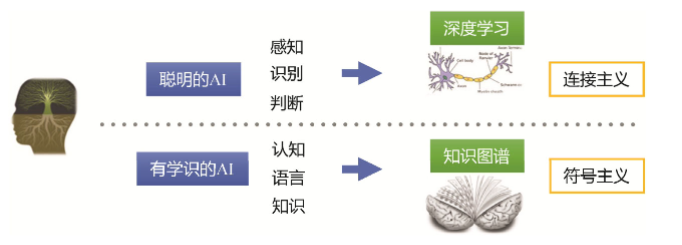
## 语言和知识

### 有学识的人工智能

故事从人工智能的起源开始。

人工智能分为多个流派，其中两个主要流派是：连接主义和符号主义。

* 连接主义：主张智能的实现应该模拟人脑的生理结构，即**用计算机模拟人脑的神经网络连接**。发展至今，就是**深度神经网络**
* 符号主义：主张智能的实现应该模拟人类的心智，即**用计算机符号记录人脑的记忆，表示人脑中的知识**，即所谓**知识工程与专家系统**[[1]](#footnote-22)等。



什么叫认知智能？认知智能有两个核心的研究命题，一个是**语言理解**，另外一个是**知识的表示与处理**。

两者相辅相成，人类通过认识世界来积累知识，而人类的知识大多数使用自然语言来描述和记录，因此正确理解语言需要知识的帮助。

### 知识的承载与表示方式

知识有很多种表达载体和存在形式，自然语言是人类知识最主要的表达载体，但是文本字符串对机器不太友好，机器难以理解人类的语言。

两大技术路线：

* 抽取技术：例如从文本中识别实体、关系和逻辑结构等。
* 语言预训练：即通过大量的文本语料训练一个神经网络大模型，文本中的知识被隐含在参数化的向量模型中，而向量化的表示和神经网络是对机器友好的。

对于机器而言，结构化数据更易于处理，图结构比字符串序列能够表达更加丰富的语义和知识。

深度学习或者更为准确地说是表示学习的兴起，表明参数化的向量和神经网络是更适合机器完成快速计算的信息载体。

### 知识图谱的模型

知识图谱的本质是一种结构化的知识表示形式。旨在利用图结构建模、识别和推断事物之间的复杂关联关系和沉淀领域知识，已经被广泛地应用于语义搜索、智能问答、语言理解、媒体理解、推理引擎和决策引擎等众多领域。

知识图谱本质上可以看作一种世界模型——World Model。

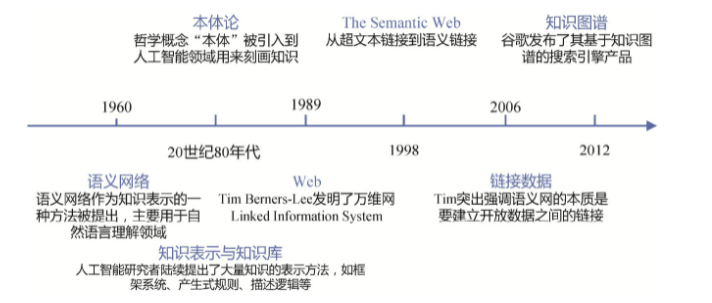
知识图谱同时拥抱机器的符号表示和向量表示，并能将两者有机地结合起来，解决搜索、问答、推理和分析等多方面的问题。

## 起源

从开始的World Wide Web，到终极目标——Web of Everythings。

知识图谱的早期理念：从经典Web发展出来的Semantic Web。

**Semantic Web也是以图和链接为中心的信息管理系统**，但不同之处是，图中的节点可以是粒度更细的事物，如一本书、个人、机构和概念等，图中的链接也标明这些事物之间的语义关系，如雇佣、朋友和作者等。

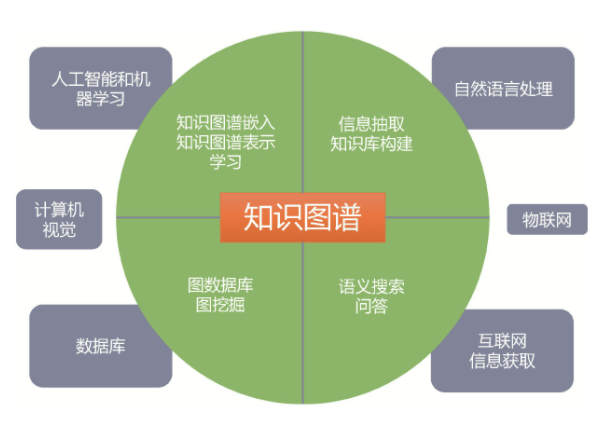


## 价值

* 语义搜索
* 智能问答
* 推荐系统
* 辅助语言语义理解
* 扩展视觉理解的深度和广度
* 辅助IoT设备互联
* 支持大数据分析

## 技术内涵

知识图谱涉及的主要技术要素：



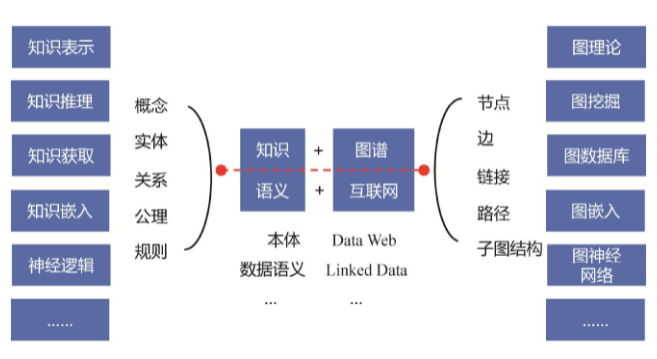
具体的交叉领域：………………..

### 两个核心的技术维度

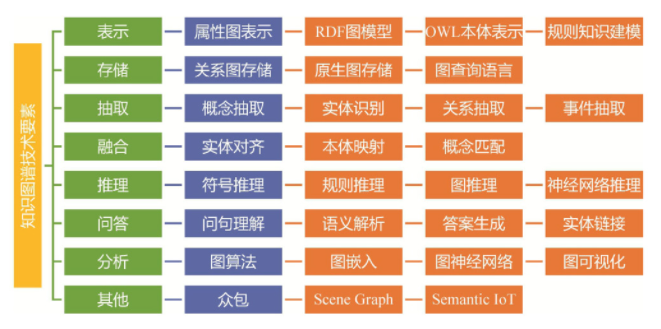
第一个是从**知识**的视角，它来源于传统AI的知识表示与推理领域，关心**怎么表示概念和实体，怎样刻画它们之间的关系，怎样进一步表示公理、规则等更加复杂的知识**。

第二个视角是从**图**的视角，它来源于知识图谱的互联网基因，关心**图中的节点、边、链接、路径、子图结构，怎样存储大规模的图数据，怎样利用图的结构对图数据进行推理、挖掘与分析等**

*知识和图谱：互补的两个维度*



### 技术栈



知识图谱的技术栈可以分为以下几个方面：表示、存储、抽取、融合、推理、问答、分析等几个方面。

之后分章节进行概览性的表示。

1. 建立一个系统，能够从专家大脑里获取知识，再通过一个推理引擎为非专家用户提供服务，如辅助诊断、判案等。从人脑获取知识的过程就叫作知识工程。 [↑](#footnote-ref-22)