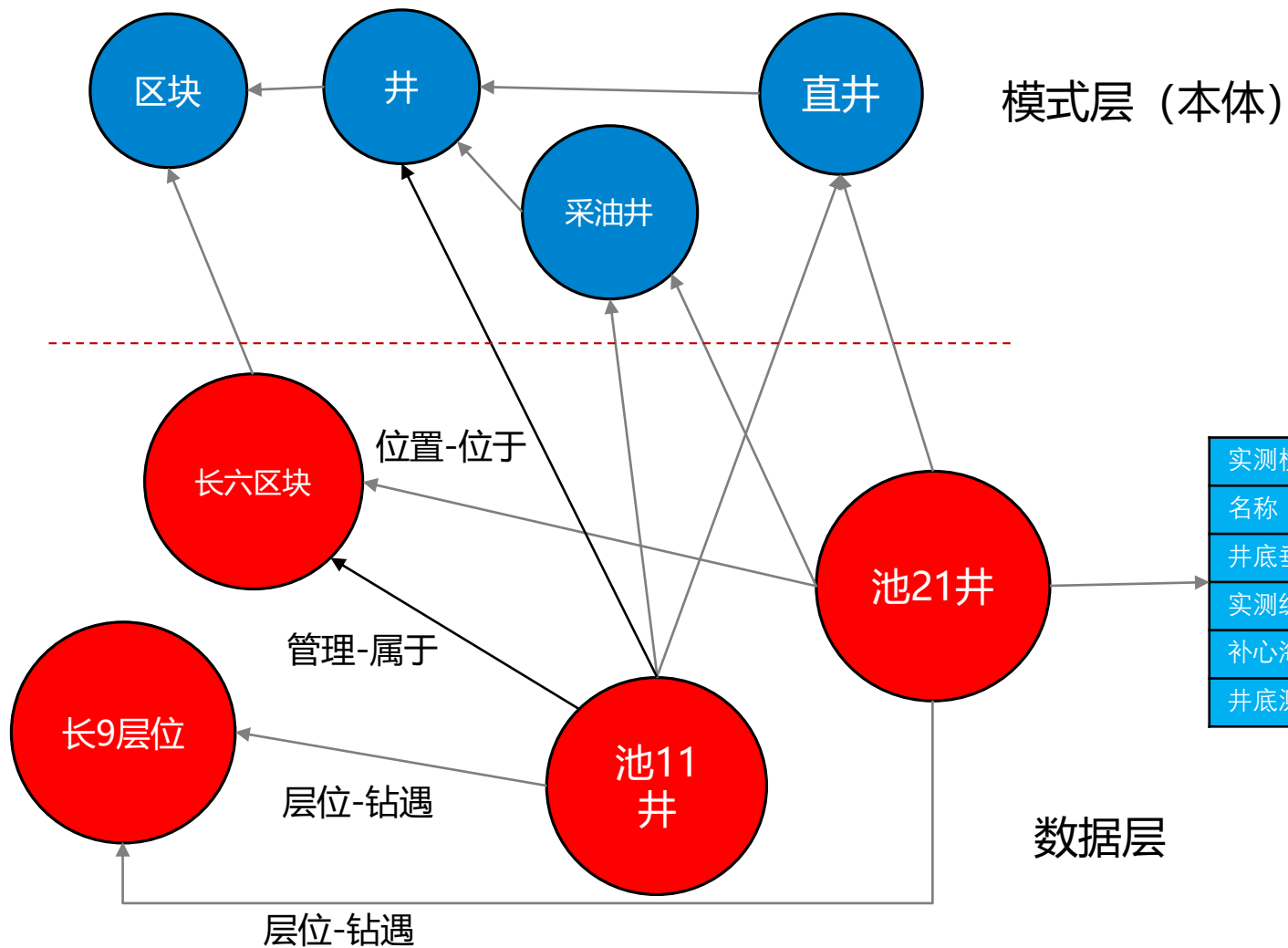
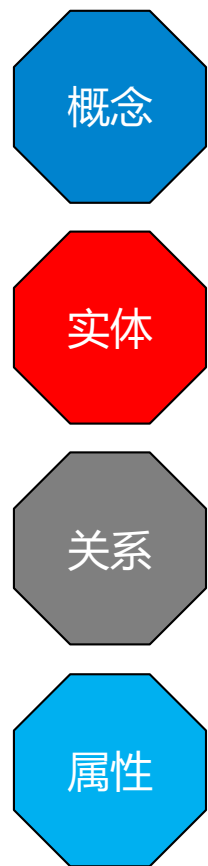
5 知识融合

