

第十一章 知识图谱

- ◆ 知识图谱定义及基本概念
- ◆ 知识图谱发展历程
- ◆ 知识图谱体系架构
- ◆ 知识图谱应用实例
- ◆ 开放知识图谱
- ◆ 未来发展方向



知识图谱定义及基本概念

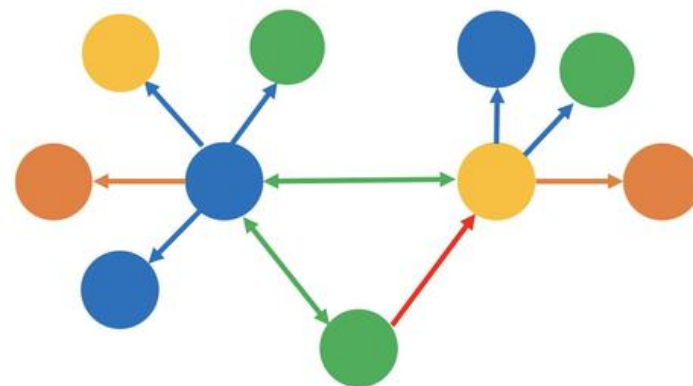
知识图谱 (Knowledge Graph)：以结构化的方式描述客观世界中实体、概念、事件以及之间的关系。

实体：客观世界的具体事务

概念：人类对客观事物的概念化描述表示

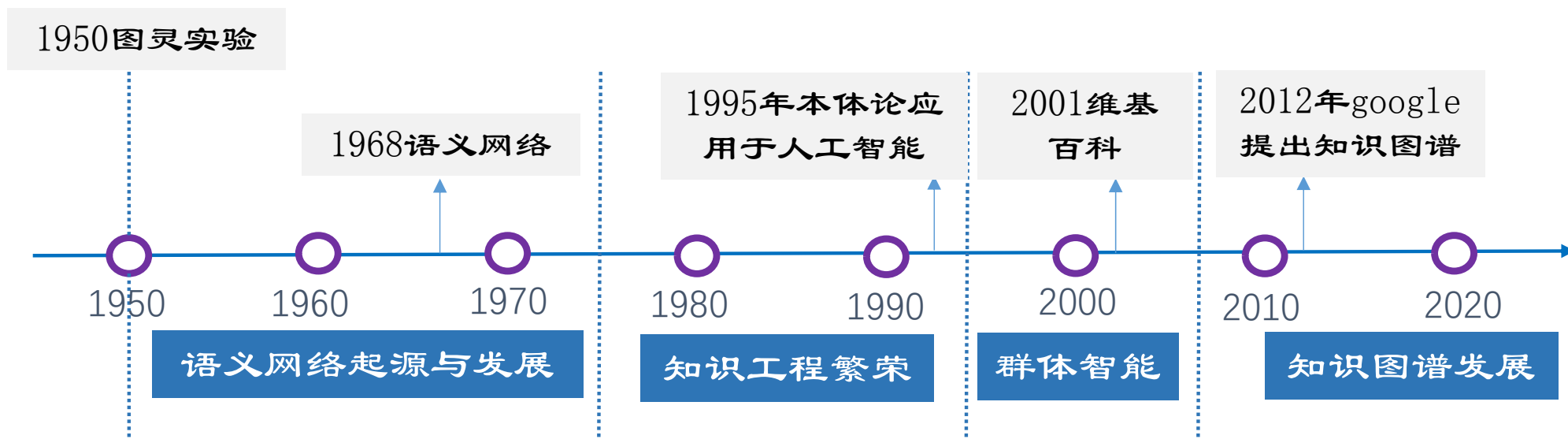
事件：发生在客观世界的活动

关系：实体、概念、事件之间客观存在的关联。



知识图谱的发展历程

2012年谷歌正式提出了知识图谱 (Knowledge Graph) 的概念，并应用于智能搜索领域。目前，知识图谱技术已在智能搜索、智能问答、网络分析、决策辅助、推荐系统等领域广泛应用。



