



人工智能原理与算法

第10章-知识表示

中科院自动化研究所 朱翔昱

知识表示

- 如何表示真实世界的各种事实
- 要把什么样的内容放进智能体的知识库中
 - 本体论工程
 - 类别和对象
 - 事件
 - 精神对象和模态逻辑
 - 类别的推理系统

知识表示的重要性

- 人工智能研究中最基本的问题之一
 - 在知识处理中总要问到：“如何表示知识？”，“知识是用什么来表示的？”
 - 怎样使机器能懂，能对之进行处理，并能以一种人类能理解的方式将处理结果告诉人们
 - 在AI系统中，给出一个清晰简洁的描述是很困难的，严格地说，AI对知识表示的认真、系统的研究才刚刚开始。
 - AI4Science中，知识的表示非常重要，Physical Informed Neural Network (PINN)

知识表示

■ 知识的种类

- 事实性知识：采用直接表示的形式，如：凡是猴子都有尾巴
- 过程性知识：描述做某件事的过程，如：面包烘焙
- 行为性知识：指导个体在特定情景下的行为，如：在图书馆保持安静
- 实例性知识：只给出一些具有某种共性的具体案例，知识藏在案例中，如：苹果、梨、杏 → 水果
- 类比性知识：即不给出外延，也不给出内涵，只给出它与其它事物的某些相似之处，如：比喻、谜语
- 元知识：有关知识的知识，如数据库的格式定义（json等），如何从知识库中找到想要的知识

知识表示

■ 知识表示的内容

□ 表示观的研究

认识论：能知道什么、如何知道、知道的界限 ...

本体论：存在的本质，包括存在的种类、结构和基本原理 ...

知识工程：知识获取、表示、存储、推理、应用 ...

□ 表示方法的研究

直接法、代替法（局部、分布，...）

- 本体论工程
- 类别和对象
- 事件
- 精神对象和模态逻辑
- 类别的推理系统

10.1 本体论工程

- 本体论（哲学）：世界上一切事物的本质属性和存在方式，包括物质、精神、意识、关系等各个方面。（唯物主义、唯心主义）
- 本体论（信息科学）：某领域中的概念及相互关系的**形式化表达**
- 如何创建这些表示，主要着重于一些不同领域都会出现的通用概念——例如**物理对象、事件、时间、信念**。表示这些抽象概念被称为**本体论工程**

10.1 本体论工程

■ 上位本体论

- 我们并不会对所有事物都写出完整描述，但是将会留下一些占位符，以使任何领域的知识都能填入

例如，定义对象，他包含一些可以被填入占位符，使得以后用户能够根据需要填入的各种类型对象的细节——诸如机器人、电视机、书本类似于面向对象编程框架的设计者定义**通用的概念（水果）**，期望用户使用这些概念定义**更具体的概念（苹果）**。

