



# 知识图谱技术与应用

---

彭鹏

湖南大学

[hnu16pp@hnu.edu.cn](mailto:hnu16pp@hnu.edu.cn)

# 课程介绍

教材：《知识图谱：概念与技术》，肖仰华等著，电子工业出版社

考核形式：

TuGraph课程作业（30%）；大作业（40%）；上台论文报告（30%）

课程网站：<https://bnu05pp.github.io/KGGraduateCourse/2025.html>

课程QQ群：



群名称:2025知识图谱技术与应用  
群 号:201839578



# 目 录

1. 什么是知识图谱?
2. 知识图谱的源起
3. 知识图谱技术的多个维度

# Part 1

## 什么是知识图谱?

# 知识图谱 (Knowledge Graph)

2012年5月16日，Google发布“知识图谱”的新一代“智能”搜索功能。



[全部](#) [图片](#) [地图](#) [视频](#) [新闻](#) [更多](#) [设置](#) [工具](#)

找到约 15,700,000 条结果 (用时 0.63 秒)

zh.wikipedia.org › zh-hans › 瓦特

[瓦特- 维基百科，自由的百科全书](#)

**瓦特**（符号：W）是国际单位制的功率单位。**瓦特**的定义是1焦耳/秒（1 J/s），即每秒钟转换，使用或耗散的（以安培为量度的）能量的速率。日常生活中更常用千瓦作为 ...

[量级](#) · [千瓦](#)

baike.baidu.com › item › 瓦特

[瓦特（功率单位）\\_百度百科](#)

**瓦特**是国际单位制的功率单位。**瓦特**的定义是1焦耳/秒（1J/s），即每秒转换、使用或耗散的（以焦耳为量度的）能量的速率。在电学单位制中，是伏特乘安培乘功率 ...

baike.baidu.com › item › 瓦特

[詹姆斯·瓦特\\_百度百科](#)

**詹姆斯·瓦特**（James Watt，1736年1月19日— 1819年8月25日），英国发明家，第一次工业革命的重要人物。1776年制造出第一台有实用价值的蒸汽机。以后又经过一 ...

代表作品: 改良蒸汽机 外文名: James Watt

[家世背景](#) · [改良蒸汽机](#) · [个人生活](#) · [人物争议](#)

## 瓦特

功率单位



瓦特是国际单位制的功率单位。瓦特的定义是1焦耳/秒，即每秒钟转换，使用或耗散的能量的速率。日常生活中更常用千瓦作为单位，1千瓦=1000瓦特，千瓦又可合写作“瓩”。在电学单位制中，瓦特 = 伏特·安培，但不是伏安。瓦特只用来表示交流电的实功率，伏安用来表示视在功率，两者的比为功率因数。单位转换： [电磁学：维基百科](#)

[反馈](#)

查看以下内容的结果：

[詹姆斯·瓦特 / 发明家](#)

# 传统互联网搜索技术

## 基于关键词字符串匹配

1. 爬取网页建立倒排索引
2. 基于倒排索引，搜索引擎首先找到了包含关键词的网页
3. 根据打分策略（如Pagerank，HITS等），对网页排序
4. 返回用户



计算机并不理解背后的语义

# 知识图谱的目的

构建知识图谱的目的，就是让机器具备认知能力，理解这个世界。

Baidu 百度 瓦特是哪个学校的校友? 百度一下

网页 资讯 视频 图片 知道 文库 贴吧 采购 地图 更多»

百度为您找到相关结果约359,000个 搜索工具

[瓦特是哪个大学毕业的\\_百度知道](#)  
2个回答 - 回答时间: 2016年12月1日  
最佳答案: 瓦特没受过正规的大学教育。 瓦特的求学经历: 1757年,格拉斯哥大学的教授提供给瓦特一个机会,让他在大学里开设了一间小修理店,这帮助瓦特...  
[更多关于瓦特是哪个学校的校友?的问题>>](#)  
百度知道 - 百度快照

[瓦特是以什么身份迈进大学的校门的?\\_百度知道](#)  
1个回答 - 回答时间: 2020年1月2日  
最佳答案: 像瓦特这样的人才正是当时的格拉斯哥大学所紧缺的,再加上迪克博士和学校教授们的帮助,学校当局终于同意在校园里给瓦特一个工作间,并且授予了他一个"大学"...  
[更多关于瓦特是哪个学校的校友?的问题>>](#)  
百度知道 - 百度快照

[瓦特?据说这是最难毕业的学校!\\_博洛尼亚](#)  
2018年12月19日 - 这座城市也因其欧洲著名学术中心的地位闻名于世——天文望远镜的发明者伽利略和以花心浪漫出名的卡萨诺瓦都是这所学校的校友。帕多瓦大学建立于1222年,当年...  
搜狐网 - 百度快照

[厉害了,进了这个英国名校,你就是亚当斯密的校友了!](#)  
2018年3月5日 - 1870年以后,格拉斯哥大学从市中心搬到了西区(West End),学校规模不断扩大,逐渐...瓦特成名后,格拉斯哥大学授予他法学名誉博士学位,属于“编外”校友。 图片9:格大...  
搜狐网 - 百度快照

瓦特是哪个学校的校友?