完成实体对齐、融合模型配置

高效创建KBQA应用



案例：油气知识计算



实测横坐标	3.81980698645996E7
名称	池155
井底垂深	2710.0
实测纵坐标	5883120.49477415
补心海拔	1571.779
井底测深	2710.0

知识计算平台



油气数据



油气信息



油气知识



油气服务

油气智能大脑

油气数据预处理

- 油气数据接入
- 油气数据可视化
- 油气数据特殊处理
- 油气公开数据集



油气AI工厂

- 油气领域特征工程
- 油气领域算法及模型开发
- 油气领域算法及模型管理



油气知识图谱

- 结构化数据知识工程
- 非结构化数据知识工程
- 油气特殊数据知识工程



AI智能服务超市

- 通用领域AI能力
- 油气领域AI能力
- 油气领域智能应用

知识计算平台

油气层识别场景

地质勘探



物探



钻井



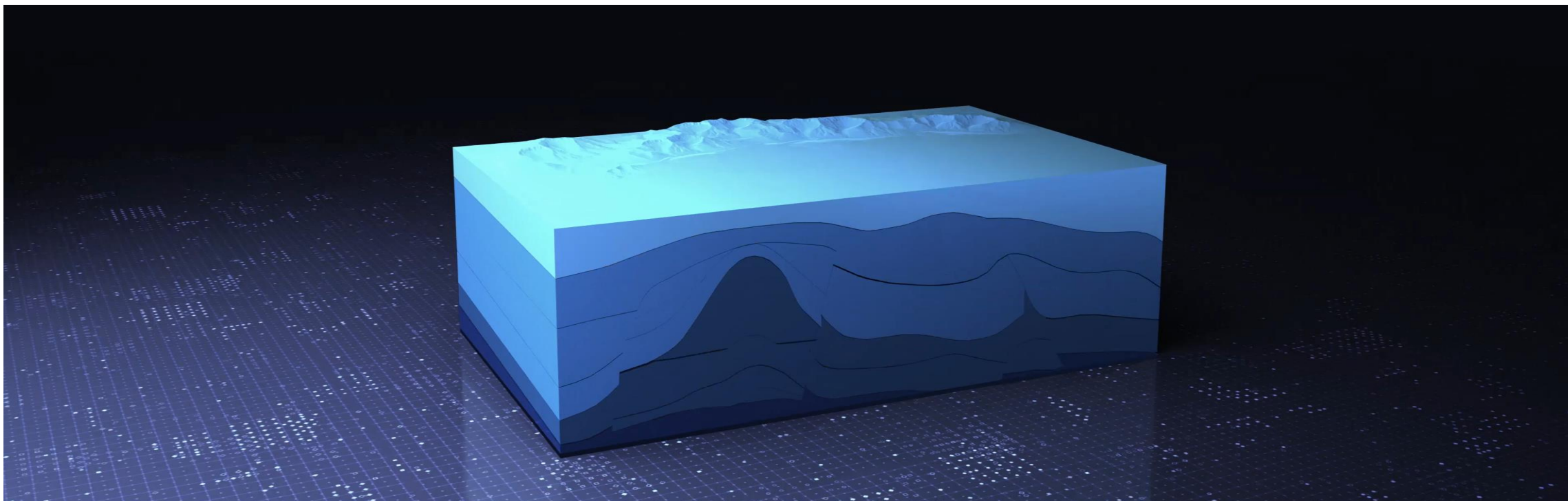
录井



测井



试油



油气层识别场景

地质勘探



物探



钻井



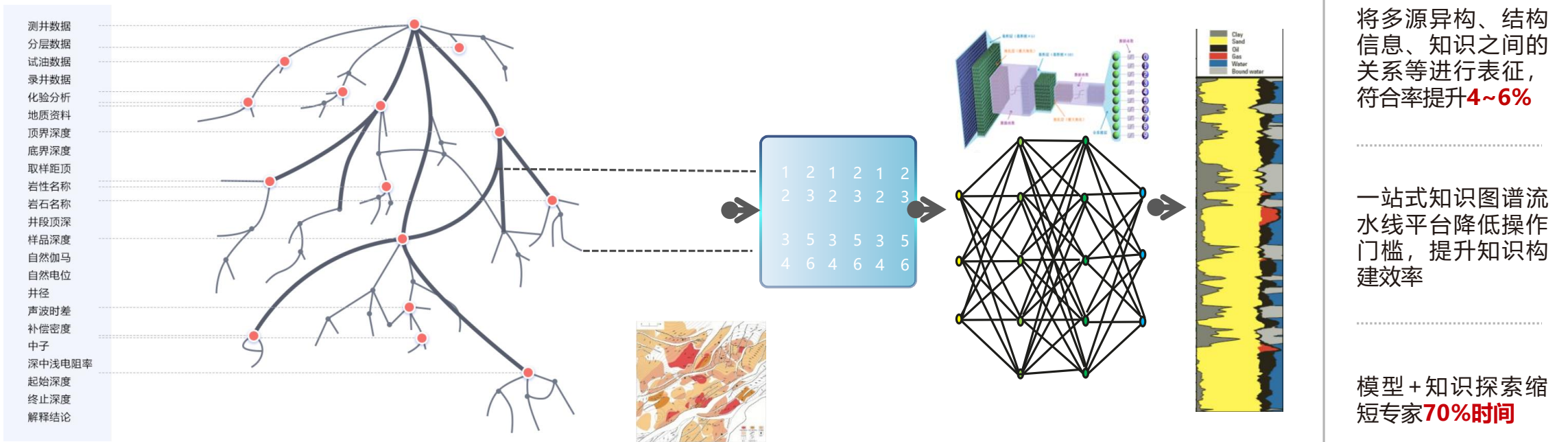
录井



测井



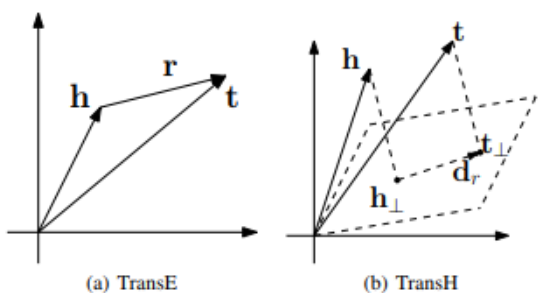
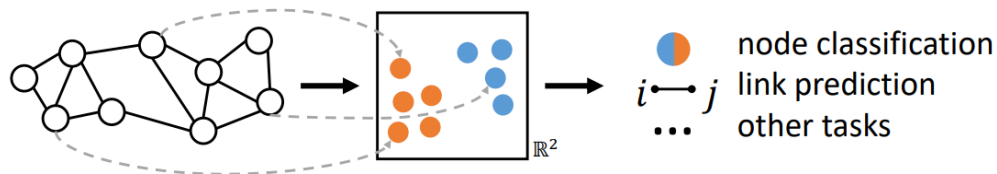
试油



