

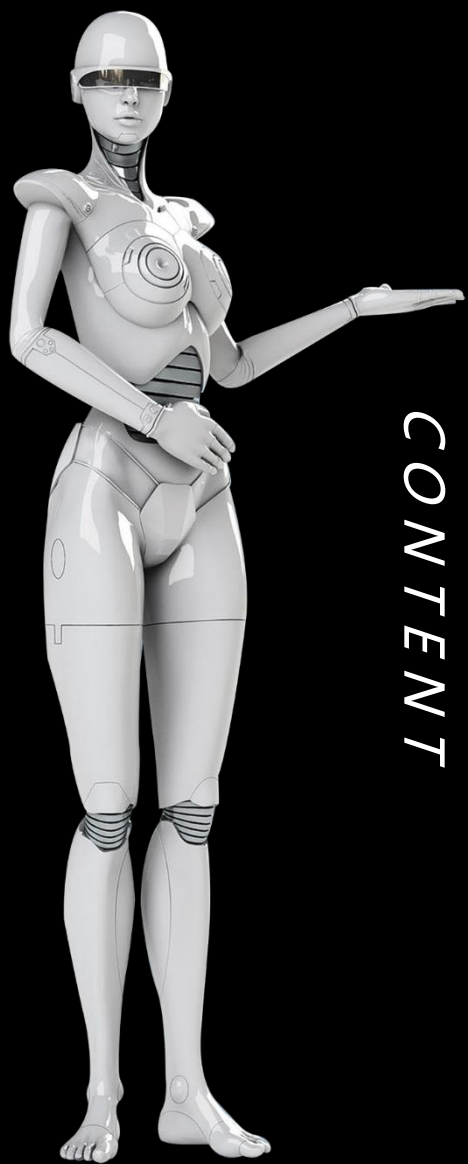
第三章 知识表示与推理

Knowledge representation and reasoning

计算机学院计科系 赵曼
zhaoman@cug.edu.cn

中国地质大学 (武汉)





CONTENT

课程内容

1. 知识表示
2. 知识推理
3. 知识图谱

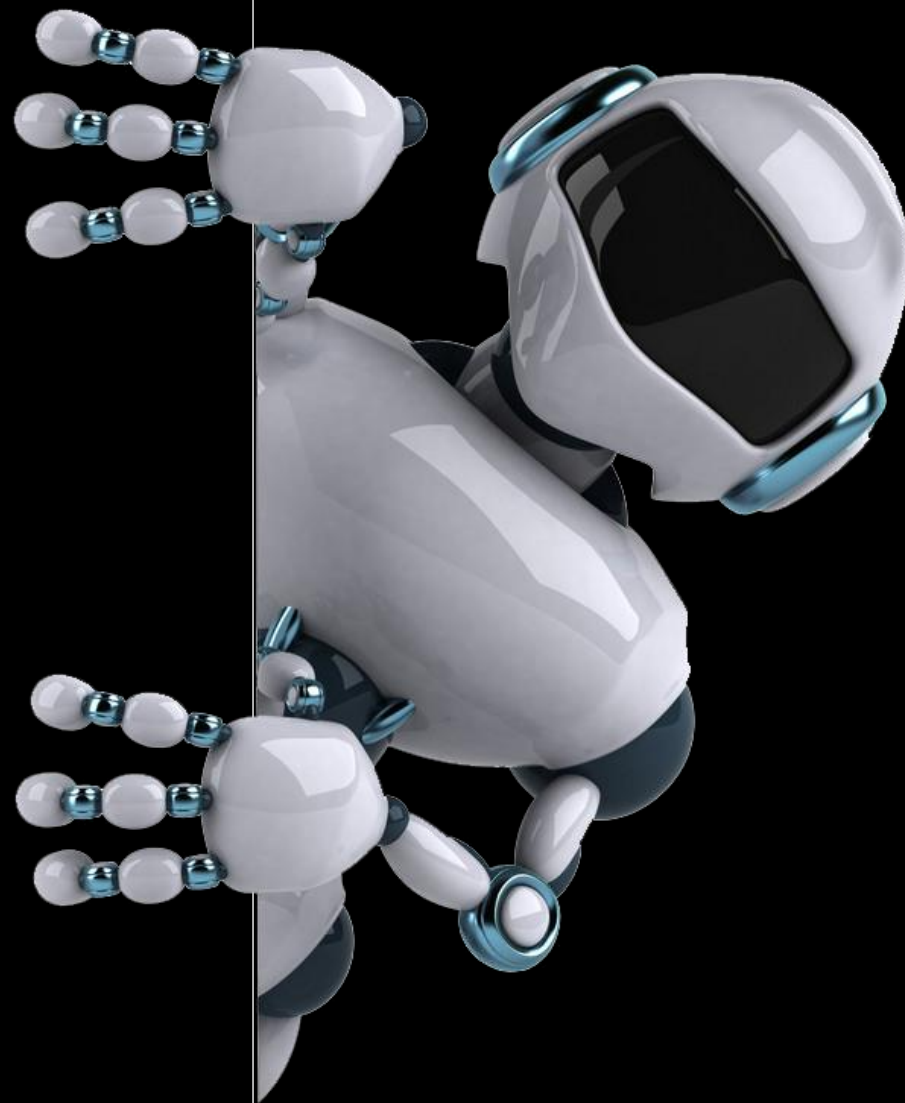




PART.01

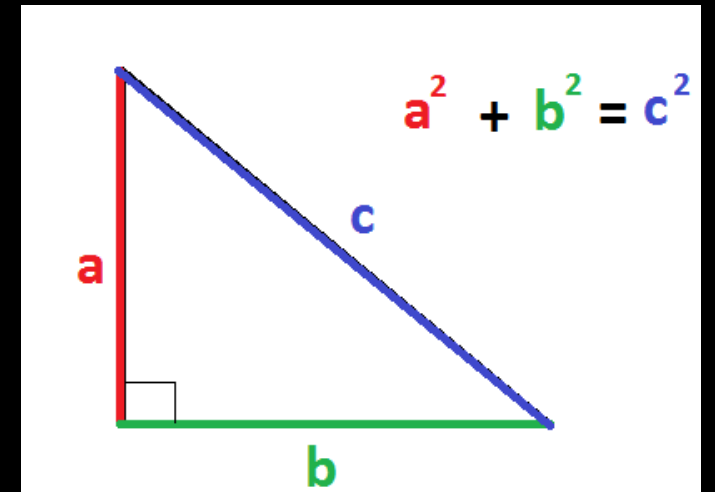
知识表示

Knowledge representation



生活中的“知识”

- “如果秋天来了，银杏树叶就会变黄”
- “地球的形状是扁球体”
- “勾股定理”
-





什么是知识

- 知识是人们在长期的生活及社会实践中、在科学研究及实验中积累起来的对客观世界的认识与经验
- 人们把实践中获得的信息关联在一起，就形成了知识



- 知识是智能的基础，为了使计算机具有智能，能够模拟人类的智能行为，就必须使它具有知识
- 人类的知识需要用适当的形式表示出来，才能存储到计算机中并被运用

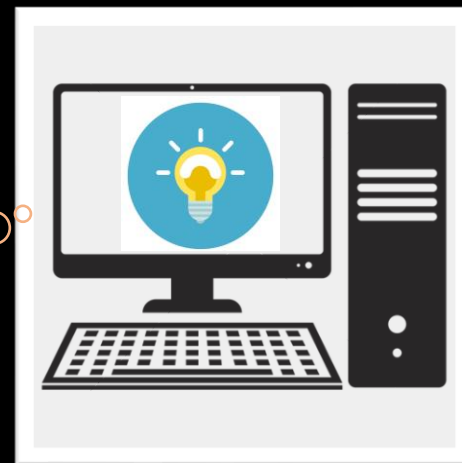
“朝霞不出门，晚霞行千里”

无法理解！



“IF：早晨有朝霞
THEN：今天的天气可能会不好”

可以理解！





规则性知识与事实性知识

- 规则性知识

- 规则性知识反映了信息间的某种因果关系，可以用“如果.....则.....”的形式来表示
- 如：“如果动物是鸟，则该动物是卵生生物”

- 事实性知识

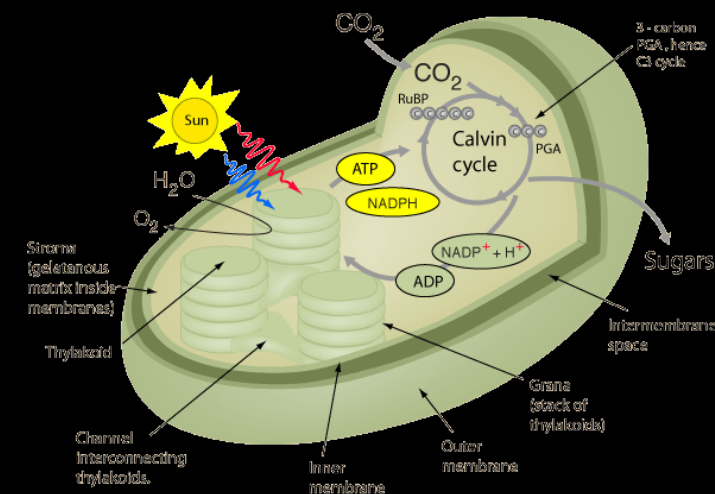
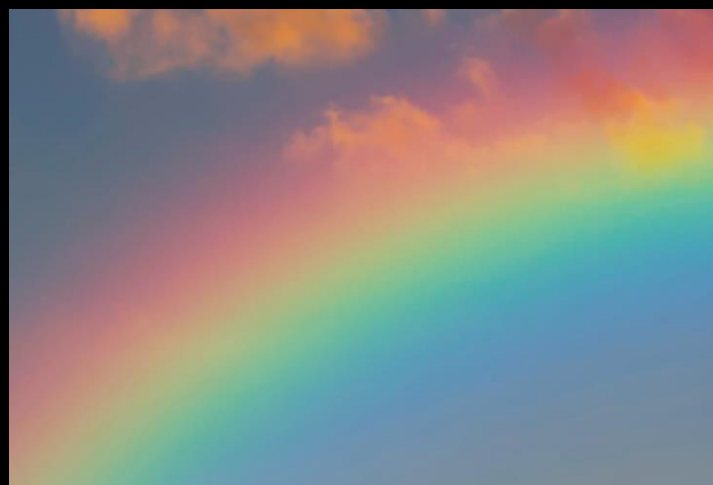
- 事实性知识反映了事物的某些性质
- 如：“篮球是圆的”，“水是透明的”