知识图谱的体系架构



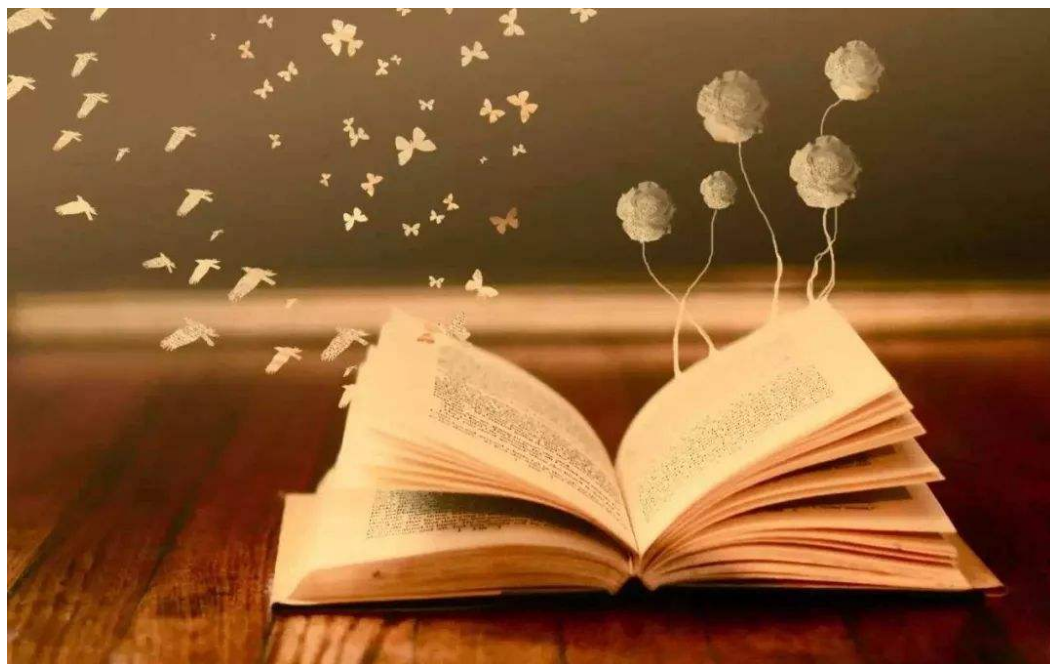
知识图谱的体系架构

知识图谱构建的关键技术

➤ 知识的获取

➤ 知识融合

➤ 知识计算



知识图谱构建的关键技术

——知识的获取

构建知识图谱的主要目的是获取大量的、让计算机可读的知识。知识大量存在于非结构化的文本数据、大量半结构化的表格和网页以及生产系统的结构化数据中。如何从非结构化、半结构化、以及结构化数据中获取知识是构建知识图谱的关键技术之一。

实体识别技术

实体关系识别技术

非结构化数据

实体链接

命名实体识别

语义相似度计算

同义词构建

语义解析

依存分析

特征提取

关键词提取

事件促发词识别

事件角色识别

潜在出题分析

.....