基于多模态知识图谱和图嵌入的 知识表征

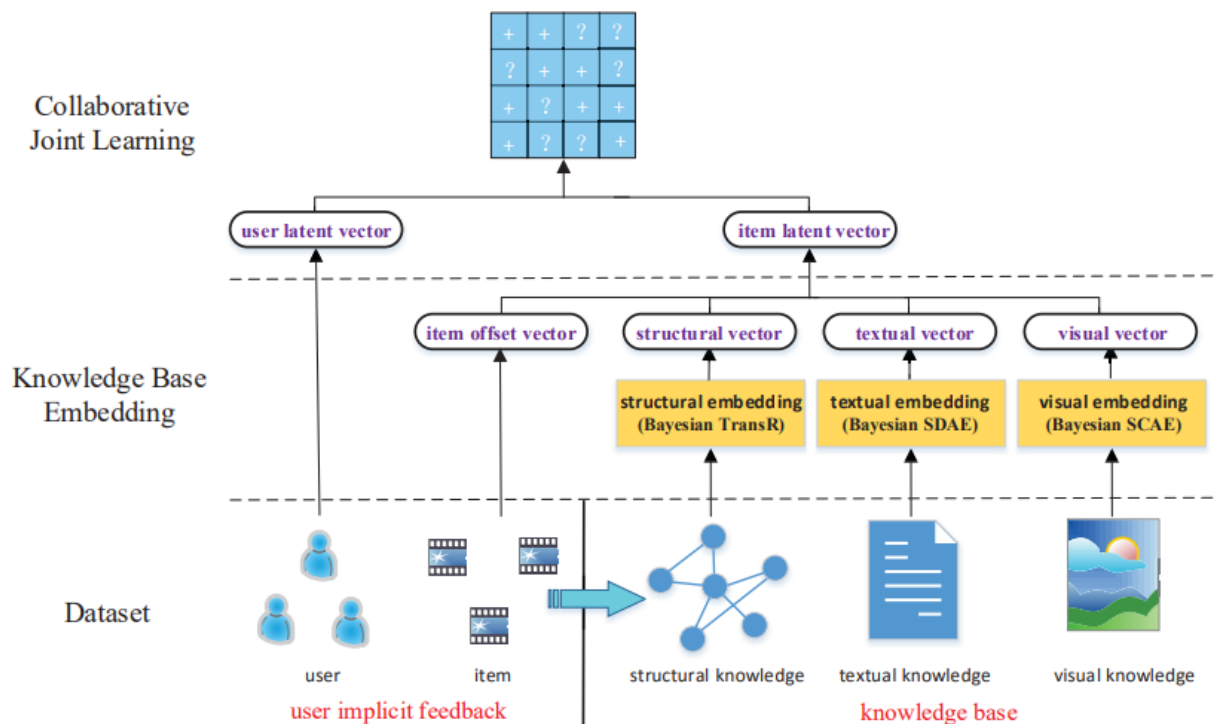
基于CNN的储层评价模型

知识表征与多模态



$$(a) \|h + r - t\| = 0$$

$$(b) \|h_{\perp} + d_r - t_{\perp}\| = 0$$



Collaborative knowledge base embedding for recommender systems, In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, pp. 353-362. ACM, 2016.