生活中的不确定性



- “下雨天路上可能会堵车”
- “空气湿度较大可能会出现彩虹”
- “能够进行光合作用的细胞可能为植物细胞”





确定性知识与不确定性知识

- 确定性知识

- 表示确定的规则或事实
- 如：“如果这个粒子为电子，则这个粒子一定带负电”

- 不确定性知识

- 事物与事物间的规则与联系是不确定和模糊的
- 如：“如果今天阴天，则明天可能会下雨”



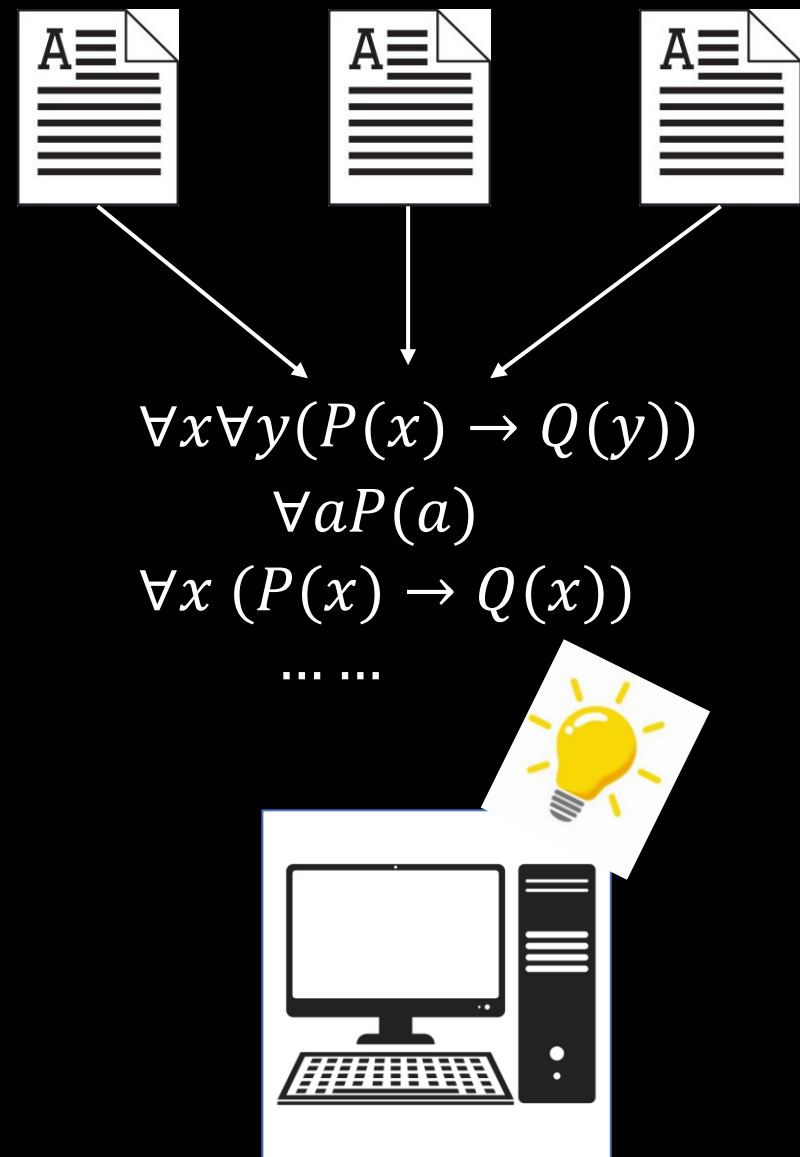


知识表示

- 概念：
 - 知识表示就是将人类知识形式化或者模型化
- 目的：
 - 能够让计算机贮存、理解并运用这些知识来解决复杂的现实世界问题
- 知识表示方法：
 - 一阶谓词逻辑表示法、产生式表示法、框架式表示法、状态空间表示法等

一阶谓词逻辑表示法

- 一阶谓词逻辑表示法是一种重要的知识表示方法，以数理逻辑为基础，是能够表达人类思维活动规律的一种精准的形式语言
- 一阶谓词逻辑表示法与人类的自然语言比较接近，可以被方便地输入到计算机中进行存储和运算
- 一阶谓词逻辑表示法是一种应用于人工智能早期的知识表示方法





个体和谓词

- **个体**：具体的事物或抽象的概念
 - 如：北京大学、姚明、地球等代表具体的事物，大学、运动员、行星等代表抽象的概念
- **谓词**：个体的性质或个体之间的的关系
 - 如：雪是白色的，地球的半径比月球的半径大
 - 用 $F(x)$ 来表示个体 x 具有某种性质 F
 - 如： $White(snow)$ 代表雪是白色的
 - 用 $G(x, y)$ 表示个体 x 和个体 y 之间具有某种关系 G
 - 如： $Long(earth, moon)$ 代表地球的半径比月球的半径大

- 量词：表示数量的词
 - **全称量词**：即所有的、全部的、任一个、每一个、都，用符号 \forall 表示(ALL)
 - $\forall x$ 表示对于所有的 x
 - **存在量词**：即至少存在一个、有一个，用符号 \exists 表示(Exist)
 - $\exists x$ 表示至少存在一个 x 或有一个 x



一阶谓词逻辑的知识表示

- “所有有理数都可以写成分数的形式”
 - x 代表数
 - $rational(x)$: 表示 x 是有理数
 - $fraction(x)$: 表示 x 可以写成分数的形式
- 上述知识通过一阶谓词逻辑就可以表示为:

$$\forall x(rational(x) \rightarrow fraction(x))$$



一阶谓词逻辑的知识表示

- “有的人喜欢吃糖”
 - x 代表人
 - $sweet(x)$: 表示 x 喜欢吃糖
- 上述知识通过一阶谓词逻辑就可以被表示为:

$$\exists x(sweet(x))$$

产生式表示法



- 一阶谓词逻辑有很强的表达能力，但也有许多缺点，如形式过于灵活，不容易统一，无法表示不确定知识等等
- 因此，在人工智能中，更常用的一种知识表示方式是产生式表示法，意思是能够根据已知条件产生新知识的式子
- 产生式这一术语最早是由美国数学家波斯特在1943年提出，如今已成为人工智能中应用最多的一种知识表示方法
- 产生式通常用于表示事实、规则以及它们的不确定性度量





确定性规则的产生式表示

- 基本形式:

IF P THEN Q

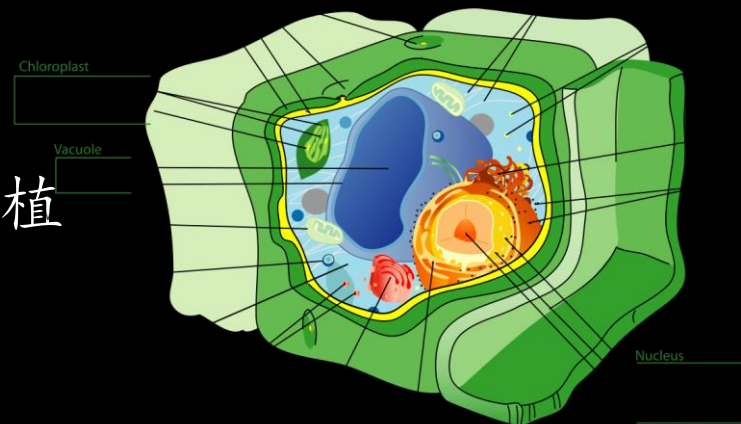
- 或者:

$P \rightarrow Q$

- P是产生式的前提，用于指出该产生式是否可用的条件，称为规则的前件
- Q是一组结论或操作，如果前提P满足，则结论Q成立或执行Q所规定的操作，称为规则的后件

确定性规则的产生式表示

- IF 细胞有细胞壁 THEN 该细胞为植物细胞
 - 细胞有细胞壁是这一产生式的前提，该细胞为植物细胞是这一产生式的结论
- IF 今天下雨 THEN 带雨伞
 - 今天下雨是这一产生式的前提，带雨伞为当今天下雨这一前提成立时需要执行的操作





不确定性规则的产生式表示

- 基本形式:

$IF\ P\ THEN\ Q$ (置信度)

- 或者:

$P \rightarrow Q$ (置信度)

- 表示当前提中的各条件都得到满足时, 结论可以相信的程度
- 这里用置信度表示知识的强度

不确定性规则的产生式表示

- IF 动物会飞 THEN 该动物是鸟 (0.5)
 - 当“动物会飞”这一条件成立时，结论“该动物是鸟”可以相信的程度为0.5
- IF 今天阴天 THEN 明天会下雨 (0.6)
 - 当“今天阴天”这一条件成立时，结论“明天会下雨”可以相信的程度为0.6

也有可能为昆虫或其他生物





确定性事实的产生式表示

- 基本形式:

(对象, 属性, 值)

- 或者:

(关系, 对象1, 对象2)

- 如:

- (北京, 气温, 3) 表示北京的气温是3摄氏度
 - 北京是描述的对象, 气温是描述对象的属性, 3是属性的值
- (朋友, 小明, 小华) 表示小明与小华是朋友
 - 小明与小华是所描述的对象, 朋友是两个对象间的关系



不确定性事实的产生式表示

- 基本形式:

(对象, 属性, 值, 置信度)

- 或者:

(关系, 对象1, 对象2, 置信度)

- 如:

- (北京, 气温, 3, 0.8) 表示北京的气温是3摄氏度这一事实的可相信程度为0.8
- (朋友, 小明, 小华, 0.1) 表示小明和小华是朋友这一事实的可相信程度为0.1

