

一图胜万言，一目了然

中国知识图谱行业研究报告

©2022.8 iResearch Inc.



前言

研究背景：

2020年，艾瑞已经发布了《2020年中国知识图谱行业研究报告》、《2020年面向人工智能“新基建”的知识图谱行业白皮书》，初步对知识图谱的概念定义、行业场景、整体市场规模与产业链等进行了梳理与分析。

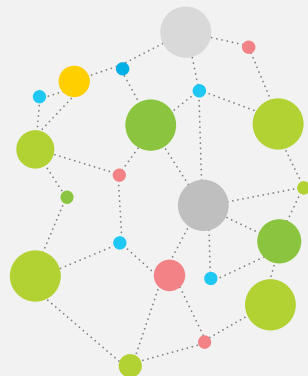
2022年，艾瑞将发布第三篇知识图谱行业报告，深入探讨知识图谱技术在通用知识图谱与行业知识图谱的企业诉求，剖析各行业的核心业务痛点，阐明知识图谱技术与应用对业务痛点的针对性与价值，对比各行业的知识图谱发展现状与潜力，展现知识图谱技术的建设重点与未来趋势。希望通过本报告，为读者呈现现阶段知识图谱行业发展的要点，提供辨析和判断知识图谱行业趋势的方法启发。当然，限于研究周期与行业理解水平，报告仍有不足之处，敬请各界读者指正。

研究对象：

本报告研究对象分为：①通用知识图谱与行业知识图谱，包含互联网、金融、政务与公安、医疗、工业与电力5大行业的行业痛点、知识图谱应用场景及价值、市场规模；②知识图谱现阶段的行业热点与建设痛点；③知识图谱的行业参与者类型及业务特点。

研究方法：

本报告通过业内资深的专家访谈、桌面研究、案例实证研究、行业对比研究、投融资数据统计与行业规模数据推算输出相应研究成果。



报告撰写

艾瑞咨询
产业数字化研究部
人工智能研究组



行业背景

感知到认知的跨越式发展，须引入发展认知技术，**知识图谱**在此形势下成为了破局的关键技术。**数字经济的持续发展**将加速知识图谱产业化进度，推动知识图谱与传统产业融合。**高性能图计算**可为图计算输送更快更准的计算能力，服务于知识图谱运算，**算力规模化部署**也为知识图谱计算的高密度、高功耗要求提供了有利发展条件。深度学习技术、NLP技术、知识图谱技术**协同并进**，尤其是**NLP技术**近几年的快速发展为知识图谱产业化提供了机会。



行业规模

2021年，知识图谱核心市场规模预计达到**107亿元**，而到2026年，相应规模将超过**296亿元**，**2021-2026年CAGR=22.5%**。**金融与公安**两大行业的知识图谱占比较高且增长速度较快，其业务与知识图谱可密切结合，同时具备建设意愿与资金投入，因而成为了市场规模的主要拉力。未来，随着政务数字化建设的完善，**政务**对知识图谱的业务需求会逐渐唤醒，成为未来市场的拉力之一。



建设重点探讨

知识图谱建设需面临的建设难点主要在于**数据治理、行业专家储备、底层图数据库存储、算法生产流程与性能待提升、客户认知待培养以及产品封装形式待优化**。攻克知识图谱的建设重难点将有利于从源头保证知识与智慧真实可靠、可用正确，储备培养深厚行业专家与技术复合型专家，升级底层图数据的存储方式，改善算法性能，为知识图谱建设减少阻碍。



趋势展望

未来，知识图谱厂商、大数据厂商、NLP厂商、互联网大厂与信息化厂商等知识图谱业内参与者将从**强化技术实力与深化行业认知**的角度出发，结合自身原有业务优势，持续深化发展行业知识图谱业务。知识图谱业务场景也将不断迭代，**行业应用场景边界拓宽，垂直应用场景被做深做透**。知识图谱生态也将继续由监管引导方、供给方、需求方、投资方、高校及科研院所融合共建，**汇聚建设合力**，促进产业生态成长壮大。