地质知识表征与知识蒸馏

中国科学: 信息科学

SCIENTIA SINICA Informationis

论文

基于地质知识蒸馏学习的油气储集层识别方法

李徵¹, 刘洪^{1*}, 王喆锋², 郑毅², 林霞³, 怀宝兴², 米兰³, 陈恩红¹

1. 中国科学技术大学大数据分析与应用安徽省重点实验室, 合肥 230027

2. 华为技术有限公司, 杭州 310051

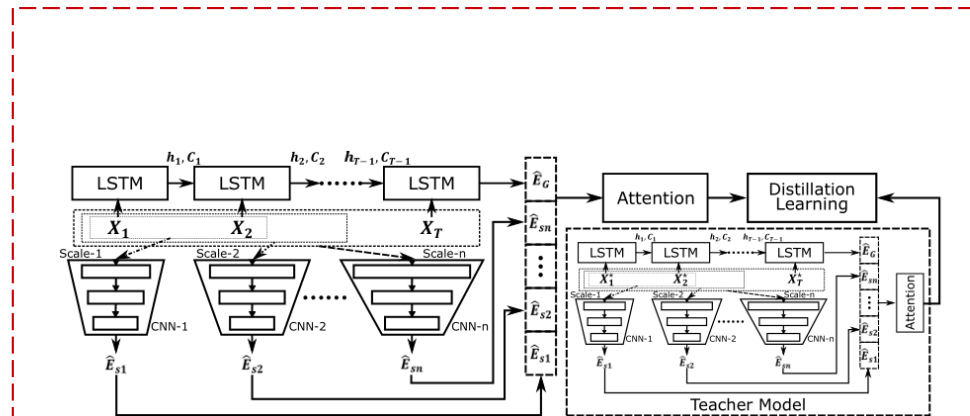
3. 中国石油勘探开发研究院, 北京 100083

* 通信作者. E-mail: qiliuql@ustc.edu.cn

国家自然科学基金 (批准号: 61922073, 61672483, U1605251) 资助项目

摘要 油气储集层识别是石油能源企业在勘测和开发业务中核心的任务之一。长期以来, 油气行业一直依靠专家人工分析海量测井数据以对地下油气储集层进行定性分析, 虽然专家解释结论有着很高的精度, 但是时间与经济成本都十分高昂。近些年来, 随着以深度学习为代表的人工智能技术的迅速发展, 智能油气储集层识别技术成为学术界和工业界共同关注的问题。然而, 真实工业环境存在严重的传感数据不一致问题, 给传统的监督学习模型带来巨大的挑战。本文针对传感器不一致情境中油气储集层识别任务展开研究, 提出多尺度地质知识蒸馏网络的方法。首先, 该方法提出一种多尺度特征自注意力融合机制来学习地质信息的多尺度动态表征。其次, 该方法设计一种地质知识蒸馏学习模型, 从非一致传感数据中学习额外的地质知识, 进一步提升模型准确度。最后, 在真实数据集上进行大量实验, 结果充分证明本文提出的模型在油气储集层识别任务上的有效性和鲁棒性。

关键词 油气储集层识别, 地质知识, 蒸馏学习, 传感数据, 深度神经网络



多尺度地质知识表征与知识蒸馏



CCDM 2020 最佳学生论文

首篇系统阐述E&P知识图谱的学术论文

Zhou XG, Gong RB, Shi FG *et al.* PetroKG: Construction and application of knowledge graph in upstream area of PetroChina. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 35(2): 368–378 Mar. 2020. DOI 10.1007/s11390-020-9966-7

PetroKG: Construction and Application of Knowledge Graph in Upstream Area of PetroChina

Xiang-Guang Zhou¹, Ren-Bin Gong¹, Fu-Geng Shi¹, and Zhe-Feng Wang²

¹PetroChina Research Institute of Petroleum Exploration and Development, Beijing 100083, China

²Huawei Technologies, Hangzhou 310007, China

E-mail: {zhouxg69, gongrb, sfg}@petrochina.com.cn; wangzhefeng@huawei.com

Received August 20, 2019; revised December 31, 2019.

Abstract There is a large amount of heterogeneous data distributed in various sources in the upstream of PetroChina. These data can be valuable assets if we can fully use them. Meanwhile, the knowledge graph, as a new emerging technique, provides a way to integrate multi-source heterogeneous data. In this paper, we present one application of the knowledge graph in the upstream of PetroChina. Specifically, we first construct a knowledge graph from both structured and unstructured data with multiple NLP (natural language processing) methods. Then, we introduce two typical knowledge graph powered applications and show the benefit that the knowledge graph brings to these applications: compared with the traditional machine learning approach, the well log interpretation method powered by knowledge graph shows more than 7.69% improvement of accuracy.

Keywords knowledge graph, natural language processing, oil and gas industry

JCSIT Journal of
Computer Science & Technology

Indexed in:

SCIE, Ei, INSPEC, JST, AJ, MR, CA,
DBLP, etc.