[格拉斯哥大学\\_百度百科](#)



格拉斯哥大学 (University of Glasgow), 简称格大, 始建于1451年, 位于英国苏格兰最大城市格拉斯哥市, 世界百强名校, 英国顶尖学府, 全球最古老的十所大学之一, 英语世界国家第四古老的大学。格大同时也是英国罗素大学集团和Universitas 2...  
[历史发展](#) [学校排名](#) [地理交通](#) [学院介绍](#) [学校建筑](#) [更多>>](#)  
<https://baike.baidu.com/>



# 知识图谱 (Knowledge Graph)

## 什么是知识图谱?

知识图谱本质上是基于图的语义网络，表示实体和实体之间的关系！

詹姆斯·瓦特，1736年1月19日 — 1819年8月25日）英国发明家，出生于苏格兰的港口小镇格林诺克，第一次工业革命的重要人物。他于1776年制造出第一台有实用价值的蒸汽机，进而开辟了人类利用能源新时代，使人类进入“蒸汽时代”……

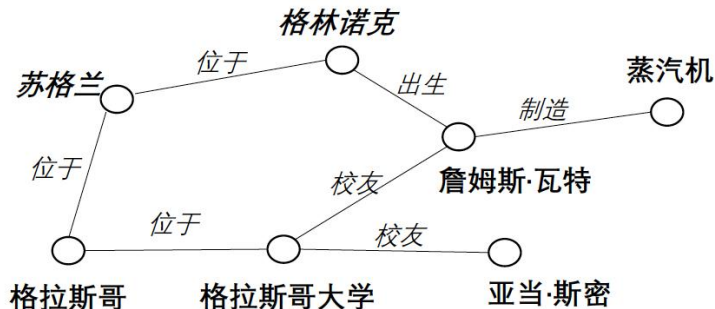
格拉斯哥大学，英国老牌名校，位于苏格兰的格拉斯哥市，始建于1451年，是全球最为古老的十所大学之一，英语世界国家第四古老大学。作为一所英国综合性古典大学，格拉斯哥大学与人类文明的发展紧密相联。经济学之父亚当·斯密、工业革命之父詹姆斯·瓦特等大批杰出校友均为社会的发展进步做出了举世瞩目的贡献。  
……

抽取

抽取的实体：  
詹姆斯·瓦特  
苏格兰  
格林诺克  
蒸汽机  
格拉斯哥大学  
苏格兰  
亚当·斯密

抽取的关系：  
出身于  
制造  
位于  
校友  
……

构建





# Part 2

## 知识图谱的源起

# 1950—1970：人工智能诞生

1956年达特茅斯会议，提出“人工智能(Artificial Intelligence, AI )”概念。

“用机器来模仿人类学习以及其他方面的智能”

“上古”流派：

- 符号主义 (Symbolism)
- 连接主义 (Connectionism)



1956，达特茅斯学  
院 知识图谱技术与应用



“人工智能来了，再过十年机器就要超越人类了！”

# 连接主义

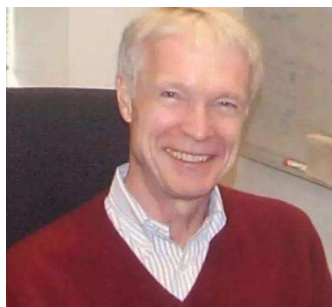
连接主义(connectionism)，又称为仿生学派(bionicsism)或生理学派(physiologism)，其主要原理为智能活动是由大量简单的单元通过复杂的相互连接后并行运行的结果。

当前典型研究：深度学习、深度神经网络

- 代表人物：



Frank Rosenblatt, 提出感知器 (1957)



John Hopfield, 提出Hopfield神经网络 (1982)

小明学骑自行车：  
经过长时间练习，小明终于学会了！  
却说不清楚“到底该怎样”骑？



# 连接主义

---

连接主义技术的主要问题是什么？

缺乏可解释性

# 符号主义

符号主义(symbolicism), 又称为逻辑主义(logicism)、心理学派(psychologism)或计算机学派(computerism), 其主要原理为认知过程就是在符号表示上的一种运算。

## 知识图谱起源于符号主义

- 代表人物:



Allen Newell

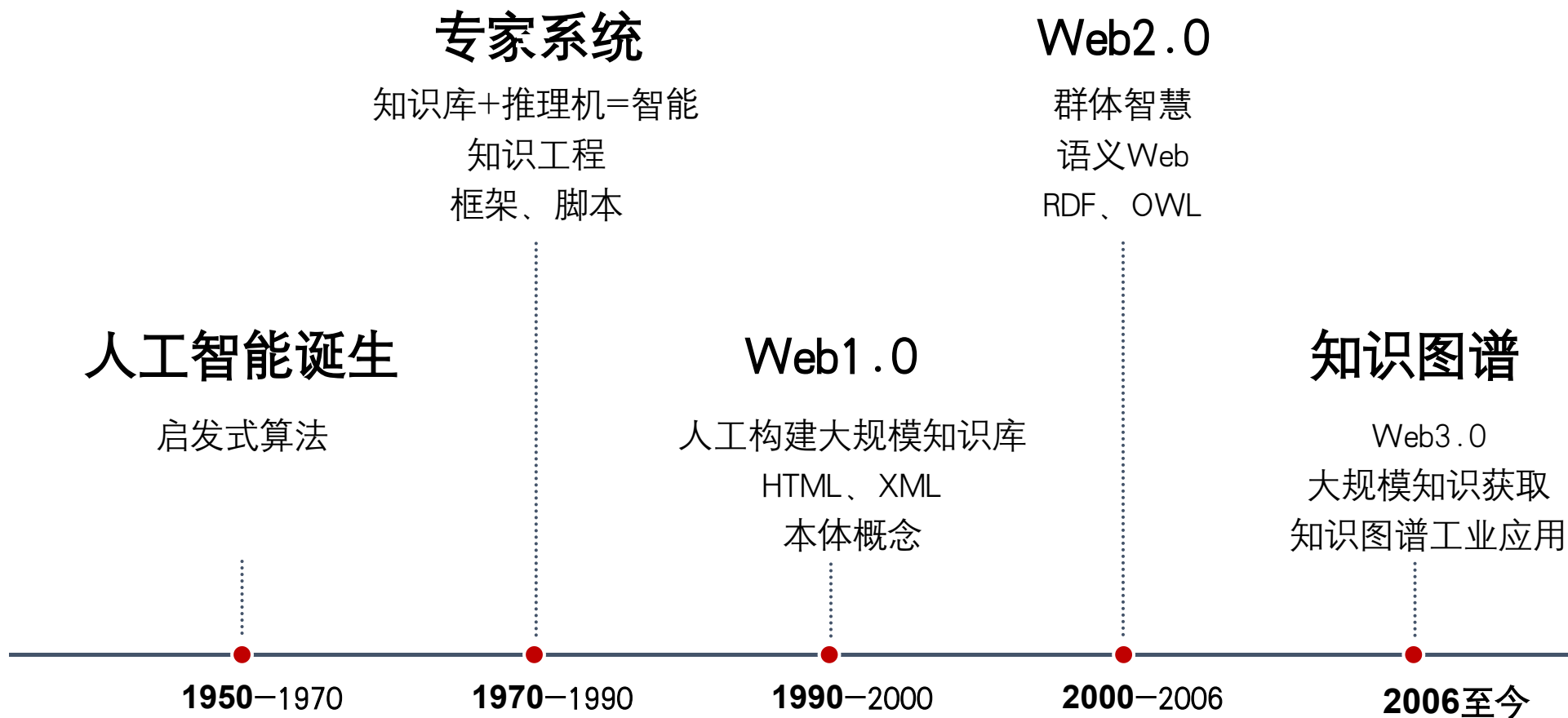


Herbert A. Simon

联合创造“The General Problem Solver”(通用问题求解程序)



# 符号主义发展历史



# 1950—1970：人工智能诞生

- 计算机有限的内存和处理速度
- 计算难度指数级增长
- 常识与推理（莫拉维克悖论）

70年代后期



人工智能发展遭遇瓶颈，出现了**第一次低谷**。

## 莫拉维克悖论：

实现人类独有的高阶智慧只需要非常少的计算能力，但是实现无意识的技能和感知却需要极大的运算能力。——**困难的问题易解，简单的问题难解**

“要让电脑如成人般地下棋是相对容易的，但是要让电脑有如一岁小孩般的感知和行动能力却是相当困难，甚至是不可能的。”



Hans Moravik

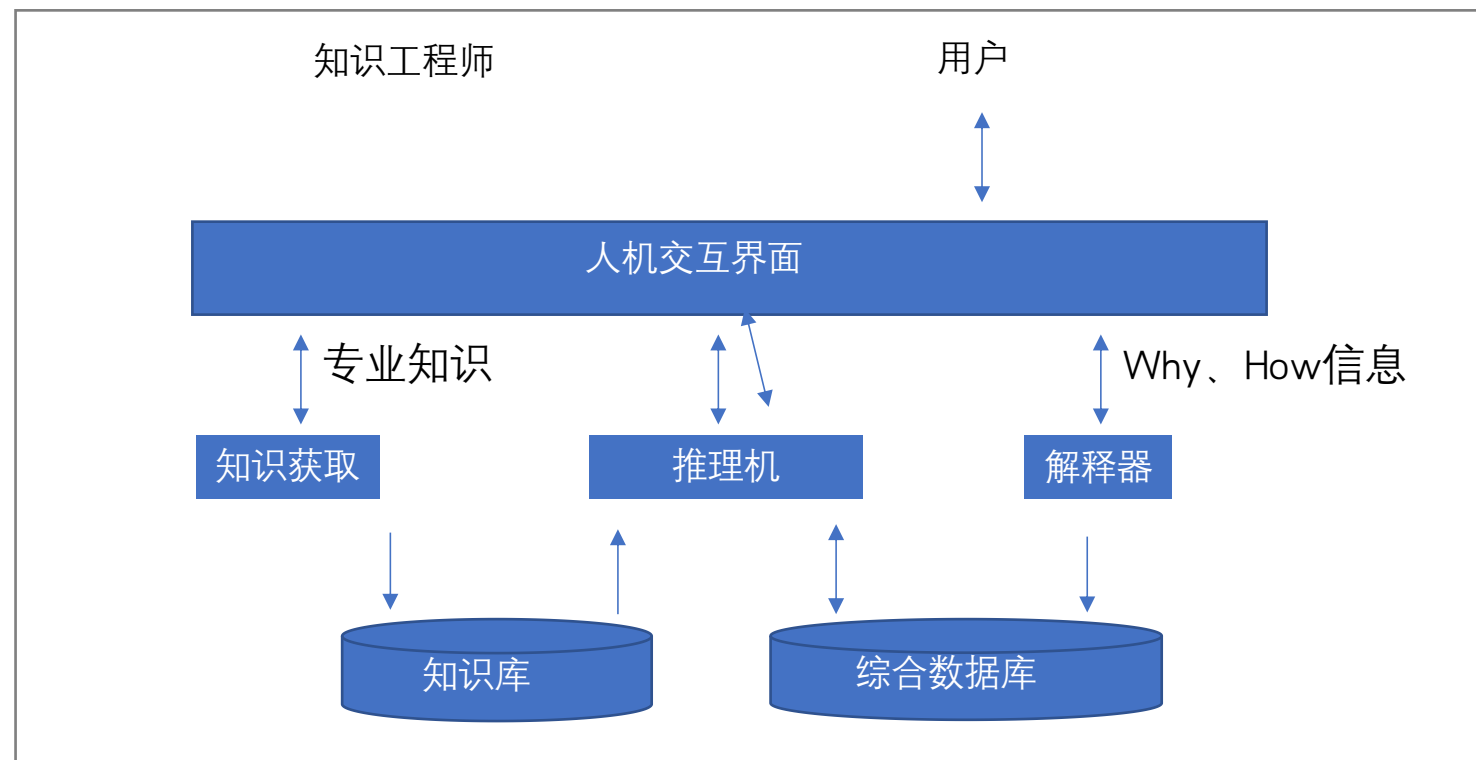


# 1970—1990：专家系统

人工智能开始转向建立基于知识的系统，通用领域——>限定领域

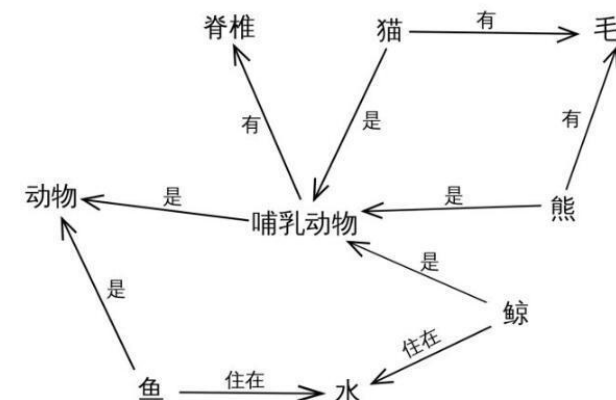
通过**知识库+推理机**实现智能。

专家系统模型：



# 1970—1990：专家系统

- 语义网络 (Semantic Network): 1970年, Herbert A. Simon正式提出, 通过有向图来表示知识, 作为知识表示的一种通用手段。