框架表示法



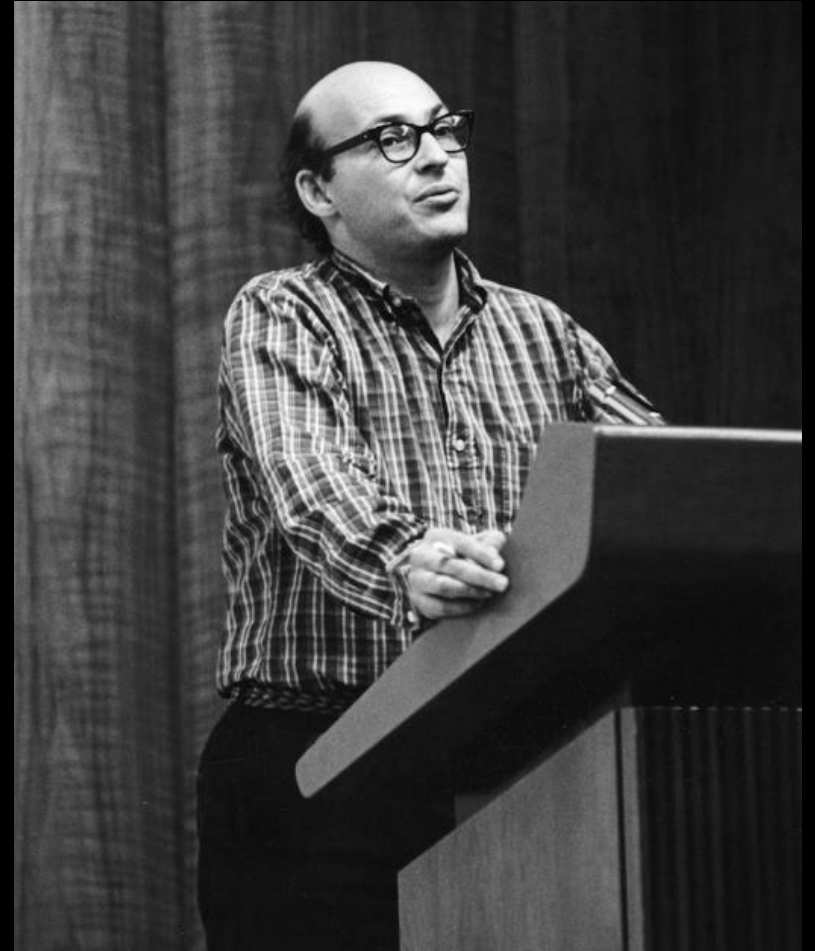
- 当向一个人描述一辆汽车的具体信息之前，他就能根据以往对“车辆”这一概念的认识，形成对这一辆车的基本印象：
 - 车一定包含：方向盘，座位，底盘和轮胎
 - 可能包含的部件：多媒体设施，GPS等等
 - 可能涉及的活动：起步，行驶，后退等等
- 尽管对这辆车的具体信息，如：座位个数、底盘高度等细节还不了解，但是通过以往见到过的车辆，已经在大脑中建立了关于“汽车”的框架



什么是“框架”



- 1975年，人工智能学者明斯基（Minsky）提出了框架理论：
 - 这一理论基于人们对现实世界中各种事物的认识都以一种类似于框架的结构存储在记忆中
 - 当面临一个新事物时，就从记忆中找到一个合适的框架，并根据实际情况对细节加以修改、补充，从而形成对当前事物的认识
- 框架表示法是一种结构化的知识表示方法，它有利于将信息组织到系统中





框架的一般结构

- 框架 (frame) 是一种描述所论对象属性的数据结构, 一个框架由若干个槽 (slot) 组成
- 每一个槽可根据实际情况划分为若干个侧面 (facet), 每一个侧面有若干个侧面值

<框架名>			
槽名1:	侧面名 ₁₁	值 ₁₁₁ , 值 ₁₁₂ , 值 ₁₁₃	约束条件1
	侧面名 ₁₂	值 ₁₂₁ , 值 ₁₂₂ , 值 ₁₂₃	约束条件2
槽名2:	侧面名 ₂₁	值 ₂₁₁ , 值 ₂₁₂ , 值 ₂₁₃	约束条件3
	侧面名 ₂₂	值 ₂₂₁ , 值 ₂₂₂ , 值 ₂₂₃	约束条件4
。 。 。			

建立汽车框架



- 汽车框架

框架名： <汽车>

座位

轮子

功能

- 对于上述框架，当把具体信息填入槽或侧面后，就得到了相应框架的一个事例框架

建立汽车框架



- 卡车事例:

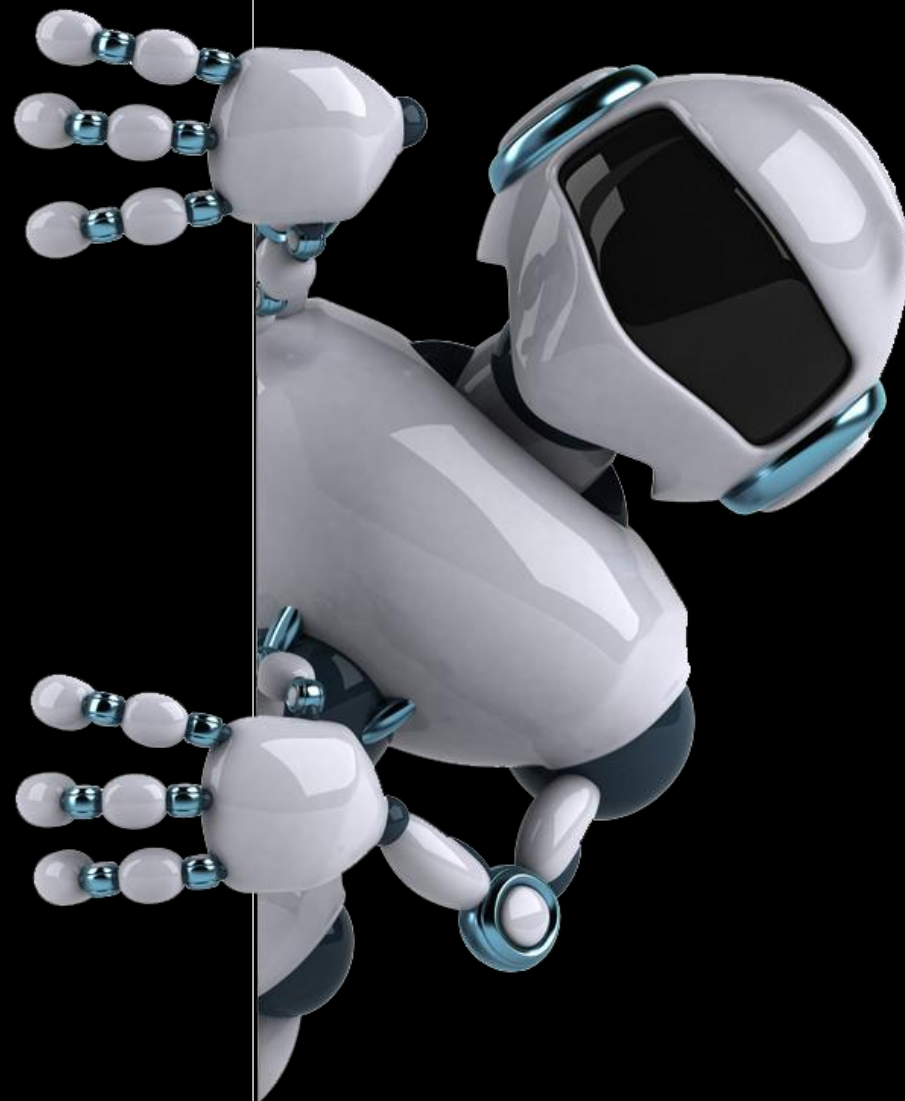
<卡车>			
槽	侧面	侧面值	约束
座位	座位数:	2	2/3/4
	座位材质:	皮革	织物/皮革
轮子	轮子数:	4	正整数
	是否防滑:	是	是/否
功能	载重:	6	正整数



PART.02

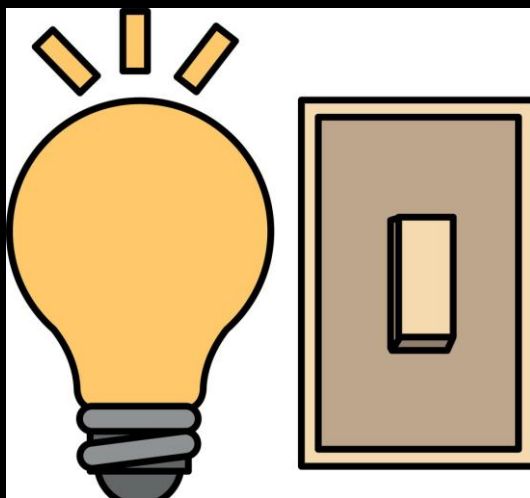
知识推理

Knowledge reasoning



什么是推理

- 推理1: 所有平面三角形的内角和都等于 180° ，这个形状是平面三角形，所以这个形状的内角和等于 180°
- 推理2: 只有开启电源开关，电灯才会亮，这盏电灯是亮的，所以电源开关是开启的
- 推理3: 会飞的动物可能是鸟，这只动物会飞，所以这只动物可能是鸟



什么是推理

- 推理是由一个或几个已知的判断（前提）推出新判断（结论）的过程
- 计算机或智能系统模拟人类的智能推理方式，依据推理控制策略，利用形式化的知识进行机器思维和求解问题的过程，叫做知识推理





正向推理

- 概念:

- 正向地使用规则，从已知条件出发向目标进行推理

- 基本思想:

1. 检验是否存在规则的前提被已知事实满足
2. 如果满足，则将该规则的结论放入已知事实中，再检查是否存在其他规则的前提被满足
3. 重复前两步过程
4. 直到目标被推出或再也没有新结论被推出



正向推理

- 已知条件:

- 已知 A, C, E 成立, 求证 F 成立

- 规则:

$$\begin{aligned} r_1: & \text{ IF } A \text{ THEN } B \\ r_2: & \text{ IF } B \text{ and } C \text{ THEN } D \\ r_3: & \text{ IF } D \text{ and } E \text{ THEN } F \end{aligned}$$

- 正向推理:

- 初始时, A, C, E 为已知事实, 根据规则 r_1 推出 B 成立, 将 B 加入已知事实中
- 根据规则 r_2 推出 D 成立, 将 D 加入已知事实中
- 根据规则 r_3 推出 F 成立, 将 F 加入已知事实中, 由于 F 是求证目标, 结果成立, 推理结束



逆向推理

- **概念:**

- 假设求解目标成立，逆向使用规则进行推理

- **基本思想:**

- 1.将目标作为假设，查看是否有某条规则支持该假设
- 2.查看这些规则的前提是否成立
- 3.如果前提成立，则假设被验证，将结论放入已知事实中，否则将前提放入假设集中，逐一验证这些假设
- 4.直到目标假设被验证或无法根据已知事实推出目标假设成立



逆向推理

- 已知条件:

- 已知 A, C, E 成立, 求证 F 成立

- 规则:

$r_1:$ IF A THEN B
 $r_2:$ IF B and C THEN D
 $r_3:$ IF D and E THEN F

- 逆向推理:

- 规则 r_3 的结论可以推出 F 成立, 检查 r_3 的前提发现 D 不在已知事实中, 将 D 加入假设集
- 将 D 作为假设, 规则 r_2 的结论可以推出 D 成立, 检查 r_2 的前提发现 B 不在已知事实中, 将 B 加入假设集
- 将 B 作为假设, 规则 r_1 的结论可以推出 B 成立, r_1 的前提 A 在已知事实中, 从而 B 成立, 从而 D 成立, 从而 F 成立