- 概念的通用框架被称为**上位本体论**，按照画图惯例：一般概念在上面，而更具体的概念在下面

10.1 本体论工程

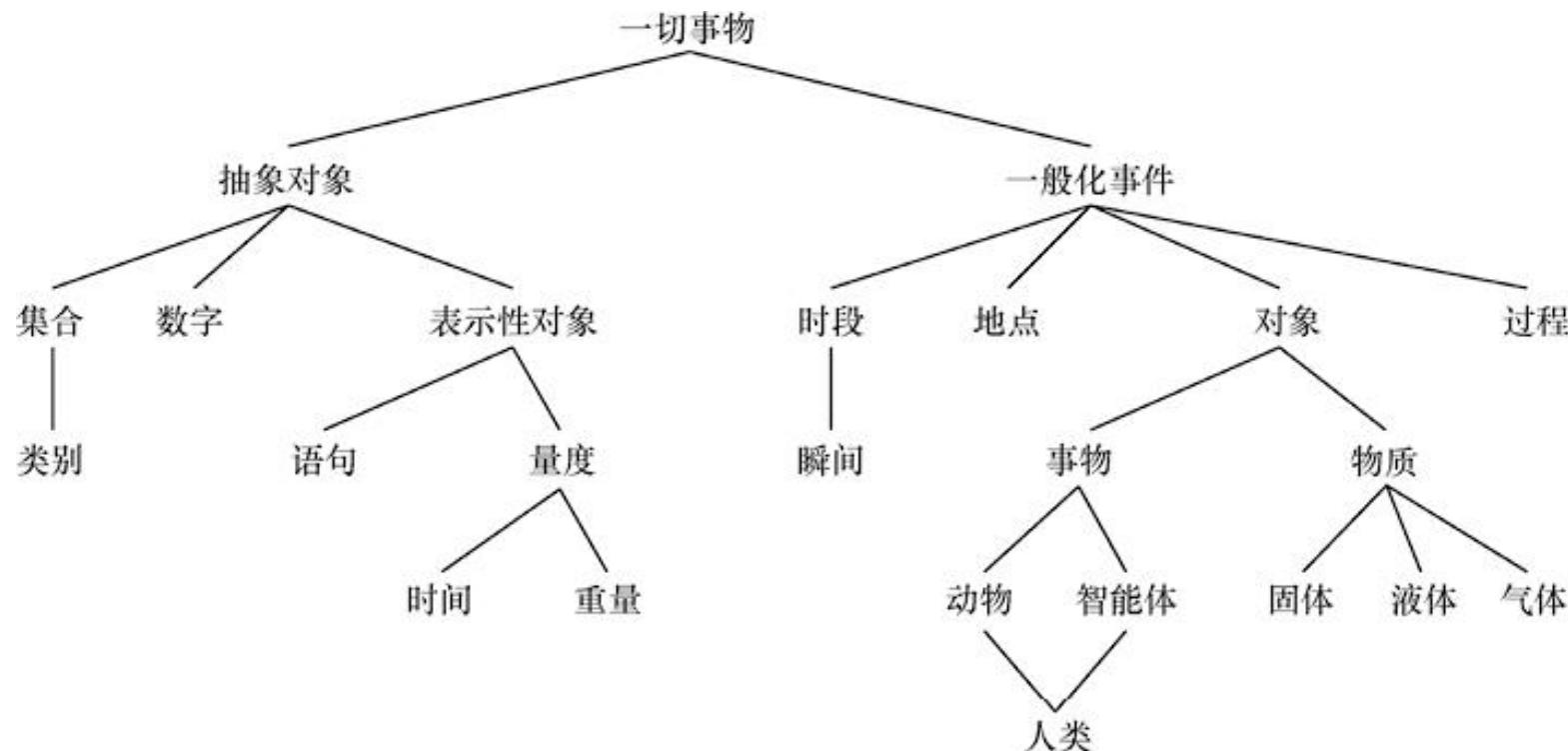


图. 世界的上层本体论。每条线表示低层概念是高层概念的一种具体化。
具体化不一定是排他的——人类既是动物，又是智能体。

10.1 本体论工程

■ 通用本体论与专用本体论

- 专用：某个领域下根据需要构建的本体论。
- 通用：在所有专用领域都应该能或多或少地适用（在增加论域特定的公理后）。这意味着它不能无视任何表示问题
- 在所有足够复杂的论域中，**不同领域的知识必须是统一的**，因为推理和问题求解可能会同时涉及数个领域。
- 对于任何专用的本体论，做些变化使其更一般化是可能的
- 一个很明显的问题出现了：所有这些本体论能收敛到一种通用的本体论吗？经过数个世纪哲学和计算的研究，答案是“有可能”
- 这一节将给出一个通用本体论，综合了这些世纪来的各种思想。

10.1 本体论工程

- “每一种本体论都是一群有共同动机，需要相互分享的人之间的合约——社会共识”
- 当竞争方面的考虑超过共享时，就不会有共同的本体论
- 利益相关者的数量越少，越容易构建本体论
- 构建通用本体论要难于构建用途有限的本体论

10.1 本体论工程

- 已有本体论根据4条路径构建
 - 通过训练有素的本体论学家或逻辑学家团队来构建本体论并写出公理
 - 通过从现有数据库中引入类别、属性和值。 (维基百科)
 - 通过分析文本文件、从中提取信息。 (通过阅读大量网页语料库自主构建，知识图谱、数据可视化)
 - 通过诱导无技能的业余人士输入常识知识。

- 本体论工程
- 类别和对象
- 事件
- 精神对象和模态逻辑
- 类别的推理系统

10.2 类别和对象

- **对象**: 现实世界中某个具体的事物。
- **类别**: 一类对象的抽象,是对对象共同特征的描述。
- 把对象组织成类别是知识表示中很重要的一个部分。虽然与这个世界之间的交互发生在个体对象层次上,但是许多推理是发生在类别层次上的
 - 例如,一位顾客可能会有买个篮球的打算,而不是买个像BB9那样特定的篮球