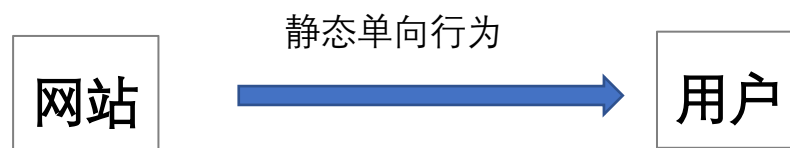
- 知识工程 (Knowledge Engineering): 1977年美国斯坦福大学计算机科学家 Edward Albert Feigenbaum教授在第五届国际人工智能会议上提出。确立了知识工程在人工智能中的核心地位。



Edward Albert Feigenbaum  
1994年图灵奖得主

# 1990—2000：万维网Web1.0

## Web 1.0：文档互联



1989年，英国科学家Tim Berners-Lee发明了万维网(World Wide Web)。

1994年，万维网联盟（World Wide Web Consortium, W3C）创建，是Web技术领域最具权威和影响力的国际中立性技术标准机构。

发布互联网内容标记语言：HTML（1997），XML（1998）

为互联网环境下大规模知识表示和共享奠定了基础。

1998年，PageRank搜索引擎技术被发明，谷歌（Google）成立。



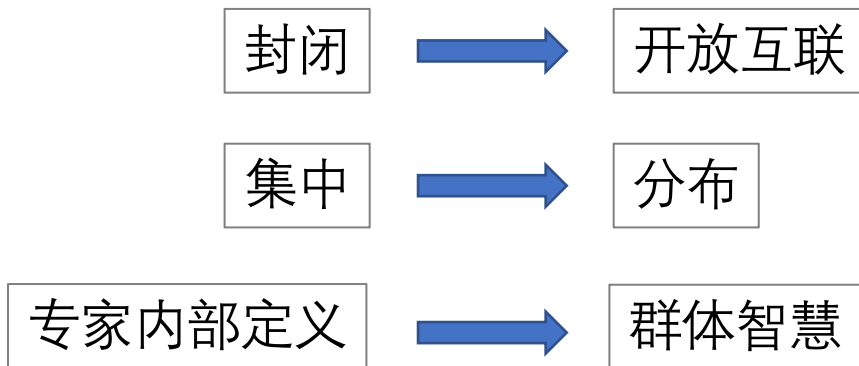
谷歌创始人：  
Lawrence Edward Page（左）  
Sergey Brin（右）