半结构化数据

包装器学习

结构化数据

ETL工具

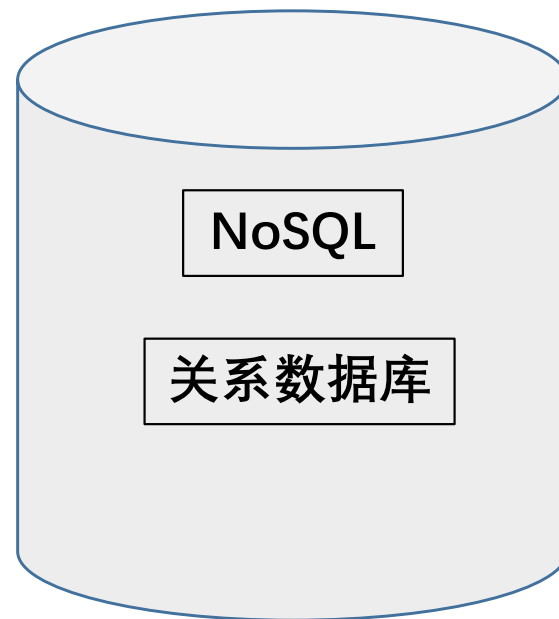
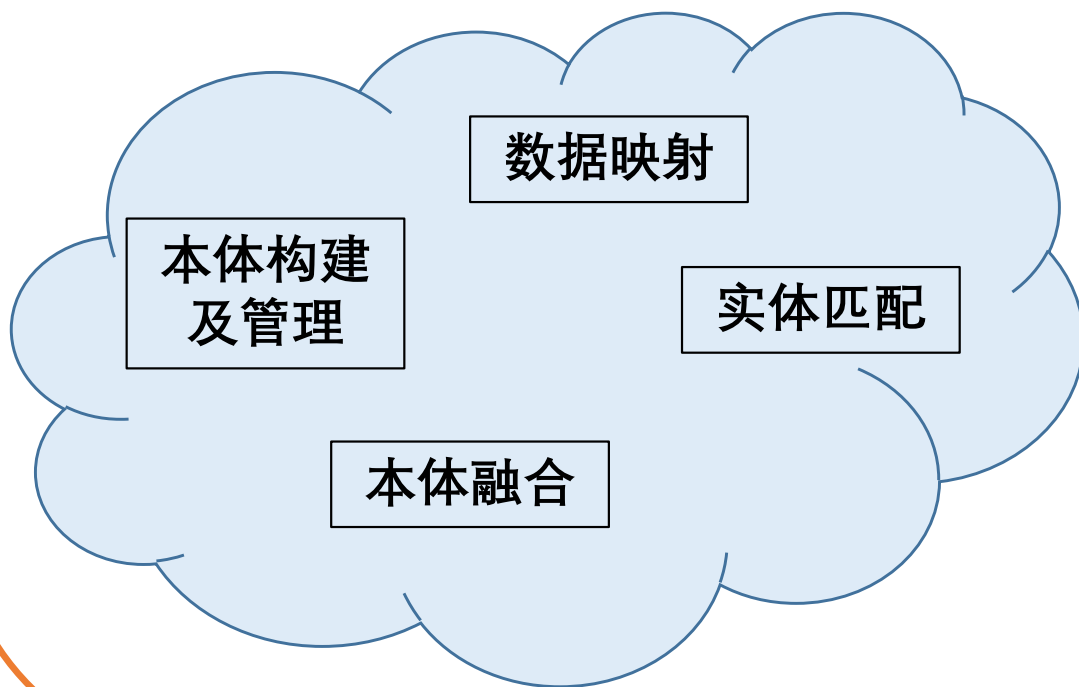
知识图谱构建的关键技术

——知识的融合

当知识从各个数据源下获取时需要提供统一的术语将各个数据源获取的知识融合成一个庞大的知识库。通过数据映射技术建立本体中术语和不同数据源抽取知识中词汇的映射关系，进而将不同数据源的数据融合在一起。