10.2 类别和对象

- 应用：一旦确定了对象的分类，就可以使用类别帮助我们理解对象。我们从感知输入推断某种对象的存在，从感知到的对象属性推断其类别归属，然后用这些类别信息对这些对象做出预测
 - 例如，从绿色和黄色的斑驳外皮、一脚长的直径、卵形、红色肉汁、黑色种子、存放在水果长廊，我们可以判断出一个对象是西瓜：由此，我们推断它可以用来做水果沙拉

10.2 类别和对象

- 用一阶逻辑表示类别有两种选择：谓词和对象。
 - 谓词：使用谓词Basketball(b)
 - 物化：将类别物化为一个对象，Basketballs。然后，我们可以用Member(b , Basketballs)（我们将它缩写为 $b \in \text{Basketballs}$ ）来说明 b 是篮球类别的一个成员。我们用Subset(Basketballs, Balls)（缩写为 $\text{Basketballs} \subset \text{Balls}$ ）来说明Basketballs是Balls的一个子类。后面将交替使用术语子类和子集。
- 类别通过继承组织知识。
 - 如果食物类别下的所有实例都可以食用，且我们断言水果是食物的子类、苹果是水果的子类，则我们可以推得所有苹果都可以食用。

10.2 类别和对象

■ 分类系统与分类层次

- 类别通过继承组织知识。
- 子类关系将类别组织成分类系统或分类层次，分类系统已经在技术领域明确地使用了数个世纪，最大的这种分类系统将大约1000万个现存和已灭绝地物种组织成单个层次系统（界、门、纲、目、科、属、种）
- 图书馆管理学发展出了所有知识领域的一个分类系统，用杜威十进制系统（Dewey Decimal system）进行编码。
(000 - 总类； 100 - 哲学与心理学； 200 - 宗教； 300 - 社会科学； 400 - 语言； 500 - 自然科学； 600 - 应用科学；
700 - 艺术； 800 - 文学； 900 - 历史与地理)

10.2 分类

分类



门类代码及名称	一级学科代码及名称	二级学科代码	二级学科名称
01 哲学	0101 哲学	010101	马克思主义哲学
		010102	中国哲学
		010103	外国哲学
		010104	逻辑学
		010105	伦理学
		010106	美学
		010107	宗教学
		010108	科学技术哲学
02 经济学	0201 理论经济学	020101	政治经济学
		020102	经济思想史
		020103	经济史
		020104	西方经济学
		020105	世界经济
		020106	人口、资源与环境经济学
	0202 应用经济学	020201	国民经济学
		020202	区域经济学
		020203	财政学
		020204	金融学
		020205	产业经济学
		020206	国际贸易学
		020207	劳动经济学
		020208	统计学
		020209	数量经济学
		020210	国防经济
		020200	社会理论

规模分

0

10.2 类别中的一阶逻辑

- 一阶逻辑通过在对象和类别之间建立联系或者在类别的成员上量化，帮助我们陈述关于类别的事实。
- 一个对象是一个类别的**成员**

$$BB_9 \in Basketballs$$

子类和成员关系是类别中最重要的关系

- 一个类别是另一个类别的**子类**

$$Basketballs \subset Balls$$

- 一个类别中的所有成员拥有某些属性

$$(x \in Basketballs) \Rightarrow Spherical(x)$$

- 一个类别的成员可以通过某些属性来识别

$$Orange(x) \wedge Round(x) \wedge Diameter(x) = 9.5'' \wedge x \in Balls \Rightarrow x \in Basketballs$$

- 一个类别作为整体拥有某些属性

$$Dogs \in \text{驯化物种}$$

类别的类别

10.2 类别和对象

■ 除子类/类别关系外，其他关系：

- 如果我们只说公猫和母猫是猫的子类，那么并没说一个公猫不能是母猫。
- **不相交**：两个或者以上类别，如果它们没有公共的成员，则称它们是不相交(disjoint)。
- 即使我们知道公猫和母猫是不相交的，我们还是不知道一个并非母猫的猫一定是公猫，除非我们进一步说明公猫和母猫构成了一个猫的完全分解(exhaustive decomposition)
- **完全分解**：两个以上子类的成员包含了所有父类成员
- **划分**：一个不相交的完全分解被称为划分(partition)

10.2 类别和对象

- 用例子说明这三个概念

- *Disjoint*({动物, 蔬菜})
 - *ExhaustiveDecomposition*({美国人, 加拿大人, 墨西哥人}, 北美人)
 - *Partition*({公猫,母猫},猫)

注意：NorthAmerican(北美人)的ExhaustiveDecomposition (完全分解)并不是一个Partition(划分)，因为有些人具有双重国籍。