# 2000—2006：群体智慧Web2.0

## Web2.0：数据互联

强调用户生成内容，易用性，参与文化和终端用户互操作性。

### 互联网知识：



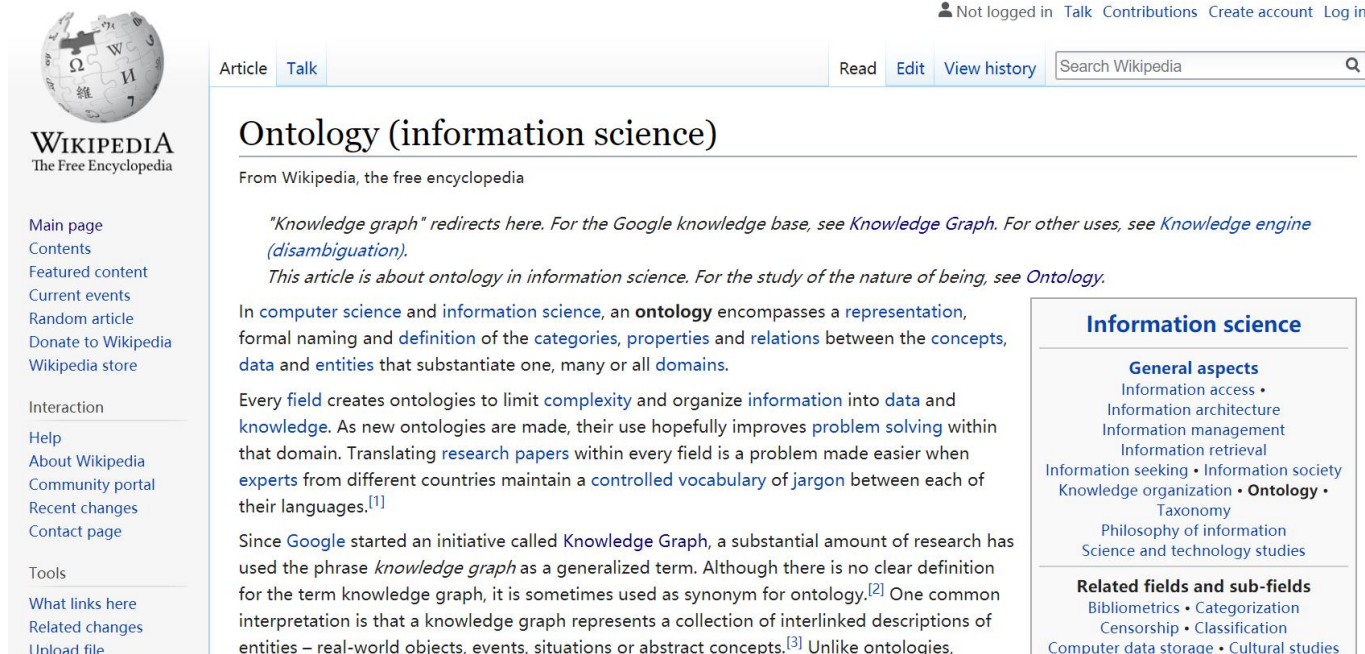
# Wikipedia

## 开放的在线多语言百科全书

- 2001年开始
- 以众包（crowdsourcing）的方式构建

## 主要特点

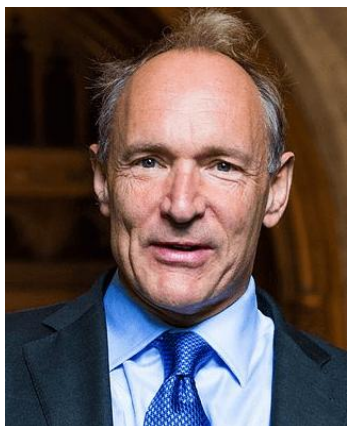
- 数据源质量高
- 500万概念
- 富含丰富语义结构的文档：
  - Infobox
  - Table
  - List
  - category



成为大规模构建知识图谱的重要数据基础！

# 语义Web

---



Tim Berners-Lee  
2016年图灵奖得主  
万维网、语义网之父

Tim Berners-Lee于2000年提出语义Web:

"I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web - the content, links, and transactions between people and computers. A "Semantic Web", which makes this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The "intelligent agents" people have touted for ages will finally materialize."

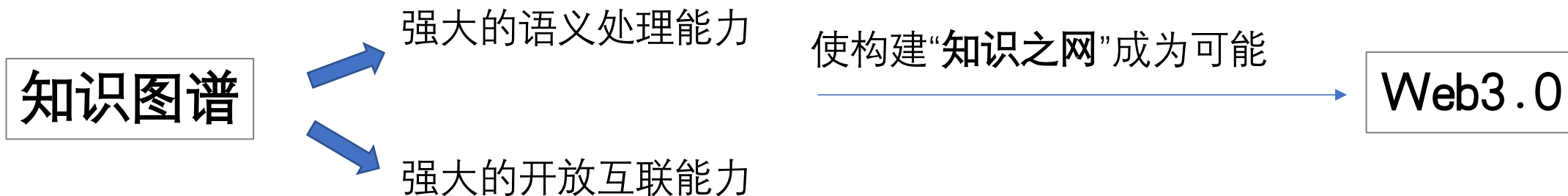
“我有一个梦想，网络中的所有计算机能够分析网络中的数据，包括内容、链接、人与计算机之间的往来。语义Web会让这一切成为可能，一旦该网络出现，日常的交易机制、事务以及我们的日常生活后都会由机器与机器之间的沟通来处理。人们吹嘘多年的“智能代理”将最终实现。”

# 2006至今：知识图谱

---

## Web3.0：知识互联

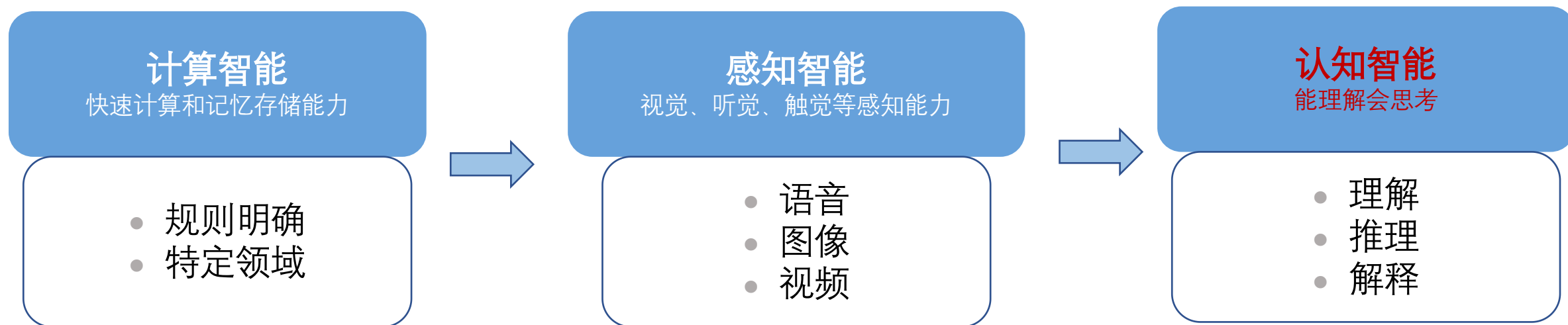
构建人与机器都可理解的万维网,使网络更加智能化。





# 知识图谱与人工智能

人工智能需要机器智能，特别是认知智能，认知智能依赖知识图谱



- **认知智能**的理解、推理、解释任务不仅需要数据，更需要**知识背景**
- 知识图谱是**知识的图谱表示**，这种知识表示方式适合理解、推理、解释
- **知识图谱**是实现**认知智能**的关键技术，是实现机器认知智能的使能器

# Part 3

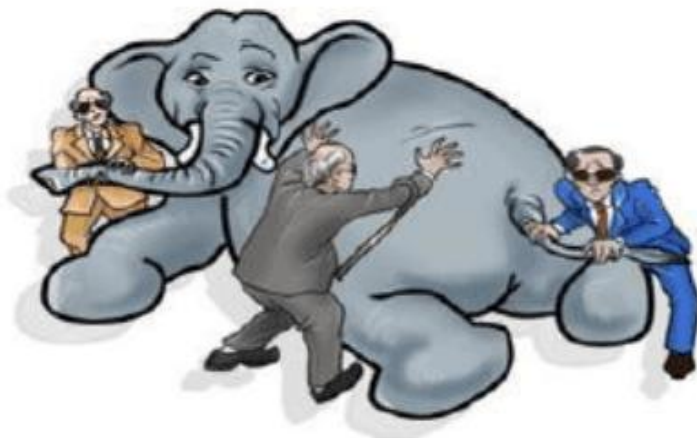
## 知识图谱技术的多个维度

# 知识图谱的相关领域发展

---

**数据库**  
图数据库

**自然语言处理**  
知识图谱构建



**机器学习**  
知识图谱分布式表示

**知识工程**  
知识图谱符号表示

# 知识图谱符号表示

---

知识图谱是Web 和大数据时代的知识工程新的发展形态。知识工程的核心：[知识库](#)和[推理引擎](#)。

- **知识抽取**：从海量的数据中通过信息抽取的方式获取知识
- **知识融合**：通过对多个相关知识图谱的对齐、关联和合并，使其称为一个有机的整体，以提供更全面知识