本体融合技术

知识库存储、管理



知识图谱构建的关键技术

——知识的计算

知识计算主要是根据图谱提供的信息得到更多隐含的知识，如通过本体或者规则推理技术可以获取数据中存在的隐含知识。

本体推理

规则推理

社区计算

相似子图
计算

路径计算

链路预测

不一致性
检测

专家系统

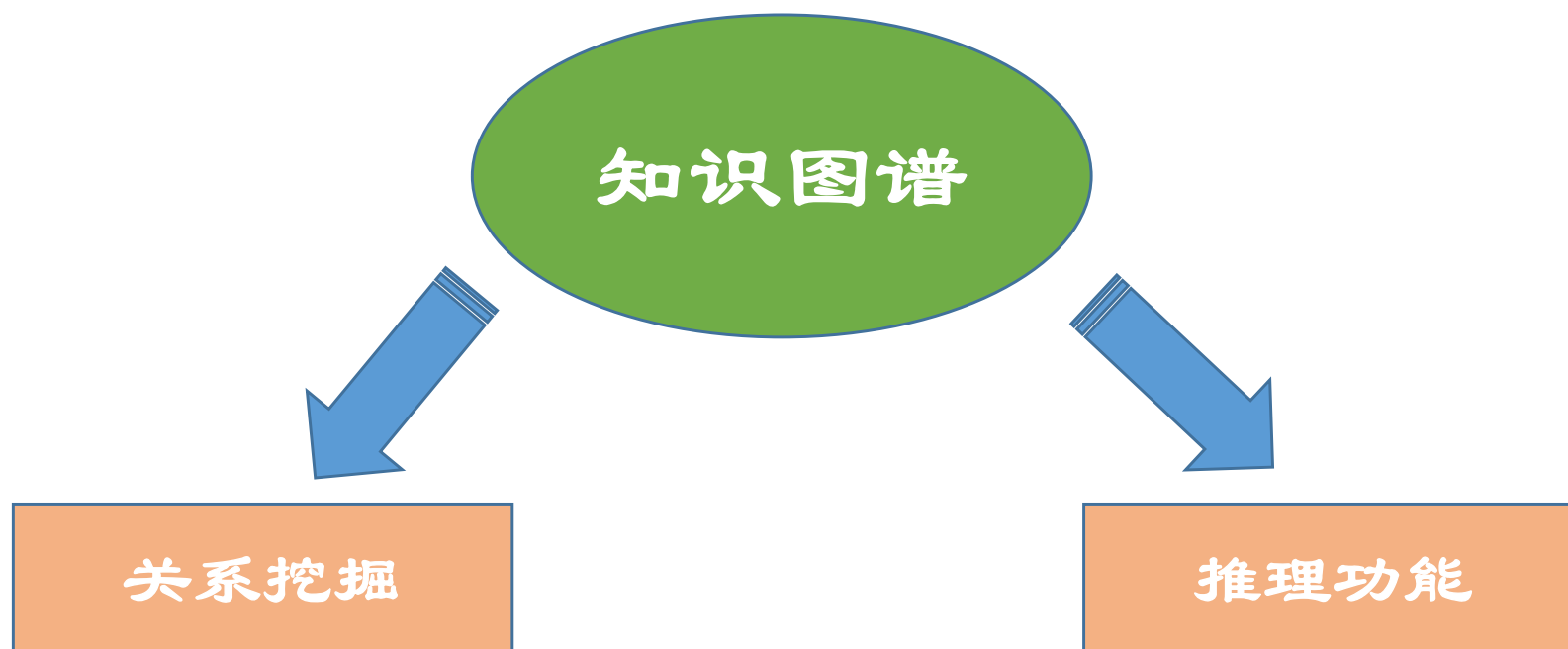
推荐系统

语义搜索

可视化分
析

问答

知识图谱的应用实例



知识图谱的应用

知识图谱应用场景地图

金融行业

银行

对公业务营销与风控

零售业务营销与风控

保险

营销与风控

证券

营销与风控

基金

营销与风控

金融

领域知识库

智能客服

舆情监控

数字化采购

政府与公共服务行业

交通

智能交通

应急管理部

应急管理

教务数字化

电子政务

公安

智能刑侦

智能经侦

智能反恐

治安管理

电信行业

运营商

领域知识库

医疗与药企行业