动物种类识别



- 专家根据掌握的知识建立关于虎、豹、鸵鸟、企鹅的分类规则，形成产生式：

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 为了判断当前图片中动物的种类，能够从图片中获取到的已知事实如下：



f_1 : 有羽毛
 f_2 : 会游泳



正向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 正向推理:

- 根据已知事实 f_1 : 有羽毛 与规则 r_5 : IF 有羽毛 THEN 鸟类 推出 f_3 : 鸟类 成立, 将 f_3 加入已知事实中



正向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳
 f_3 : 鸟类

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 正向推理:

- 根据已知事实 f_2 : 会游泳 和 f_3 : 鸟类 与规则 r_6 : IF 鸟类 and 会游泳 THEN 企鹅 推出事实 f_4 : 企鹅 成立, 将 f_4 加入已知事实中



正向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳
 f_3 : 鸟类
 f_4 : 企鹅

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 正向推理:

- 由于 f_4 是求解目标, 结果成立, 推理结束



逆向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 逆向推理:

- 假设该动物的种类是企鹅, 得到假设 h_1 : 企鹅, 将假设 h_1 加入假设集



逆向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳

- 假设集:

h_1 : 企鹅

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 逆向推理:

- 规则 r_6 : IF 鸟类 and 会游泳 THEN 企鹅 的结论可以推出假设 h_1 成立, 检查规则 r_6 的前提, 发现 鸟类 不在已知事实中, 得到假设 h_2 : 鸟类, 将假设 h_2 加入假设集



逆向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳

- 假设集:

h_1 : 企鹅
 h_2 : 鸟类

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 逆向推理:

- 规则 r_5 : IF 有羽毛 THEN 鸟类的结论可以推出假设 h_2 : 鸟类 成立, 检查规则 r_5 的前提 有羽毛 与已知事实 f_1 符合, 所以假设 h_2 成立, 将事实 f_3 : 鸟类 加入已知事实中



逆向推理

- 已知事实:

f_1 : 有羽毛
 f_2 : 会游泳
 f_3 : 鸟类

- 假设集:

h_1 : 企鹅

- 规则:

r_1 : IF 有毛 THEN 哺乳动物
 r_2 : IF 哺乳动物 and 食肉 THEN 食肉动物
 r_3 : IF 食肉动物 and 黑条纹 THEN 虎
 r_4 : IF 食肉动物 and 黑斑 THEN 豹
 r_5 : IF 有羽毛 THEN 鸟类
 r_6 : IF 鸟类 and 会游泳 THEN 企鹅
 r_7 : IF 鸟类 and 善于奔跑 THEN 鸵鸟

- 逆向推理:

- 规则 r_6 : IF 鸟类 and 会游泳 THEN 企鹅 的前提与已知事实 f_2 和 f_3 符合, 所以假设 h_1 成立, 得到目标假设成立, 推理结束

非确定性推理

- 现实生活中的许多事实与规则都是非确定性的，需要非确定性的推理方法
- **概念：**
 - 非确定性可以理解为在缺少足够信息的情况下做出判断
 - 非确定性推理就是从不确定性初始证据出发，通过运用不确定性的知识，最终推出具有一定程度的不确定性但却是合理或者近乎合理的结论





规则运算

- 规则的置信度可以理解为当规则的前提为真时结论的置信度
- 当规则的前提不一定为真时，规则结论的置信度可以用如下方式计算
- 已知：

IF A THEN B

规则的前提为真时结论的置信度为： $CF(B, A)$

前提A成立的置信度为： $CF(A)$

- 则最终结论B成立的置信度为：

$$CF(A) \times CF(B, A)$$

规则运算



- 已知:

IF 堵车 *THEN* 迟到
 $CF(\text{迟到}, \text{堵车}) = 0.8$
 $CF(\text{堵车}) = 0.5$

- 则最终迟到的置信度为:

$$0.8 \times 0.5 = 0.4$$

- 从规则中得到迟到的置信度为0.4





规则合成

- 通常情况下，得到同一个结论的规则可能不止一条，但是从不同规则得到同一个结论的置信度可能不同
- 如：

IF 堵车 *THEN* 迟到
 $CF(\text{迟到}, \text{堵车}) = 0.8$
 $CF(\text{堵车}) = 0.5$

IF 发生事故 *THEN* 迟到
 $CF(\text{迟到}, \text{发生事故}) = 1.0$
 $CF(\text{发生事故}) = 0.1$

- 根据第一条规则可以得到迟到的置信度为：0.4
- 根据第二条规则可以得到迟到的置信度为：0.1

- 迟到这一结论的置信度究竟有多少？

- 已知：

IF A_1 *THEN* B

$$CF(B, A_1) = b_1$$

$$CF(A_1) = a_1$$

IF A_2 *THEN* B

$$CF(B, A_2) = b_2$$

$$CF(A_2) = a_2$$

- 则：

结论B成立的置信度为： $a_1 \times b_1 + a_2 \times b_2 - a_1 \times b_1 \times a_2 \times b_2$

规则合成



- 已知:

IF 堵车 *THEN* 迟到
 $CF(\text{迟到}, \text{堵车}) = 0.8$
 $CF(\text{堵车}) = 0.5$

IF 发生事故 *THEN* 迟到
 $CF(\text{迟到}, \text{发生事故}) = 1.0$
 $CF(\text{发生事故}) = 0.1$

- 迟到的置信度为:

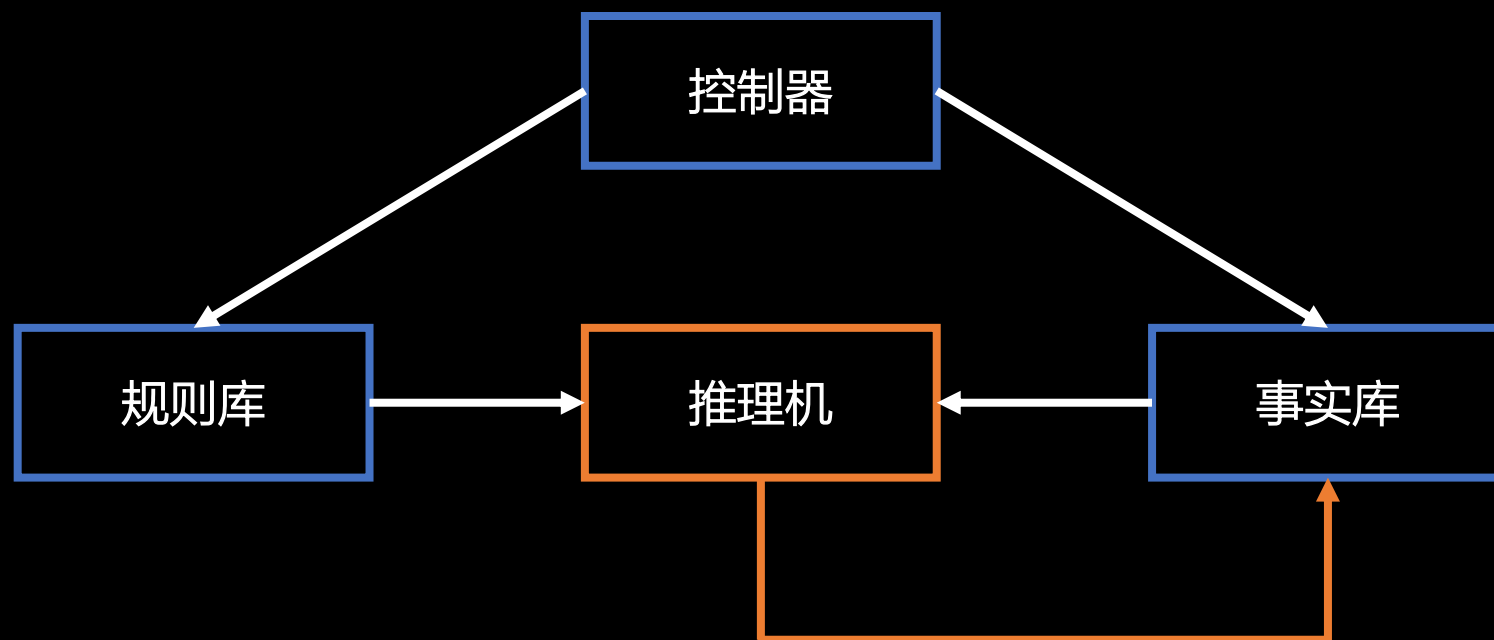
$$0.8 \times 0.5 + 1.0 \times 0.1 - 0.8 \times 0.5 \times 1.0 \times 0.1 = 0.46$$

- 从规则中得到迟到的置信度为0.46



产生式系统

- 将一组产生式放在一起，相互配合，一个产生式生成的结论可以供另一个产生式作为已知事实使用，这样来求得问题的解，这样的系统称为产生式系统
- 产生式系统是知识表示与知识推理的一个综合运用





产生式系统

- 规则库

- 存放相应领域内知识的产生式集合

- 事实库

- 存放已有的事实，以及通过推理得到的新的事实
- 事实库在推理进行的过程中是不断变化的

- 推理机

- 推理机读取事实库和规则库，将事实与规则的前提进行匹配，以产生新的事实

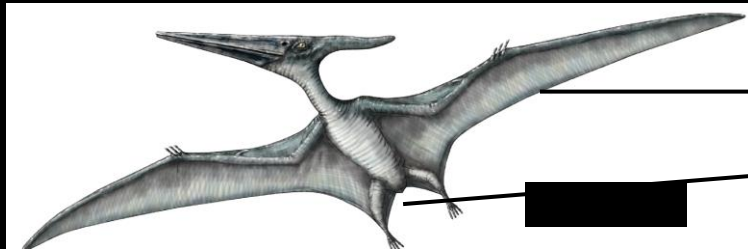
恐龙种类识别



- 专家根据掌握的恐龙的相关知识建立了关于无齿翼龙、翼手龙、三角龙和霸王龙的分类规则，创建了规则库：

r_1 : IF 有翅膀 THEN 会飞
 r_2 : IF 会飞 and 没有尾巴 THEN 无齿翼龙
 r_3 : IF 会飞 and 有尾巴 THEN 翼手龙
 r_4 : IF 没有翅膀 and 食草 THEN 三角龙
 r_4 : IF 没有翅膀 and 食肉 THEN 霸王龙