研究范围界定及赛道评估

1

行业场景分析

2

优秀案例实践

3

建设重点解读

4

行业趋势洞察

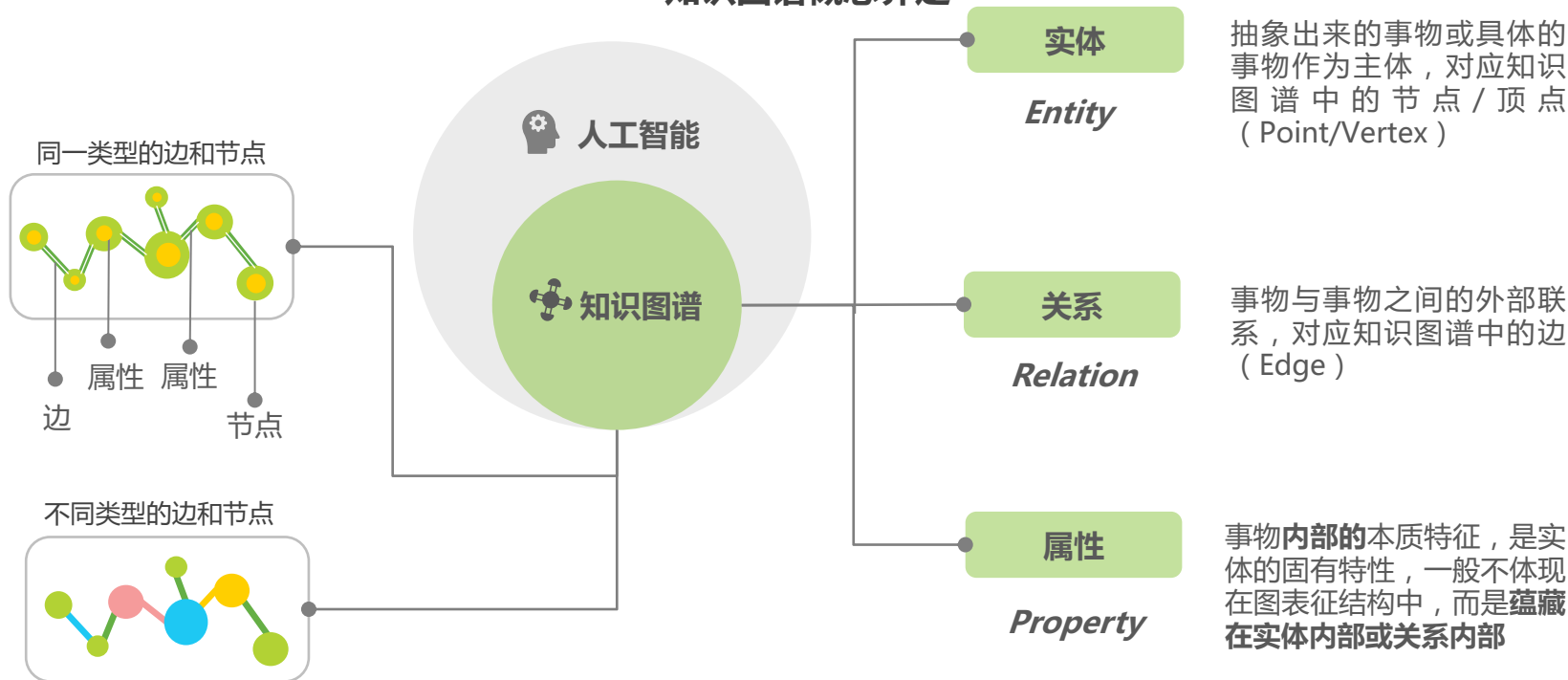
5

概念与研究范围界定

描绘实体之间关系的数据结构语义网络

知识图谱是人工智能的一大底层技术，是描绘实体之间关系的语义网络，自带语义、逻辑含义和规则，通过三元组即“实体×关系×属性”集合的形式来描述事物之间的关系。知识图谱将非线性世界中的知识信息结构化、可视化，辅助人类进行推理、预判、归类。知识图谱中的图并非图像概念，而是类似化学分子式的结构，一个知识图谱往往存在多种类型的实体与关系。知识结构网络化、网络结构复杂、网络由三元组构成、数据主要由知识库承载是知识图谱的四大基本特征。一般而言，知识图谱的数据以文本化数据为主，数据化数据为辅。