药企

医药翻译

数字化营销

医院

智能导诊

零售行业

品牌商

智慧门店

智慧营销

能源与工业行业

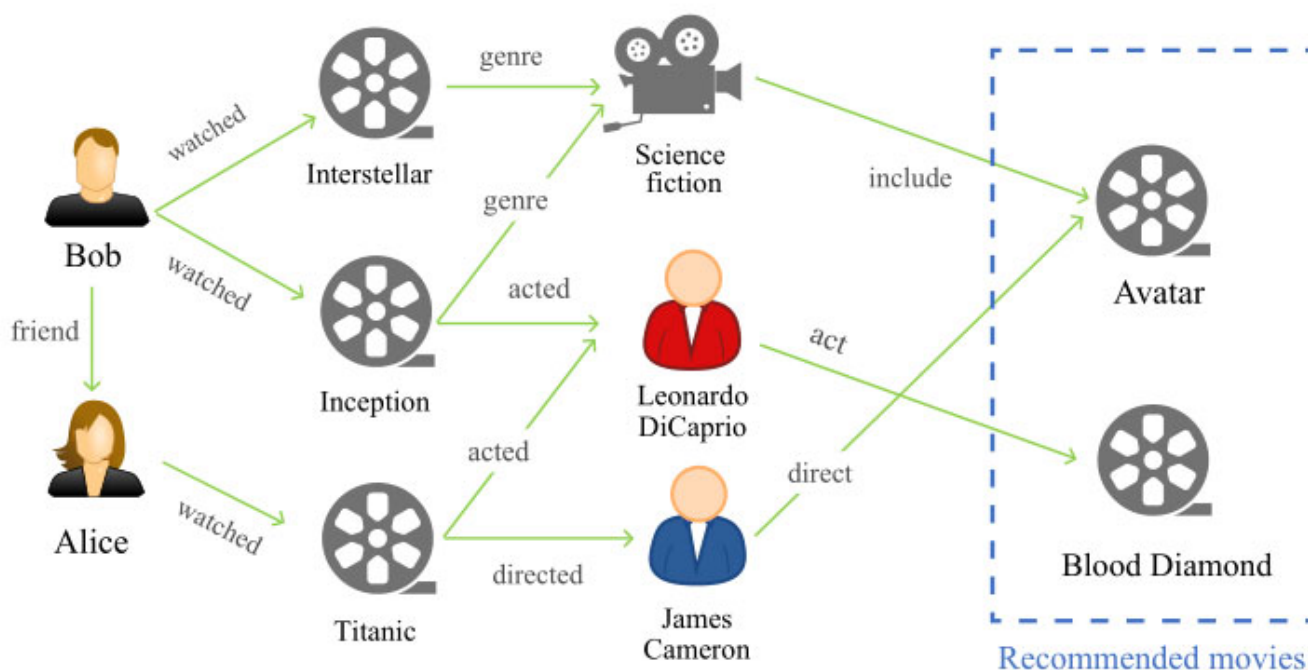
能源与工业

领域知识库

知识图谱的应用实例

◆ 知识图谱在推荐系统中的应用

基于知识图谱的推荐系统 (KG-based recommendation system, KGRS) 引起研究者的广泛兴趣，主要是把知识图谱作为辅助信息整合到推荐系统中，这样的做法带来两个方面的优势，其一是能够提高推荐系统的准确性，其二是能够为推荐系统提供可解释性。



知识图谱的应用实例

◆ 知识图谱在情报分析的案例

股票投研情报分析

通过知识图谱相关技术从招股书、年报、公司公告、券商研究报告、新闻等半结构化表格和非结构化文本数据中批量自动抽取公司的股东、子公司、供应商、客户、合作伙伴、竞争对手等信息，构建出公司的知识图谱。通过此图谱做更深层次的分析 and 更好的投资决策。