- 为了判断当前图片中恐龙的种类，能够从图片中抽取出事实库：



f_1 : 有翅膀
 f_2 : 没有尾巴



推理机的工作

- 事实库:

f_1 : 有翅膀
 f_2 : 没有尾巴

- 规则库:

r_1 : IF 有翅膀 THEN 会飞
 r_2 : IF 会飞 and 没有尾巴 THEN 无齿翼龙
 r_3 : IF 会飞 and 有尾巴 THEN 翼手龙
 r_4 : IF 没有翅膀 and 食草 THEN 三角龙
 r_4 : IF 没有翅膀 and 食肉 THEN 霸王龙

- 匹配推理:

- 产生式 r_1 : IF 有翅膀 THEN 会飞 的前提与事实 f_1 : 有翅膀一致, 得到 会飞 的新事实, 向事实库中添加 f_3 : 会飞



推理机的工作

- 事实库:

f_1 : 有翅膀
 f_2 : 没有尾巴
 f_3 : 会飞

- 规则库:

r_1 : IF 有翅膀 THEN 会飞
 r_2 : IF 会飞 and 没有尾巴 THEN 无齿翼龙
 r_3 : IF 会飞 and 有尾巴 THEN 翼手龙
 r_4 : IF 没有翅膀 and 食草 THEN 三角龙
 r_4 : IF 没有翅膀 and 食肉 THEN 霸王龙

- 匹配推理:

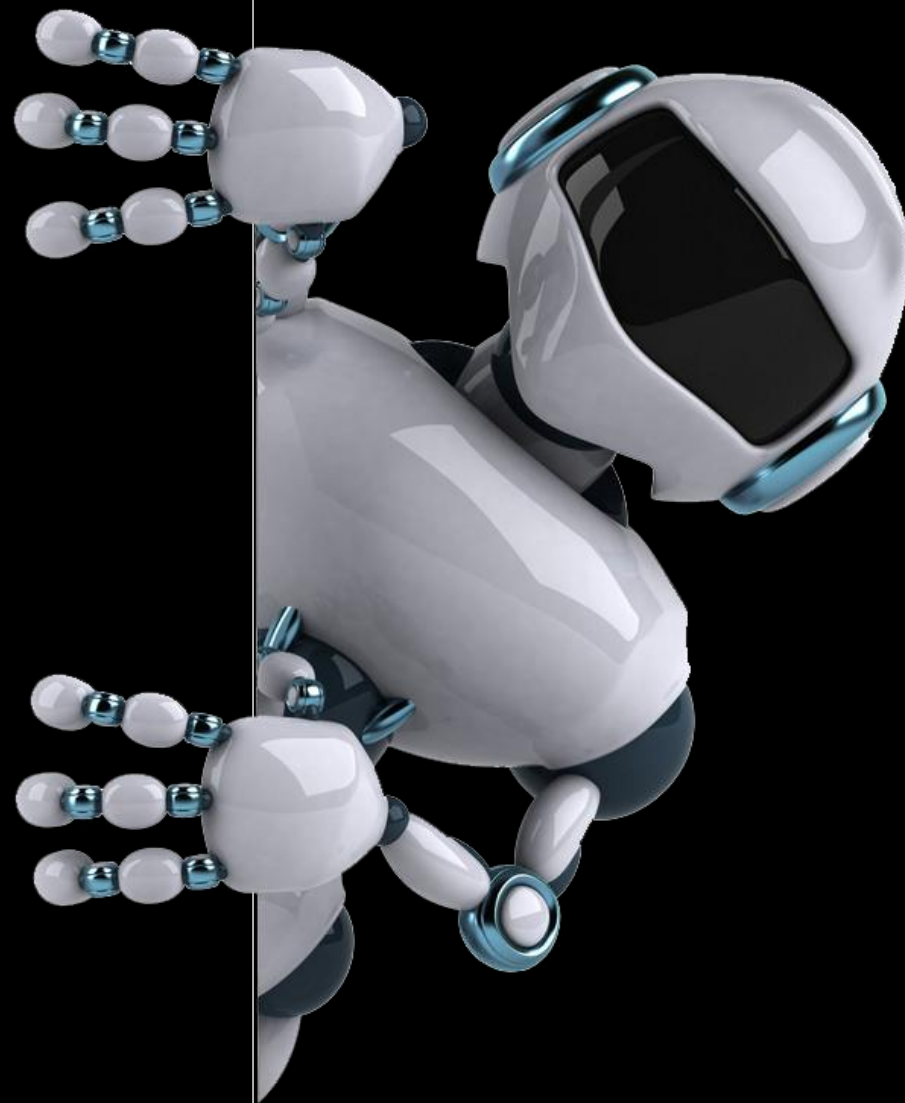
- 产生式 r_2 : IF 会飞 and 没有尾巴 THEN 无齿翼龙的前提与事实 f_2 和 f_3 一致, 得到新事实 f_4 : 无齿翼龙, 得到明确分类结论, 推理结束



PART.03

知识图谱

Knowledge Graph



- 数据库是以一定方式储存在一起的数据集合
- 知识库是用于知识管理的一种特殊的数据库，以便于有关知识的采集、整理以及提取
- 知识库中存储的知识是求解问题所需领域的知识的集合，包括基本事实、规则和其他有关信息
- 知识库中的知识数据以结构化的方式进行存储



知识库举例



- 李白:

- (李白, 性别, 男)
- (李白, 朝代, 唐)
- (李白, 身份, 诗人)
- (李白, 好友, 孟浩然)
- (李白, 代表作, 《蜀道难》)

.....

实体

关系

实体

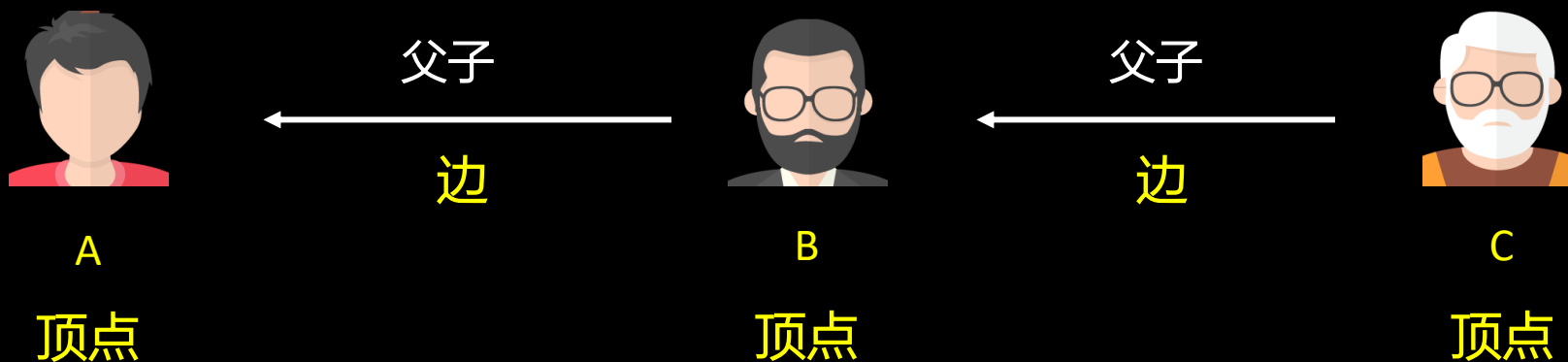
结构化存储
的知识数据



关系图

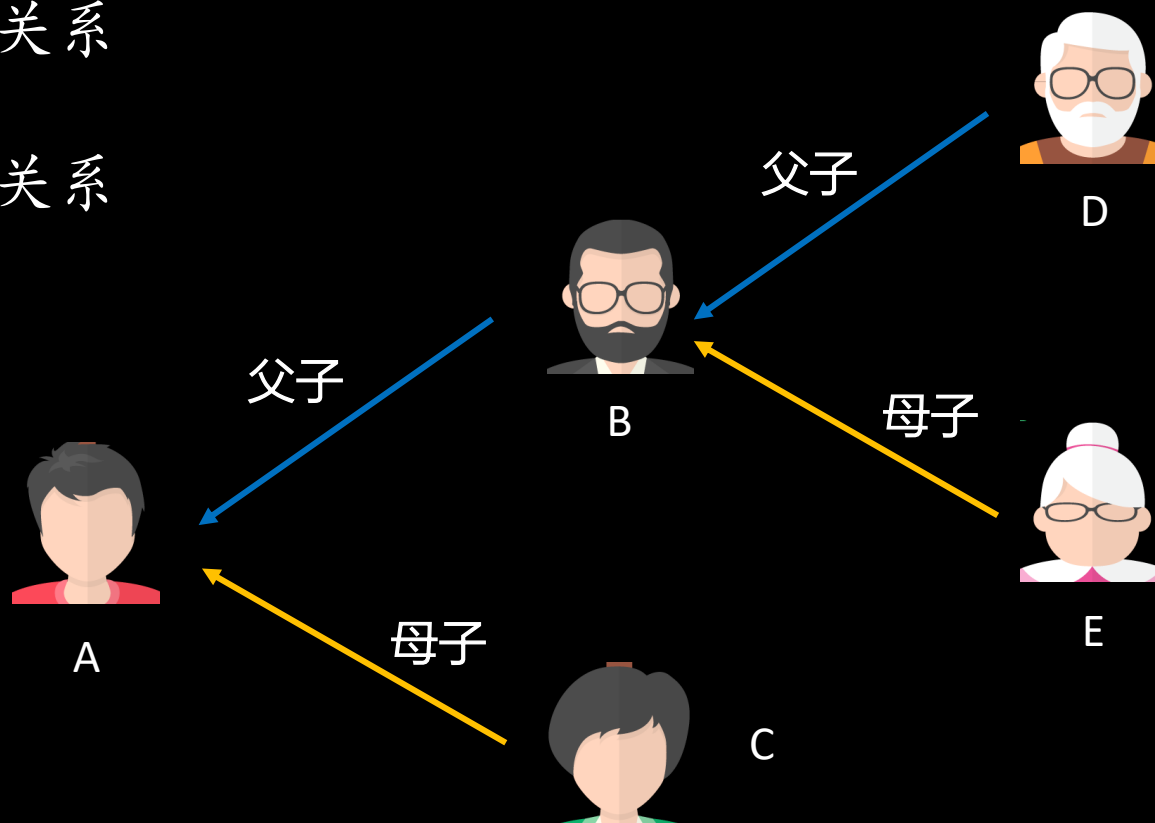


- 图是由若干给定的顶点及连接两顶点的边所构成的图形
- 顶点用于代表事物，连接两顶点的边用于表示两个事物间具有某种关系
- 如：
 - B与A是父子关系，C与B是父子关系