传统知识工程 Vs. 以知识图谱为代表的新一代知识工程

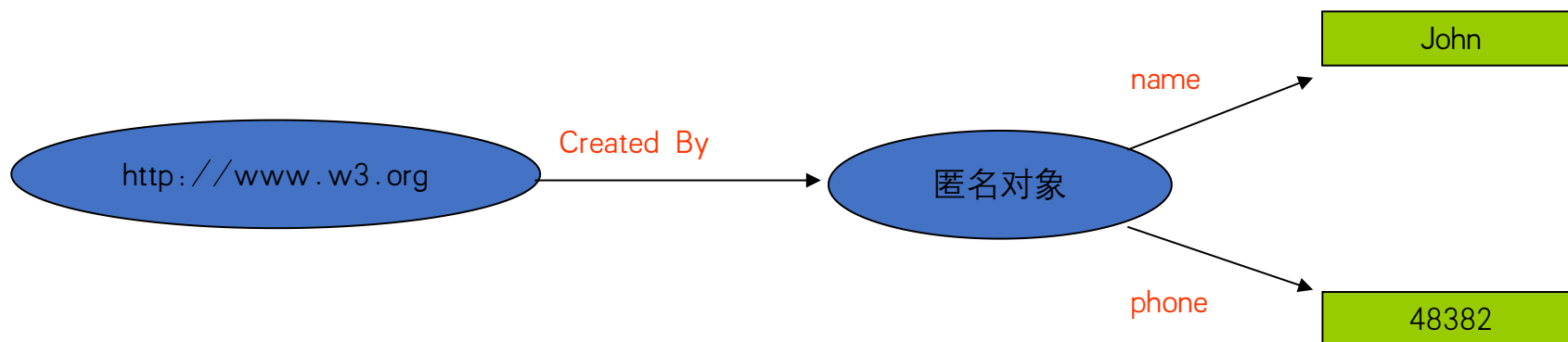
“Knowledge is the power in AI” —— Edward Albert Feigenbaum

# 知识图谱数据模型

## RDF (Resource Description Framework)

RDF定义了一个简单的模型，用于描述资源，属性和值之间的关系。资源是可以由URI标识的所有事物，属性是资源的一个特定的方面或特征，值可以是另一个资源，也可以是字符串。总的来说，一个RDF描述就是一个三元组：<主语、谓词、宾语>。

□用有向图表示的RDF示例：



# 知识图谱数据模型

## RDF (Resource Description Framework)

RDF定义了一个简单的模型，用于描述资源，属性和值之间的关系。资源是可以由URI标识的所有事物，属性是资源的一个特定的方面或特征，值可以是另一个资源，也可以是字符串。总的来说，一个RDF描述就是一个三元组：<主语、谓词、宾语>。

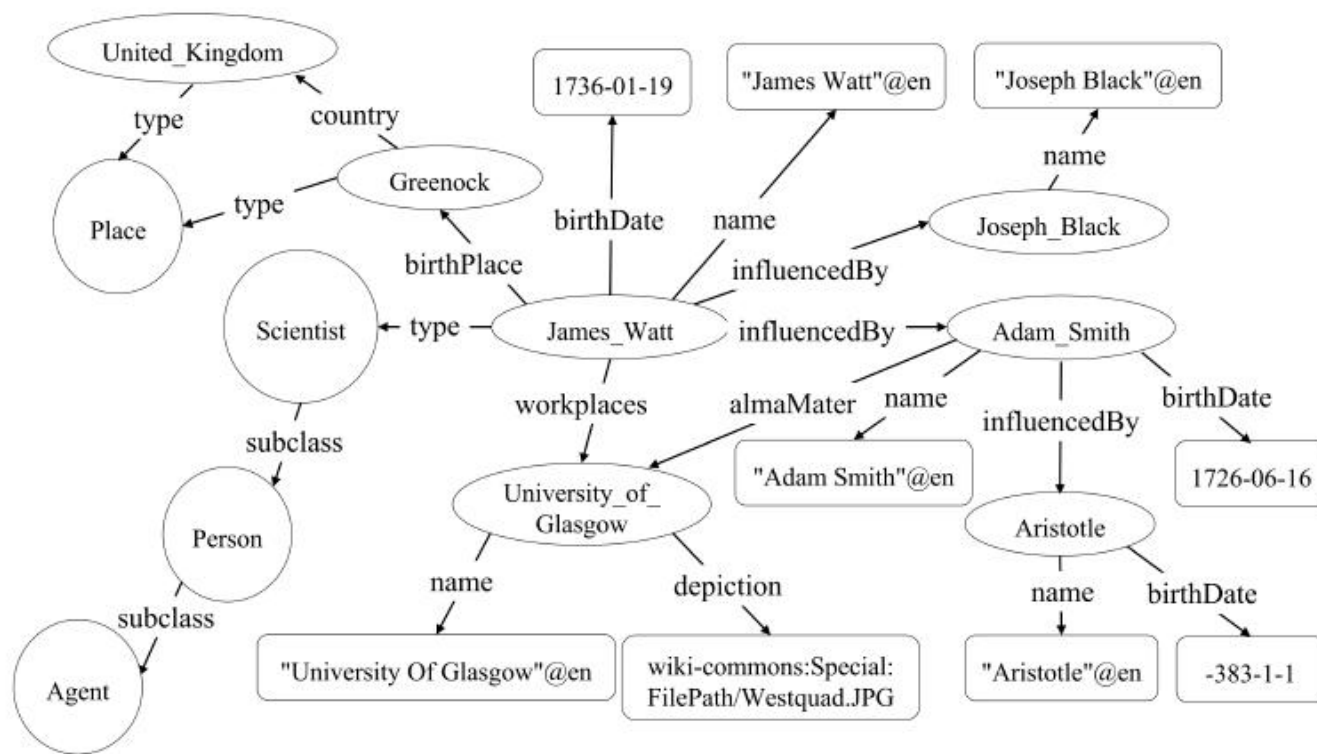
### □用列表表示的RDF示例：

Subject	Predicate	Object
http://w3.org/	created_by	#anonymous
# anonymous	name	"John"
# anonymous	phone	"477738"