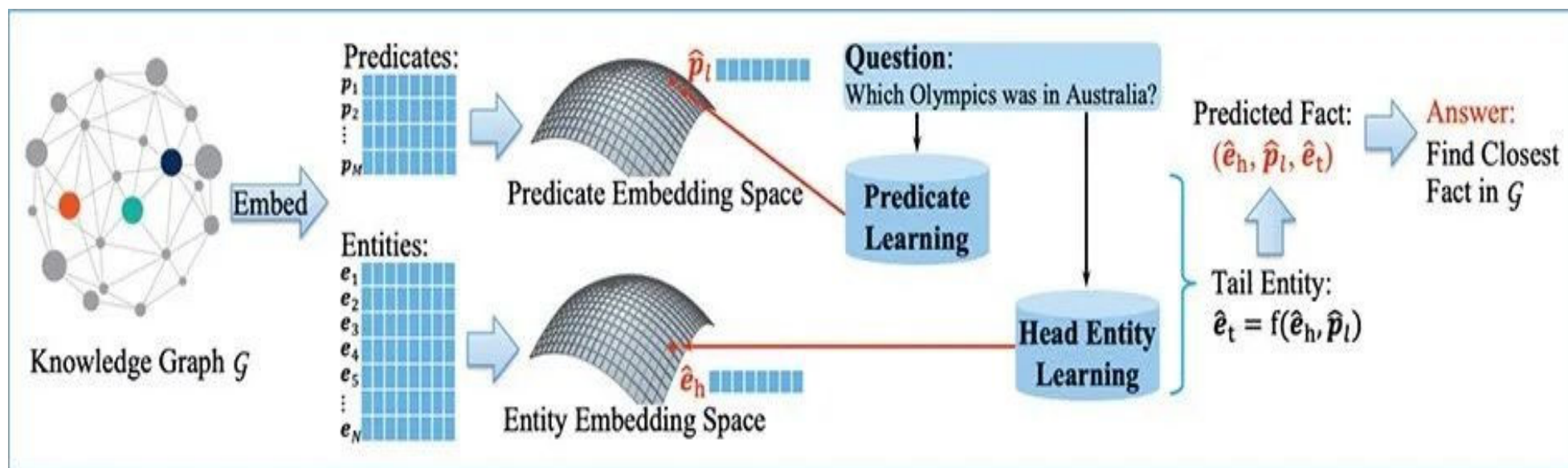
公安情报分析

通过融合企业和个人银行资金交易明细、出行、住宿、工商等信息构建初步的“资金账户-人-公司”关联知识图谱。同时从案件描述、笔录等非结构化文本中抽取人、事、物、组织、卡号、时间、地点等信息，链接并补充到原有的知识图谱中形成一个完整的证据链。辅助公安刑侦、经侦、银行进行案件线索侦查和挖掘同伙。

知识图谱的应用实例

◆ 知识图谱在问答系统中的应用

由于知识图谱本身是对机器友好的结构，如果有了相应的问询语句，即可很容易地在知识图谱中查询到最终的答案。因此，难度主要集中在如何将自然语言问题解析为知识图谱内存在的关系或者实体的合法查询。

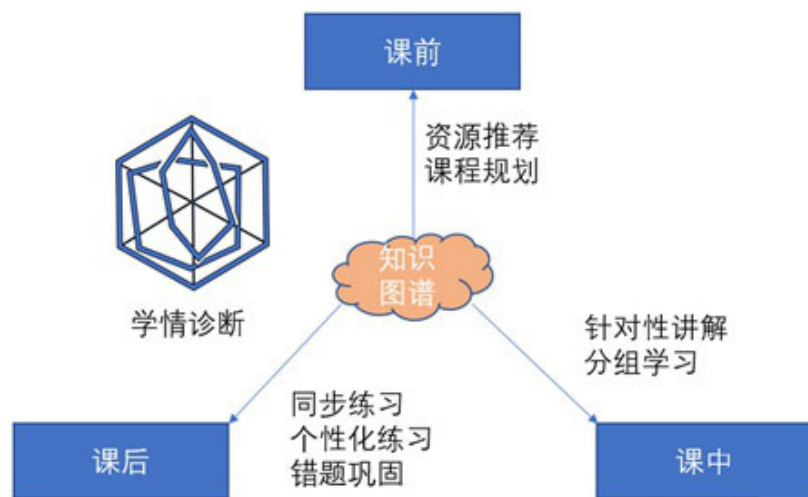


知识图谱的应用实例

◆ 知识图谱在智慧教育的应用

教育知识图谱，以学科知识为核心，建立各个学科的知识点概念建立层级关系，知识点与知识点之间的关联关系，不同知识点之间的前后序关系，构成学科知识图谱。——**帮助学生构建知识体系，加深知识理解**

学科知识图谱，可以跟教学资源（教材，试题，讲义，教学视频，试卷等）构建关联，进而通过用户信息和学习记录，建立知识点与用户之间的关联。——**学情研判、学习资源个性化推荐**



智能备课

智慧课堂

深度阅读

答疑机器人

.....