多关系图

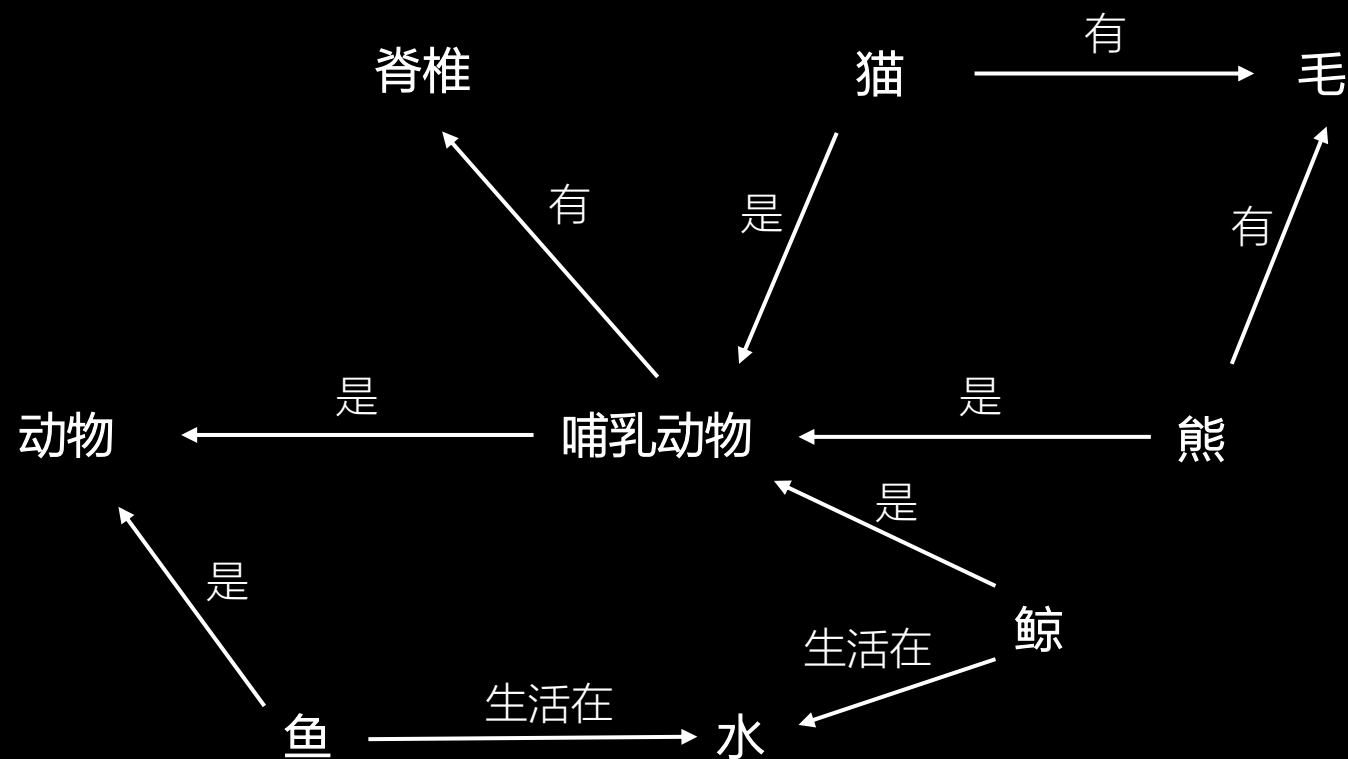
- 多关系图中一般包含多种类型的顶点和多种类型的边
- 当需要描述更多事物时，他们之间的关系可能就不仅仅局限于一种，如：
 - B与A是父子关系，C与A是母子关系
 - D与B是父子关系，E与B是母子关系





语义网络

- 语义网络是一种用关系图/多关系图来表示知识的结构化方式
- 语义网络中顶点表示概念，边表示的是这些概念之间的语义关系



知识图谱的概念

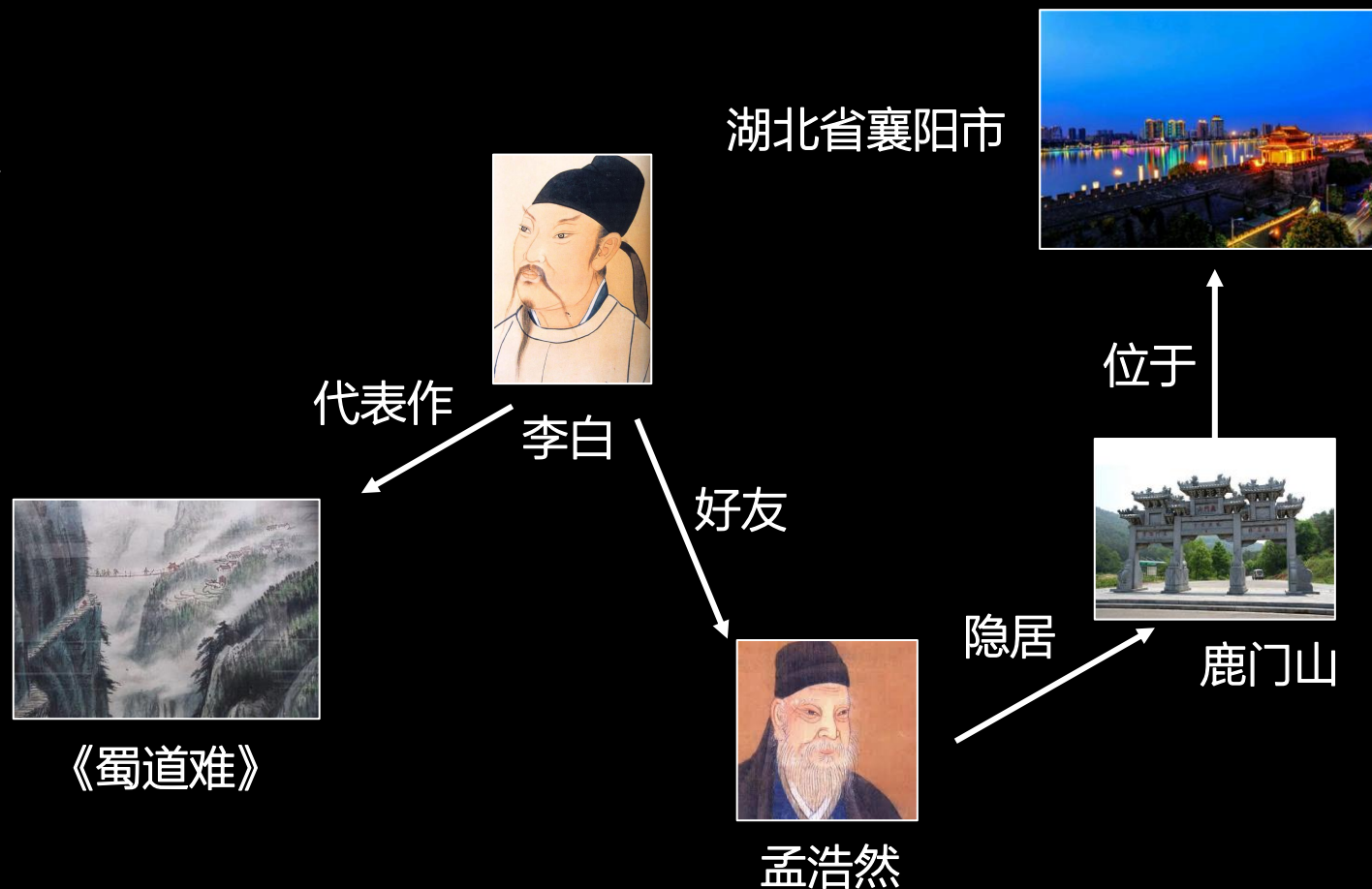
- 知识图谱本质上是一种用语义网络来进行知识表示的知识库，图的顶点代表实体或概念，边代表实体/概念之间的关系
- 从功能上看，知识图谱以结构化的形式描述客观世界中概念、实体间的复杂关系，将信息表达成更接近人类认知世界的形式