# 知识图谱数据模型



RDF所表示的瓦特知识图谱：





# 知识图谱数据模型



主体	属性	客体
dbr:James_Watt	rdfs:label	"James Watt"@en
dbr:James_Watt	dbo:birthDate	"1736-01-19"^^xsd:date
dbr:James_Watt	dbo:birthPlace	dbr:Greenock
dbr:James_Watt	rdf:type	dbo:Scientist
dbr:James_Watt	dbo:influencedBy	dbr:Joseph_Black
dbr:James_Watt	dbo:influencedBy	dbr:Adam_Smith
dbr:James_Watt	dbp:workplaces	dbr:University_of_Glasgow
dbr:Adam_Smith	rdfs:label	"Adam Smith"@en
dbr:Adam_Smith	dbo:birthDate	"1723-06-16"^^xsd:date
dbr:Adam_Smith	dbo:almaMater	dbr:University_of_Glasgow
dbr:Adam_Smith	dbo:influencedBy	dbr:Aristotle
dbr:Joseph_Black	rdfs:label	"Joseph Black"@en
dbr:University_of_Glasgow	name	"University Of Glasgow"@en
dbr:Aristotle	rdfs:label	"Aristotle"@en
dbr:Aristotle	dbo:birthDate	"-383-1-1"^^xsd:date
dbr:Greenock	dbo:country	dbr:United_Kingdom
dbr:Greenock	dbo:type	dbo:Place
dbr:United_Kingdom	dbo:type	dbo:Place
dbo:Scientist	dbo:subClassOf	dbo:Person
dbo:Person	dbo:subClassOf	dbo:Agent

# 知识图谱构建

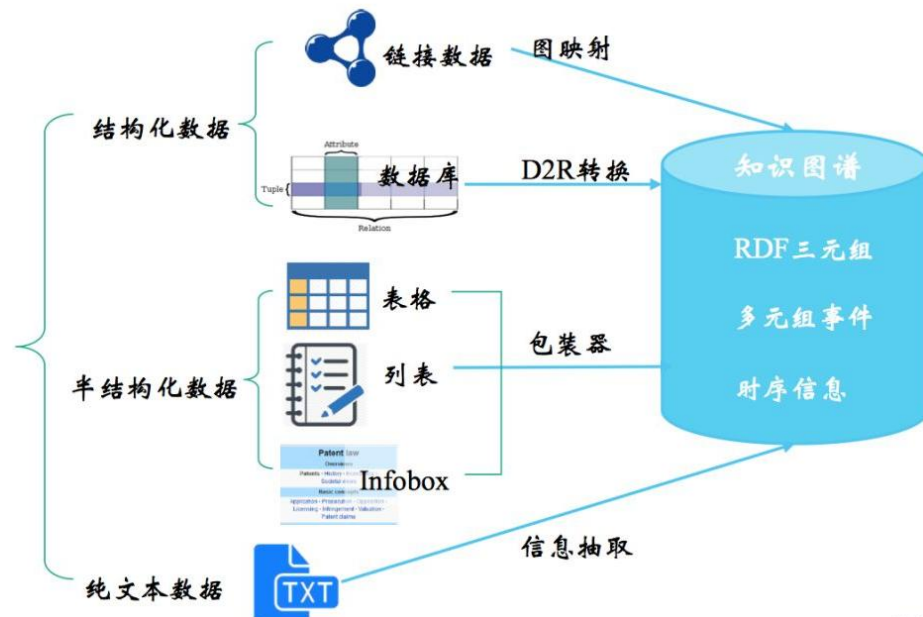
## 知识抽取

❑ **知识获取的目标**是从海量的文本数据中通过信息抽取的方式获取知识，其方法根据所处理的数据源的不同而不同。分为：

- 结构化数据
- 半结构化数据
- 非结构化文本数据

❑ **文本信息抽取**：从非结构化文本数据中进行知识抽取

- 实体识别
- 实体消歧
- 关系抽取
- 事件抽取



# 大规模知识抽取

## Yago(Yet Another Great Ontology)

2007年，由德国马普研究所发起  
融合WordNet和Wikipedia：

- 从Wikipedia的结构中抽取信息
- 利用人工采样评估
- 超过1亿事实和100种关系



Content	Entities of public Interest
Format	TSV ,RDF ,XML ,N3 ,Web Interface
Sources	Wikipedia, WordNet, Geonames
Main Strength	Focus on Precision, geotemporal annotations ,multilingual
Precision	95%
Technique	Extraction from Wikipedia + matching with WordNet & Geonames + consistency checks
Size	Entities: 3 m (+ geonames -> 10m) Facts: 120m ( + geonames -> 460m) Relations: 100, Classes: 200k, Languages: 200
License	Creative Commons BY-SA
URL	<a href="http://www.yago.com/">http://www.yago.com/</a>
References	[Suchanek, WWW 2007][Hoffart, WWW 2011] [deMelo CIKM 2010]

# 大规模知识抽取

## DBPedia

2007年开放。

目标是构建一个社区，通过社区成员定义和撰写准确的抽取模板，进而从维基百科中抽取结构信息，并将其发布到Web上。

社区通过人工的方式构建分类：

- 280个类别
- 覆盖约50%的维基百科实体



Content	Entities of public Interest
Format	RDF, API, SPARQL
Sources	Wikipedia, WordNet, YAGO
Main Strength	Focus on coberage, interlinking with other data sets
Technique	Extraction from Wikipedia +manual supervision by the community
Size	Entities: 3.5 m (in manual taxonomy: 1.7m) Facts: 670m Attributes: 9k (manually defines: 1k) Manual Classes: 280
License	CC-BY-SA & GNU FDL
URL	<a href="http://dbpedia.org">http://dbpedia.org</a>
References	[Auer, ISWC 2007], [Bizer09, JWS 2009]]