国内外开放知识图谱

- 目前国内外有很多高质量的开放知识图谱平台，例如 DBpedia、Yago、Wikidata、BabelNet、ConceptNet 和 Microsoft Concept Graph 等。

DBpedia是一个大规模的多语言百科知识图谱，可视为是维基百科的结构化版本。DBpedia使用固定的模式对维基百科中的实体信息进行抽取。DBpedia目前拥有 127 种语言的超过两千八百万个实体与数亿个 RDF 三元组，并且作为链接数据的核心，与许多其他数据集均存在实体映射关系。DBpedia支持数据集的完全下载。

国内外开放知识图谱

Yago是一个整合了维基百科与 WordNet的大规模本体。目前，Yago 拥有 10 种语言约 459 万个实体，2400 万个 Facts，Yago 中 Facts的正确率约为 95%。Yago支持数据集的完全下载。

Wikidata 是一个可以自由协作编辑的多语言百科知识库，它由维基媒体基金会发起，期望将维基百科、维基文库、维基导游等项目中结构化知识进行抽取、存储、关联。目前 Wikidata 目前支持超过 350 种语言，拥有近 2500 万个实体及超过 7000 万的声明，并且目前 Freebase 正在往 Wikidata 上进行迁移以进一步支持 Google 的语义搜索。Wikidata 支持数据集的完全下载。

国内外开放知识图谱

BabelNet 是目前世界范围内最大的多语言百科同义词典，它本身可被视为一个由概念、实体、关系构成的语义网络 (Semantic Network)。BabelNet 目前有超过 1400 万个词目，每个词目对应一个 synset。目前 BabelNet 又整合了 Wikidata、GeoNames、OmegaWiki 等多种资源，共拥有 271 个语言版本。关于数据集的使用，BabelNet 目前支持 HTTP API 调用，而数据集的完全下载需要经过非商用的认证后才能完成。

ConceptNet 是一个大规模的多语言常识知识库，其本质为一个以自然语言的方式描述人类常识的大型语义网络。自 1999 年开始通过文本抽取、众包、融合现有知识库中的常识知识以及设计一些游戏从而不断获取常识知识。ConceptNet 目前支持数据集的完全下载。

国内外开放知识图谱

Microsoft Concept Graph 是一个大规模的英文 Taxonomy，其中主要包含的是概念间以及实例（等同于上文中的实体）概念间的 IsA 关系，其中并不区分 instanceOf 与 subclassOf 关系。目前，Microsoft Concept Graph 拥有约 530 万个概念，1250 万个实例以及 8500 万个 IsA 关系（正确率约为 92.8%）。目前支持 HTTP API 调用，而数据集的完全下载需要经过非商用的认证后才能完成。

中文目前可用的大规模开放知识图谱有 Zhishi.me、Zhishi.schema 与 XLore 等。

- Zhishi.me 是第一份构建中文链接数据的平台，与 DBpedia 类似。支持数据集的完全下载
- XLore 是一个大型的中英文知识图谱。所有数据可以通过在线 SPARQL Endpoint 查询得到。

国内外开放知识图谱

中文开放知识图谱联盟 (OpenKG) 旨在推动中文知识图谱的开放与互联，推动知识图谱技术在中国的普及与应用，为中国人工智能的发展以及创新创业做出贡献。目前，联盟已经搭建有 OpenKG.CN 技术平台，已有 35 家机构入驻。吸引了国内最著名知识图谱资源的加入，如 Zhishi.me, CN-DBPedia, PKUBase。并已经包含了来自于常识、医疗、金融、城市、出行等 15 个类目的开放知识图谱。