知识图谱举例

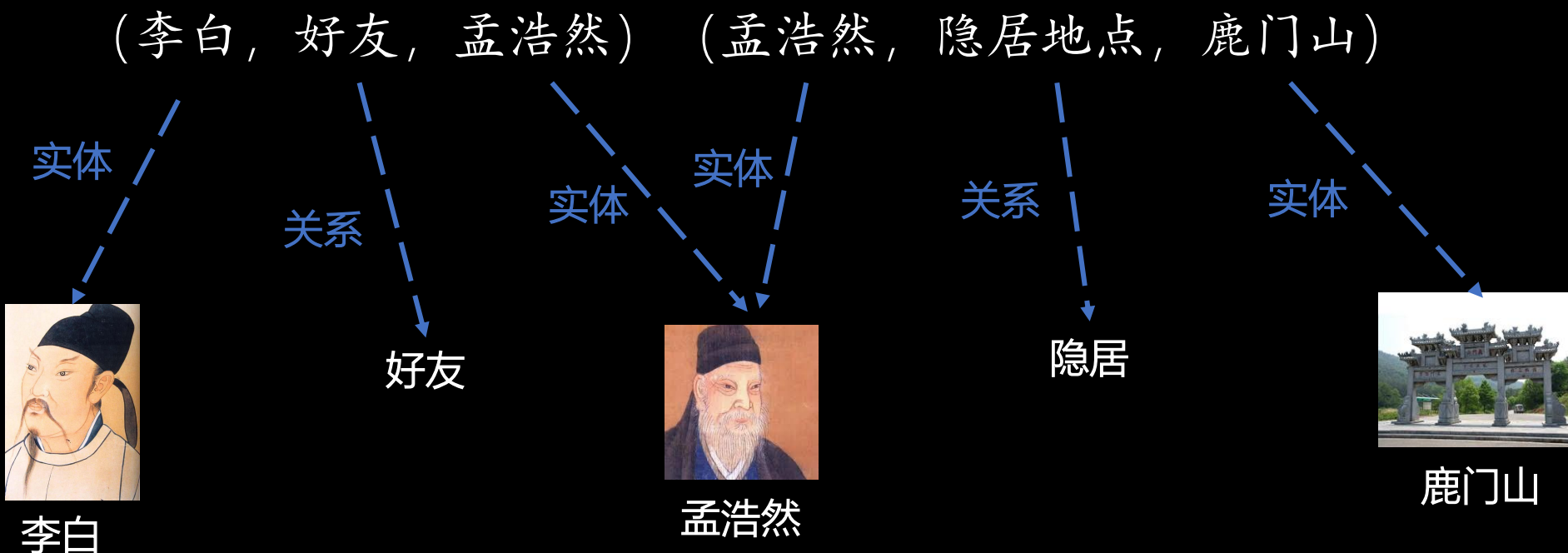


- 中国唐朝诗人李白的代表作是《蜀道难》
- 李白的好友是孟浩然
- 孟浩然隐居的地点是鹿门山
- 鹿门山地处湖北省襄阳市



知识图谱的生成

- 通过将知识数据中的概念/实体与关系进行相互的匹配和推理，将概念/实体转化为图的顶点，概念/实体间的关系转化为图的边，进而构建出相对应的知识图谱



知识数据的来源

- 知识数据的来源主要有以下两种：
 - 通过知识库中结构化存储的知识数据
 - 通过从网页上抓取的自然语言片段，经过信息抽取后得到的知识数据





结构化知识数据的来源

- Wikidata

- 维基数据是一个可协同编辑的知识库，由维基媒体基金会托管

值

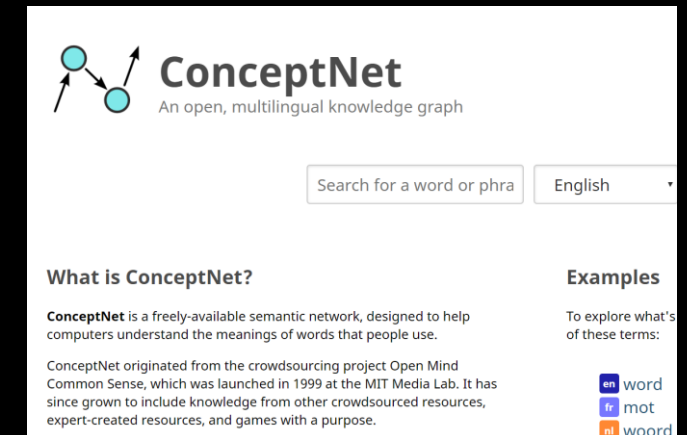
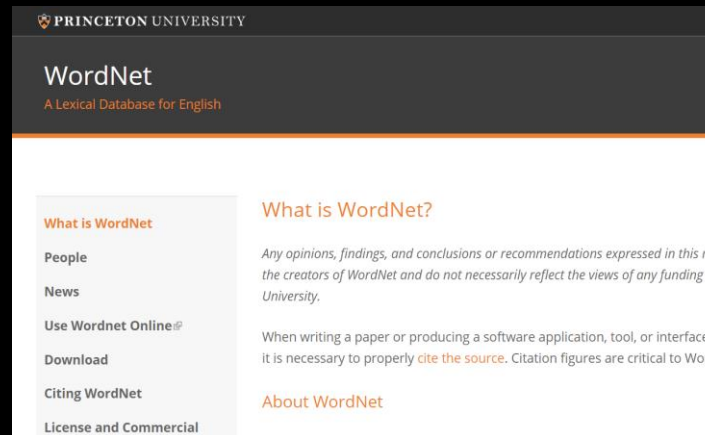
属性

信息来源

The screenshot displays Wikidata data for Li Bai (李白) with three property-value pairs. Annotations include: 'sex or gender' (属性) pointing to the property label, 'male' (值) pointing to the value, 'country of citizenship' (属性) pointing to the property label, 'Tang Empire' (值) pointing to the value, '1 reference' (信息来源) pointing to the reference list, and 'China Biographical Database' (信息来源) pointing to the specific reference source.

Property (属性)	Value (值)	Reference (信息来源)
sex or gender	male	3 references
country of citizenship	Tang Empire	1 reference: China Biographical Database
birth name	李白 (Chinese)	0 references


- • • • •





从网页抓取的自然语言数据

- 通过搜索“李白”的相关资料，可以从网页上抓取到如下数据：
- “李白深受黄老列庄思想影响，有《李太白集》传世，诗作中多以醉时写的，代表作有《望庐山瀑布》、《行路难》、《蜀道难》、《将进酒》、《梁甫吟》、《早发白帝城》等多首。”（节选自：百度知道）



[搜索答案](#) [我要提问](#)

李白深受黄老列庄思想影响，有《李太白集》传世，诗作中多以醉时写的，代表作有《望庐山瀑布》、《行路难》、《蜀道难》、《将进酒》、《梁甫吟》、《早发白帝城》等多首。

李白的诗想象丰富奇特，风格雄浑奔放，色彩绚丽，语言清新自然，他被誉为“诗仙”。他是唐代伟大的浪漫主义诗人，创造了古代浪漫主义文学、高峰歌行体和七绝，达到后人难及的高度。

而李白诗歌的语言，有的清新如同口语，有的豪放，不拘声律，近于散文，但都统一在“清水出芙蓉，天然去雕饰”的自然美之中。这和他自觉地追求自然美有关。他继承陈子昂的文学主张，以恢复



从网页抓取的自然语言数据

- 从上述数据中，首先可以抽取出如下事物：
 - 李白、黄老列庄思想，《李太白集》、代表作、《望庐山瀑布》等
- 接下来，可以从上述数据中抽取出事物之间的关系，如：
 - （黄老列庄思想，影响，李白）
 - （李白，代表作，《蜀道难》）
- 通过上述的操作，就完成了由非结构化知识数据到结构化知识数据的转变
- 通过抽取出的结构化的知识数据，便可以创建知识图谱



知识图谱的应用：搜索引擎

- “The world is not made of strings, but made of things”

——谷歌知识图谱负责人：Amit Singhal

- 知识图谱的概念最初由谷歌于2012年提出，将知识图谱应用于搜索引擎，目的是增强信息检索能力，为用户提供更加智能的检索结果

The image shows a knowledge graph profile for the Chinese poet Li Bai (李白). At the top, there is a grid of images including a large portrait of Li Bai and several smaller traditional Chinese paintings. Below the images, the name "Li Bai (李白)" is displayed, followed by the title "Chinese poet". A brief biographical text states: "Li Bai, also known as Li Bo, courtesy name Taibai, was a Chinese poet acclaimed from his own day to the present as a genius and a romantic figure who took traditional poetic forms to new heights." This is followed by a link to the Wikipedia page. Below the biography, key facts are listed: "Born: May 19, 701 AD, Suyab", "Died: November 30, 762 AD, Xuan Zhou", "Spouse: Zong (m. 744 AD–762 AD)", "Nicknames: Tai Bai, Poet Knight-Errant, banished immortal, MORE", and "Children: Li Poli, Li Pingyang, Li Boqin". A section titled "Books" displays five book covers: "Quiet Night Thought", "Bright Moon, White Clouds", "Selected 300 Poems of Li Bai", "I Didn't Notice the Mountain...", and "300 Tang Poems". A link "View 5+ more" is provided. At the bottom, a section titled "People also search for" shows several small portrait images of other figures, with a link "View 15+ more".

搜索引擎



- magi.com

- Magi是由Peak Labs研发的基于机器学习的信息抽取和检索系统
- 它能将自然语言文本中的知识提取成结构化的数据，为人类用户和其他人工智能提供可解析、可检索、可溯源的知识体系

Mag[i] 知识图谱

知识图谱 99

描述

“基于图的数据结构” “语义网络” “语义网络的知识库” “各种大规模的知识库”

“显示知识发展进程与结构关系的一系列各种不同的图形” “关系网络”

属性 仅显示部分热门信息，更多请点击或直接搜索问题

提出	2012年 ...	组成	边 ...
应用	智能搜索 ...	主要功能	识别实体之间缺失的链接 ...
价值	理解数据的内在含义 ...	作用	定义关系 ...
包括	客户数据	通过把过去的知识积累进行整合 ...

主要学习来源

知识图谱在互联网金融行业的应用_平台事件_互金知识_网贷之家
www.wdzj.com · 2018年5月3日

转型AI产品经理需要掌握的硬知识二：AI常见概念和算法梳理 - 简书
www.jianshu.com · 2017年12月20日

AI产品经理必修课：知识图谱的入门与应用_实体
www.sohu.com · 2019年9月4日

你不得不看的六篇知识图谱落地好文
www.infoq.cn · 2017年11月20日

智能SOA：勾画企业“演进图谱”_CNKI学问
xuewen.cnki.net · 2007年12月12日

保险业“大阅兵”：你所不知的保险科技生力军战力指数|保险_新浪财...
finance.sina.com.cn · 2019年9月30日

知识图谱 | 机器之心
www.ilqzhixin.com

知识图谱的应用：问答系统

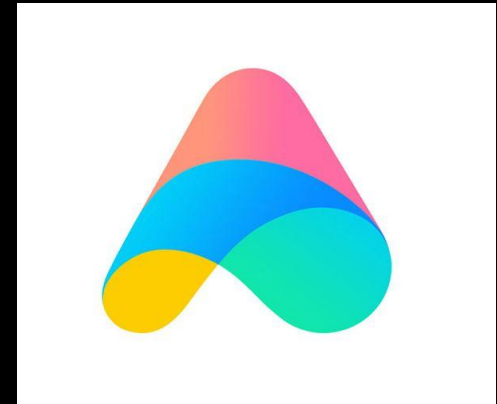
- 问答系统是指让计算机自动回答用户所提出的问题，是信息服务的一种高级形式
- 不同于现有的搜索引擎，问答系统返回用户的不再是若干相关文档，而是精准的、单一的语言形式的答案



苹果公司：Siri



微软：Cortana



小米：小爱同学



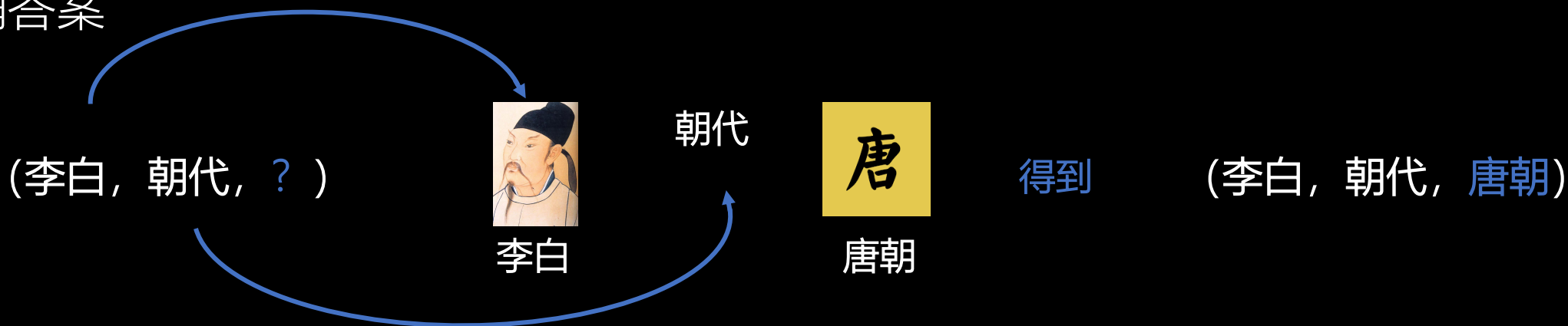
问答系统

- 基于知识图谱的问答主要是通过对自然问句的解析，再从知识图谱中寻找答案的过程
 - 提问分析：将用户提问中的语义、意图提取出来

李白是哪个朝代的人？

(李白, 朝代, ?)