# 大规模知识抽取

## Freebase

2007年Metaweb公司发布。

2010年被Google收购。

大规模协同构建知识库。

从Wikipedia和其他数据源(如 IMDB、MusicBrainz

中导入知识

核心思想:

- 在Wikipedia中, 人们编辑文章
- 在Freebase中, 人们编辑结构化知识

Content	Entities of public Information
Format	RDF, API
Construction	By the community Data import from public sources
Sources	Wikipedia, Libraries, WordNet, MusicBrainz...
Main Strength	Free and large
Size	Facts: several: millions Entities: 20m
License	CC-BY(Creative Commons Attribution)
URL	<a href="http://www.freebase.be/">http://www.freebase.be/</a>



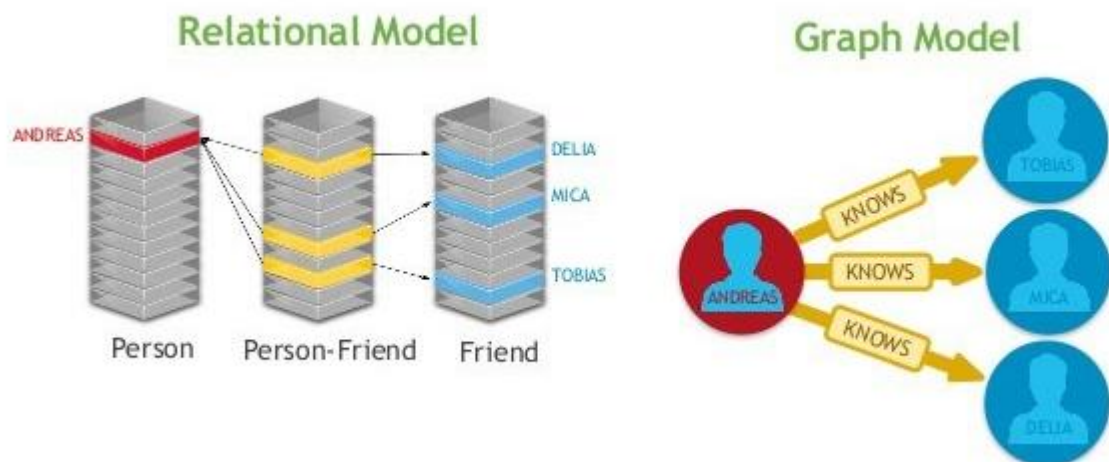
# 知识图谱管理

知识图谱本质上是多关系图，通常用“**实体**”来表达图里的**结点**、用“**关系**”来表达图里的**边**。

**关系型数据库**：实体与实体之间的关系通常都是利用外键来实现，对关系的查询需要大量join操作



**图数据库**：图模型建模实体（结点）和实体之间的关系（边），在对关系的操作上有更高的性能



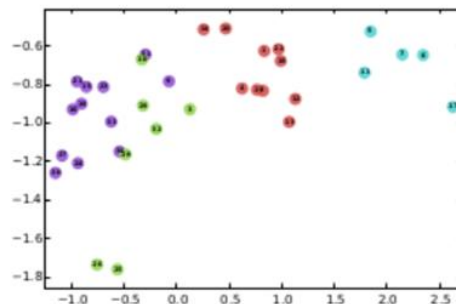
# 知识图谱的分布式表示

## 知识表示学习：

- **背景**：基于网络形式的知识表示存在数据稀疏问题和计算效率问题。
- **知识表示学习**（representation learning）：主要是面向知识图谱中的实体和关系进行表示学习，使用建模方法将实体和向量表示在低维稠密向量空间中，然后进行计算和推理。
- **优点**：显著提升计算效率，有效缓解数据稀疏，实现异质信息融合。
- **应用**：知识图谱补全、相似度计算、关系抽取、自动问答、实体链指



(a) Input: Karate Graph



(b) Output: Representation

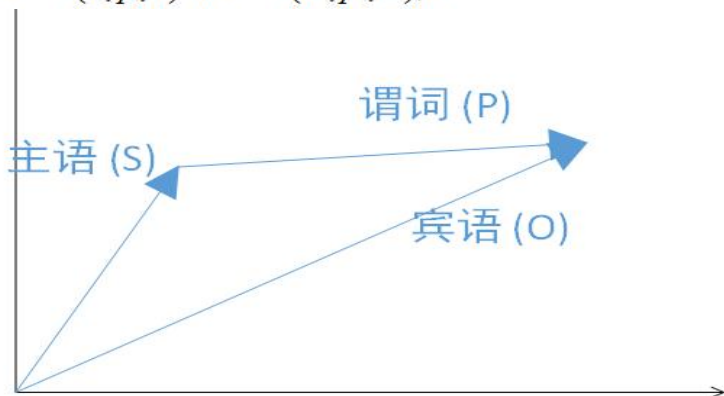


# 知识表示学习

## 知识表示学习:

- **举例:** 知识表示代表模型: TransE [Bordes et al., NIPS 13]。
- 对每个事实(Subject, Predicate, Object), 将其中的predicate作为从subject到object的翻译操作。
- 每个Subject/Predicate/Object, 都映射成一个多维向量
- 优化目标:  $S+P=O$

$$\Gamma = \sum_{(s,p,o) \in S} \sum_{(s',p',o') \notin S} [r + d(s+p,o) - d(s'+p',o')]_+$$



S	P	O
China	Capital	Beijing
Canada	Capital	Ottawa
.....	.....	.....

$$\begin{aligned} \text{Beijing} - \text{China} \\ \approx \\ \text{Ottawa} - \text{Canada} \end{aligned} = \text{Capital}$$

# THANK YOU

---