知识计算带来的改变



知识聚合

沉淀专业知识，培养油气人才

Precipitates professional knowledge
Helps to cultivate oil and gas talents



降本增效

简化业务流程，缩短工作时间

Simpler business process
Shorter working time

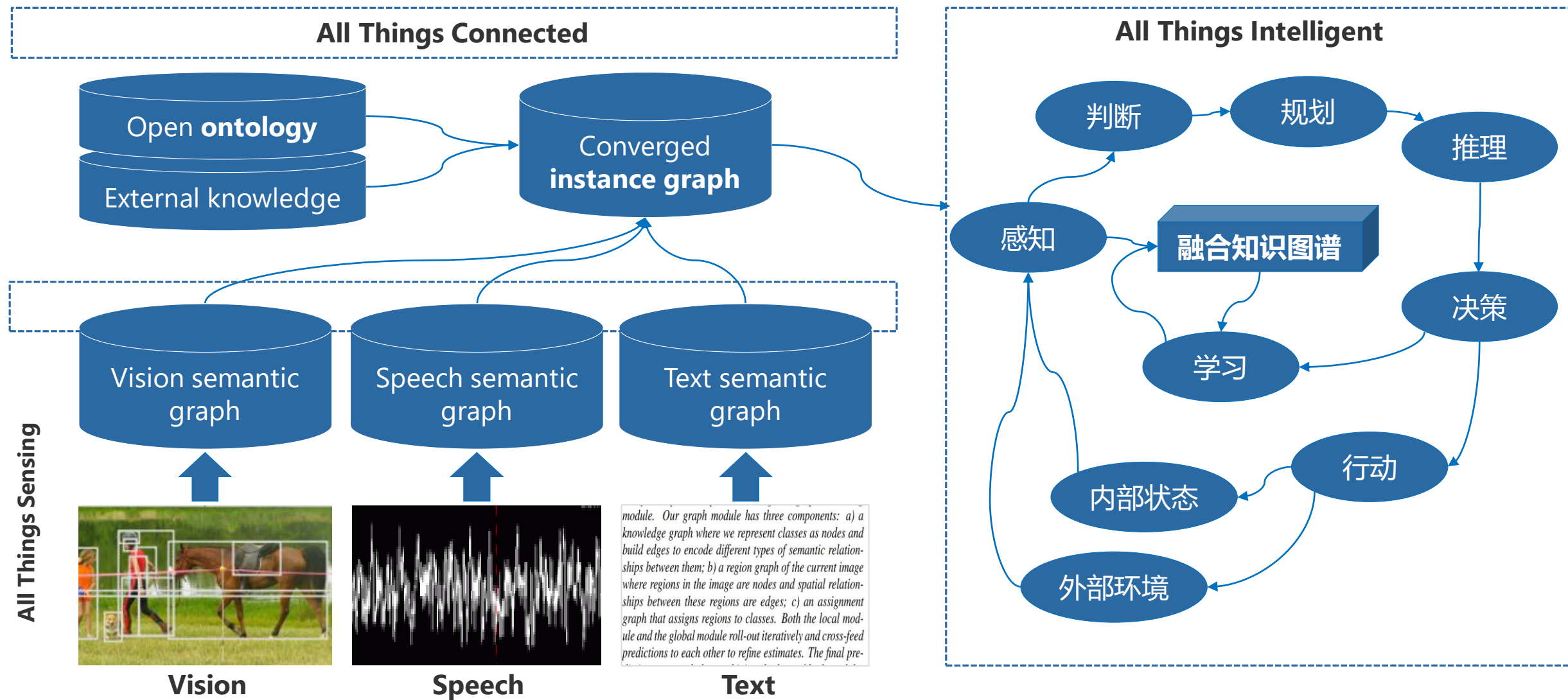


增储上产

增加探明储量，保障能源安全

Increased proven reserves
Ensured energy security

多模态知识图谱——ALL IN ONE



欢迎扫码关注



华为云知识图谱



华为云AI公众号