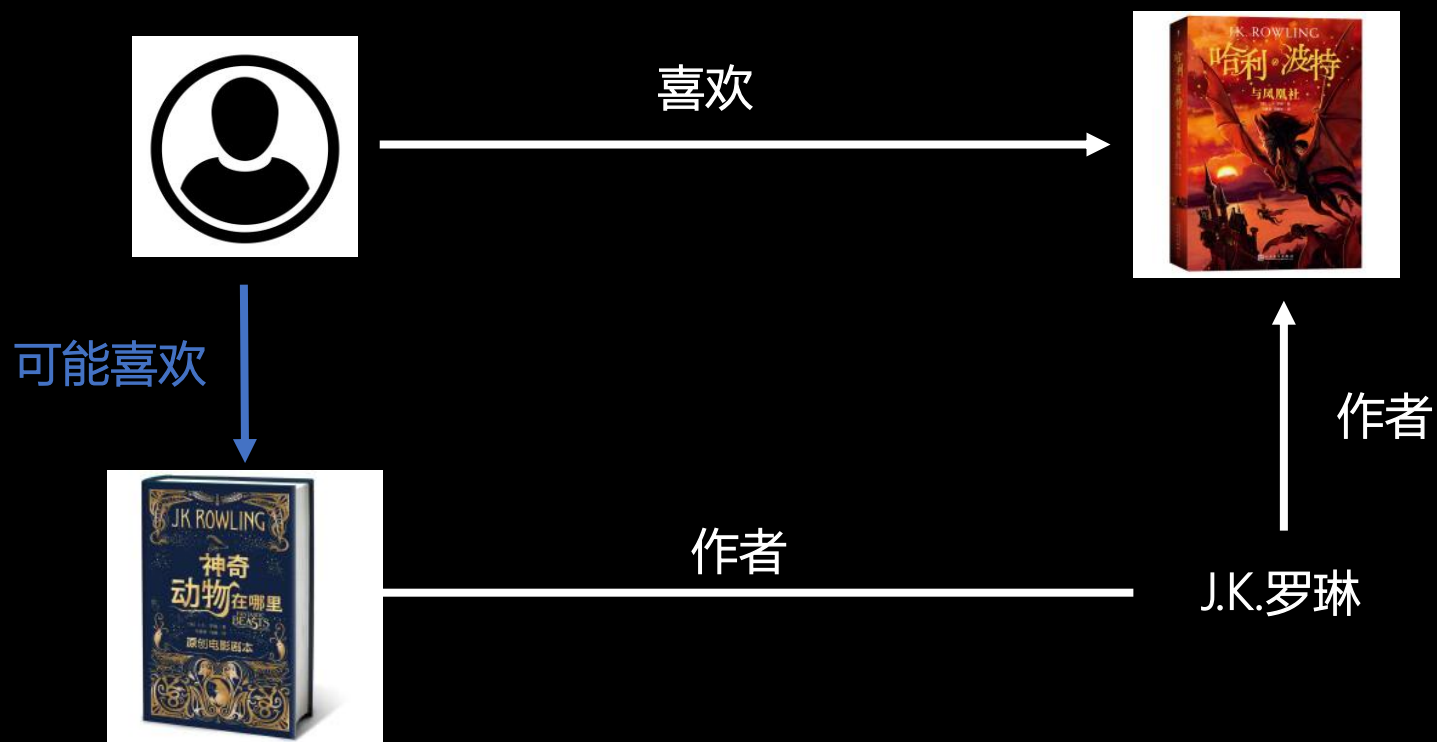
- 答案推理：将提取出的问题与知识图谱中的知识进行匹配和推理，获取正确答案





知识图谱的应用：推荐系统

- 通过将知识图谱引入推荐系统，可以让推荐结果更加精确、多元，并且解释性更强
- 知识图谱在物品之间引入了多种联系，可以深层次发现用户的兴趣

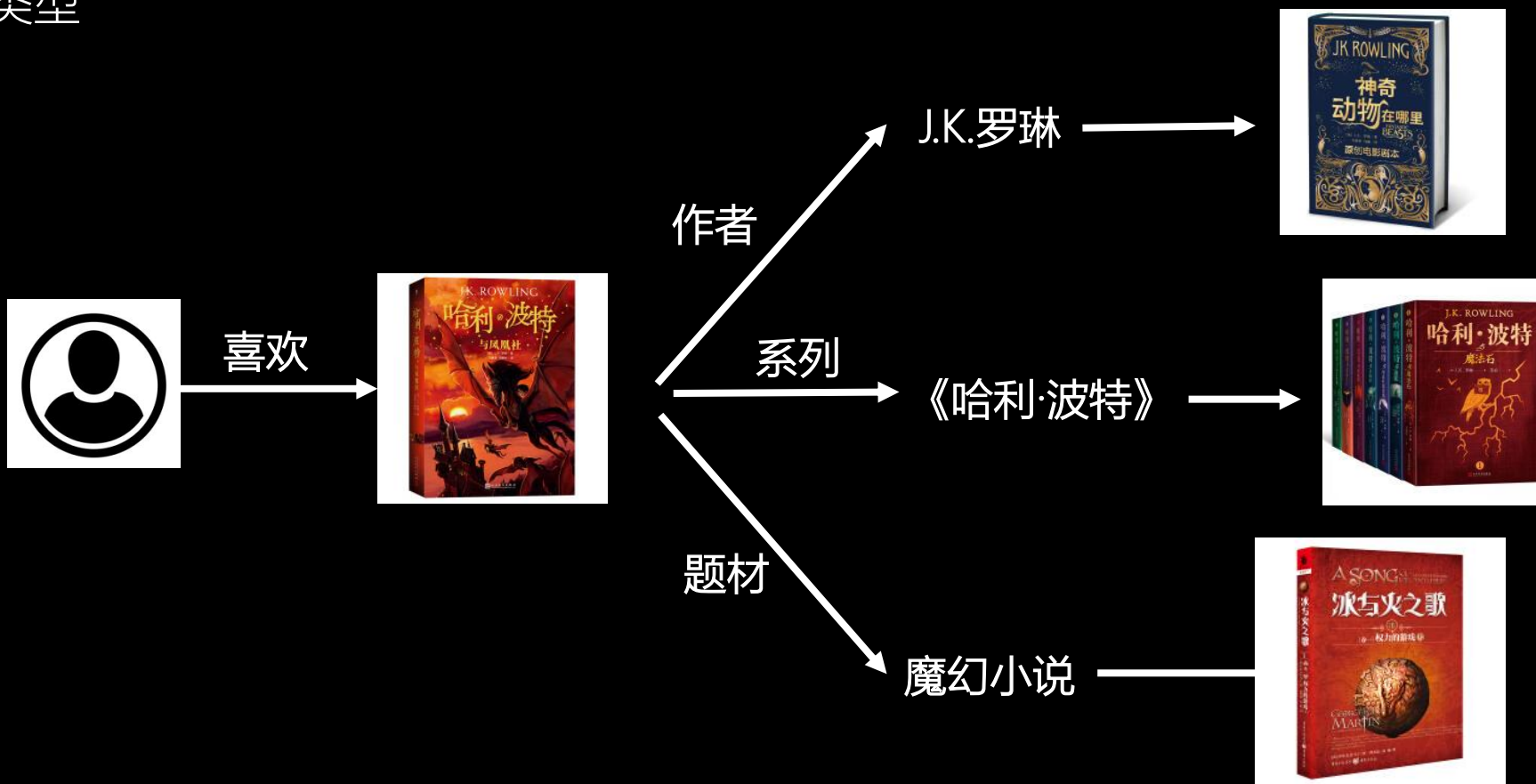


推荐系统



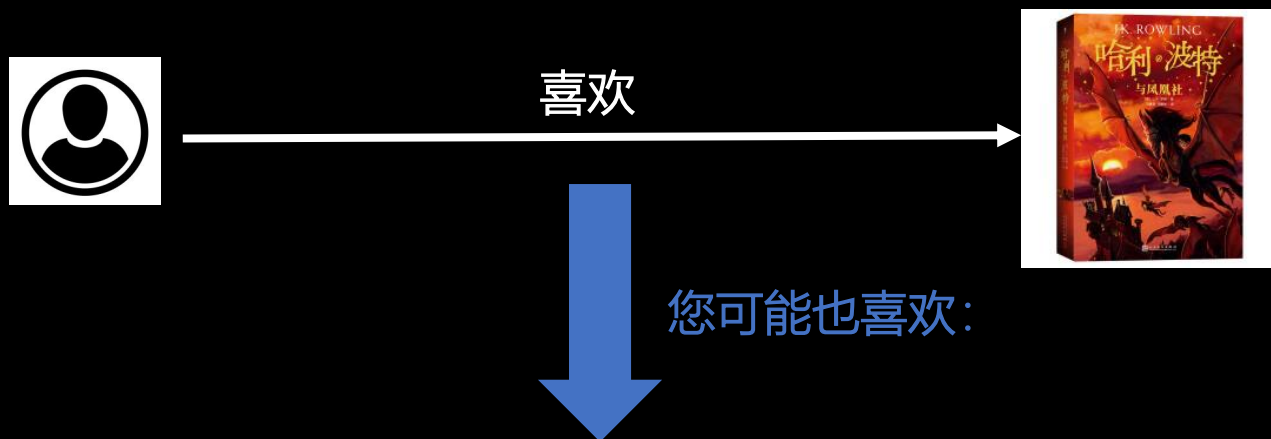
- 多元推荐:

- 知识图谱在物品之间建立了多元的联系, 有利于推荐结果的发散, 避免推荐结果局限于单一的类型



- 可解释性:

- 知识图谱在用户的历史数据与推荐结果之间建立了联系, 增加了用户对推荐结果的满意度



《神奇动物在哪里》, 因为它们有相同的作者
《哈利·波特与魔法石》, 因为它们都属于《哈利·波特》故事系列
《冰与火之歌》, 因为它们有相同的题材