知识图谱概念界定



来源：艾瑞咨询研究院根据公开资料自主研究绘制。

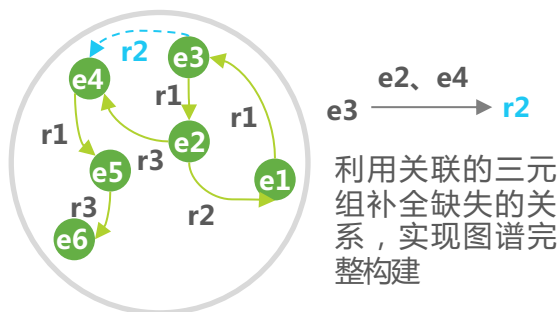
主要任务介绍

五大任务，复杂关联推理为突破重点

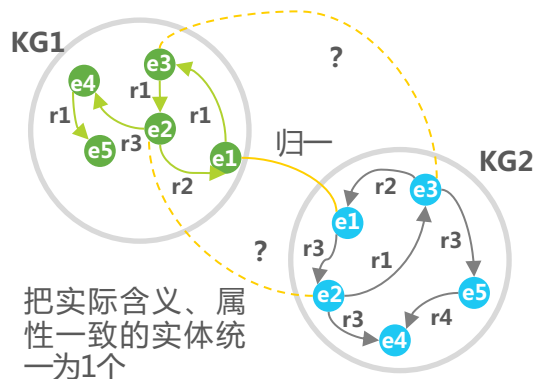
一般而言，知识图谱需执行的主要任务包括知识图谱构建与补全、实体统一（消歧）、实体分类、知识检索问答（简单推理）、复杂关系推理。现阶段的复杂关系推理需要更多依赖人类预测与推断各种可能的情况，并优先推荐可能性大的情况。