未来发展方向 —— 认知图谱

认知图谱可以被解释为“基于原始文本数据，针对特定问题情境，使用强大的机器学习模型动态构建的，节点带有上下文语义信息的知识图谱”。

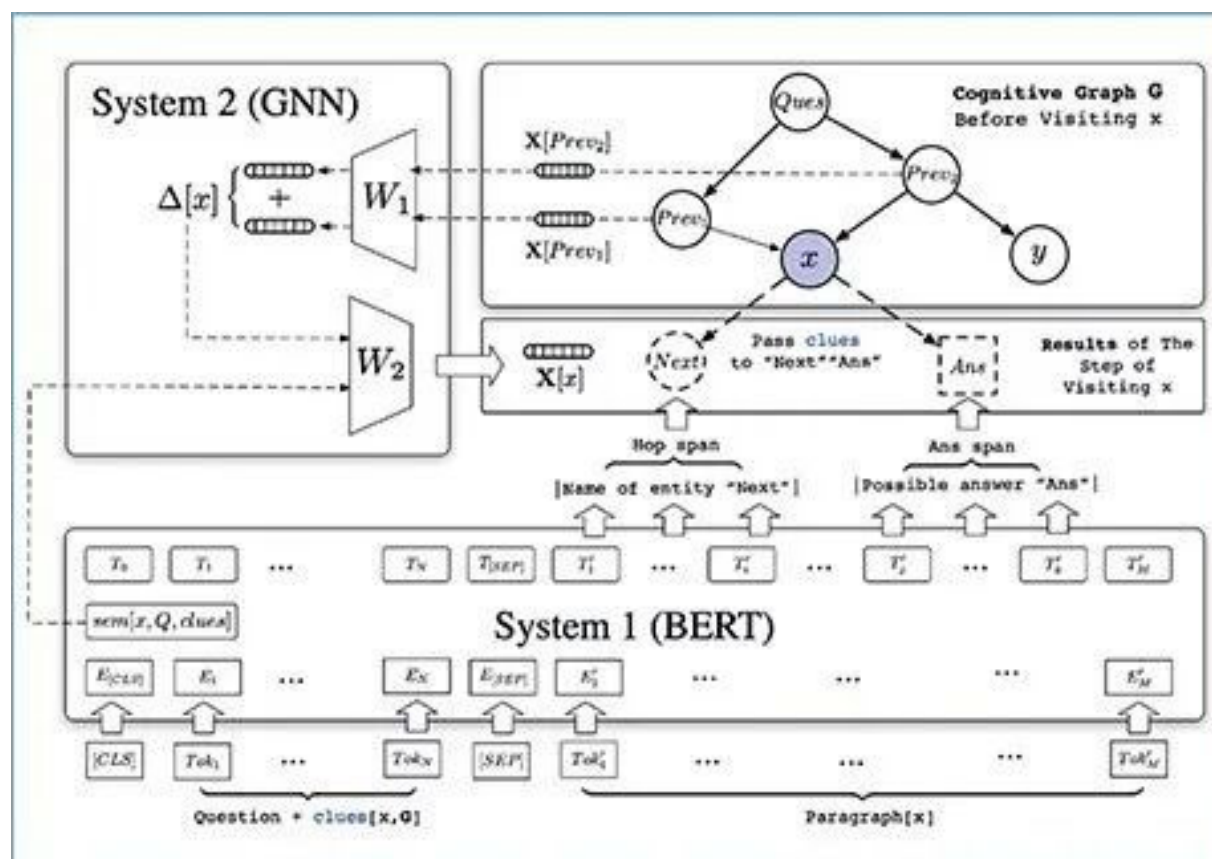
认知图谱主要有三方面创新，分别对应人类认知智能的三个方面：

1. **（长期记忆）直接存储带索引的文本数据，使用信息检索算法代替知识图谱的显式边来访问相关知识。**
2. **（系统1推理）图谱依据查询动态、多步构建，实体节点通过相关实体识别模型产生。**
3. **（系统2推理）图中节点产生的同时拥有上下文信息的隐表示，可通过图神经网络等模型进行可解释的关系推理。**

本质上，认知图谱的改进思路是减少图谱构建时的信息损失，将信息处理压力转移给检索和自然语言理解算法，同时保留图结构进行可解释关系推理。

未来发展方向 —— 认知图谱

认知图谱问答提出了一种新颖的迭代框架：算法使用两个系统来维护一张认知图谱，系统1在文本中抽取与问题相关的实体名称并扩展节点和汇总语义向量，系统2利用图神经网络在认知图谱上进行推理计算。



Thanks !