知识图谱主要任务

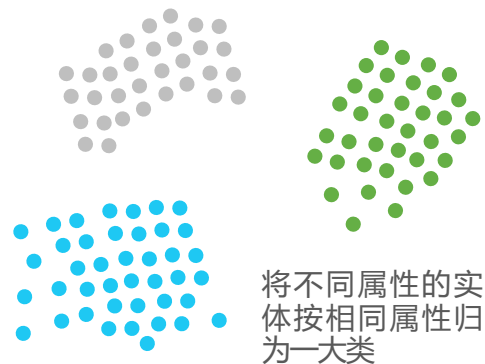
Task1 : 知识图谱构建与补全



Task2 : 实体统一（消歧）



Task3 : 实体分类

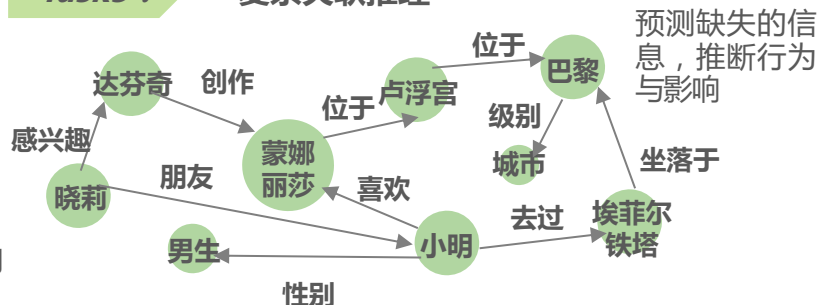


Task4 : 知识检索问答（简单推理）

- 水和二氧化碳反应可以生成什么？
- 《复仇者联盟》黑寡妇的扮演者是谁？
- 并联电路各支路电压的特征？

- 碳酸
- 斯嘉丽 约翰逊
- 各支路电压相同

Task5 : 复杂关联推理

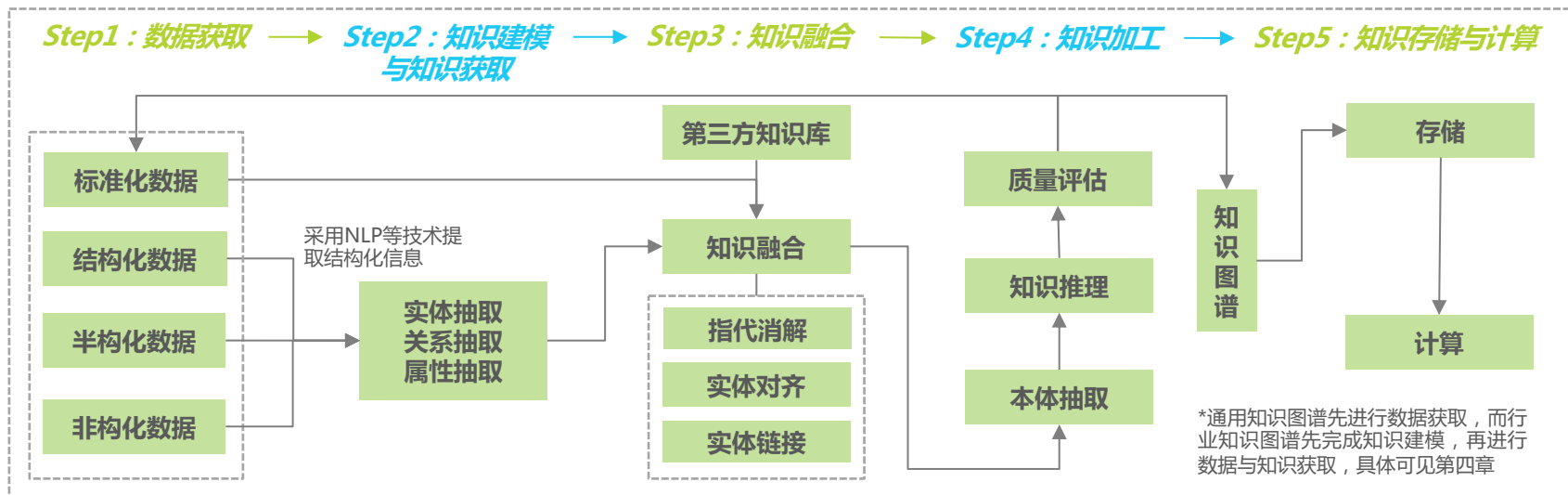


技术架构原理

技术架构分五步走，旨在构建实体语义网络

- **数据获取**：主要获取半结构化数据，为后续的实体与实体属性构建做准备。结构化数据则为数值属性做准备。
- **知识获取**：①从文本数据集中自动识别出命名实体，包括抽取人名、地名、机构名等；②从语料中抽取实体之间的关系，形成关系网络；③从不同的信息源中采集特定的属性信息。
- **知识融合**：①完成指示代词与先行词的合并；②完成同一实体的歧义消除；③将已识别的实体对象，无歧义地指向知识库中的目标实体。
- **知识加工**：①构建知识概念模块，抽取本体；②进行知识图谱推理，并对知识图谱的可信度进行量化评估，评估过关的知识图谱流入知识图谱库中存储，评估不过关的知识图谱返回一开始的数据环节进行调整，而后重复相同环节直到评估过关。
- **知识存储与计算**：存储是为了快速查询与运用知识，需支持底层数据描述与上层计算，有的主体计算包含在存储中。

知识图谱技术架构图



来源：艾瑞研究院根据公开资料、专家访谈自主研究绘制。

主流产品类型

成熟产品：通用互联网知识图谱；起步产品：行业知识图谱

知识图谱的产品类型以通用知识图谱与行业知识图谱为典型代表。通用知识图谱经过开拓性构建阶段后，逐渐演变为通用互联网知识图谱，形成搜索引擎、智能推荐、智能问答三大产品类型，产品发展较为成熟。行业知识图谱处于起步期，但其价值及效果逐渐被客户所认可，是知识图谱当前乃至未来一段时期内的发展热点。

通用知识图谱与行业知识图谱

强调知识广度：通用知识图谱



有哪些产品类型？



搜索引擎、智能推荐、智能问答



覆盖哪些知识？



通用/垂类泛知识、百科常识、泛学科领域知识

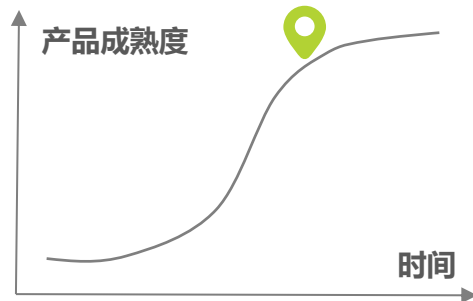


用于哪些场景？



互联网的信息检索、商品/服务/视频/交友推荐、KBQA、KGQA等

产品处于成熟期



强调知识深度：行业知识图谱

预览已结束，完整报告链接和二维码如下：

https://www.yunbaogao.cn/report/index/report?reportId=1_45105