10.2 类别和对象

- 这三个谓词按如下定义：

- $Disjoint(s) \Leftrightarrow (\forall c_1, c_2 \ c_1 \in s \wedge c_2 \in s \wedge c_1 \neq c_2 \Rightarrow \text{Intersection}(c_1, c_2) = \{\})$
- c_1, c_2 是集合 s 中的不同类，他们不共享成员
- $ExhaustiveDecomposition(s, c) \Leftrightarrow (\forall i \ i \in c \Leftrightarrow \exists c_2 \ c_2 \in s \wedge i \in c_2)$
- 对 c 中的成员，一定属于集合 s 中的一类
- $\text{Partition}(s, c) \Leftrightarrow Disjoint(s) \wedge ExhaustiveDecomposition(s, c)$

- 类别也可以通过提供成员的充分必要条件来定义。如，
单身汉是未婚成年男性：

- $x \in Bachelors \Leftrightarrow \text{Unmarried}(x) \wedge x \in Adults \wedge x \in Males$

10.2 物理组成

- 某个对象可以是另一个对象的一部分 (Part of)
 - 鼻子是头的一个部分
 - 罗马尼亚是欧洲的一个部分
 - 这一章是本书的一个部分
- 对象与对象之间的关系：使用一般的partOf关系表述一个物体是另一个物体的部分。对象可以组织成PartOf(部分)层次结构
 - *PartOf(Bucharest, Romania)*, 布加勒斯特是罗马尼亚的一部分
 - *PartOf(Romania, EasternEurope)*, 罗马尼亚是东欧的一部分
 - *PartOf(EasternEurope, Europe)*, 东欧是欧洲的一部分
 - *PartOf(Europe, Earth)*, 欧洲是地球的一部分

10.2 物理组成

- PartOf 关系是传递的和自反的；即
 - $\text{PartOf}(x, y) \wedge \text{PartOf}(y, z) \Rightarrow \text{PartOf}(x, z)$
 - $\text{PartOf}(x, x)$
- 可以得出结论 $\text{PartOf}(\text{Bucharest}, \text{Earth})$

10.2 物理组成

- 复合对象 (composite objects) 的类别经常是通过各部分之间的结构关系刻画的。例如，一个两足动物身体上有两条腿。
- “恰好” 有两条腿，这个关系描述起来比较复杂；

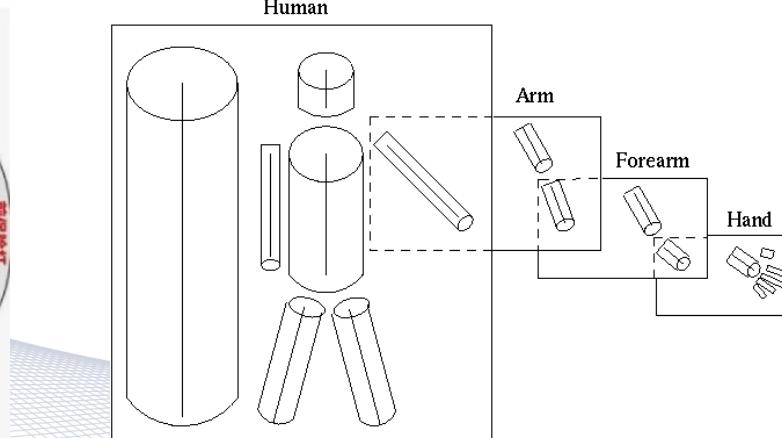
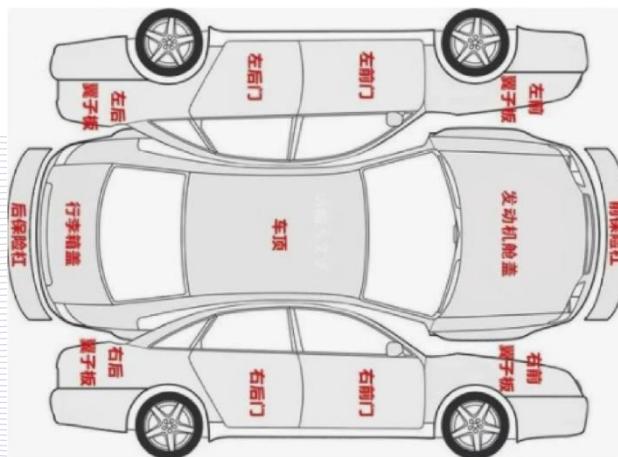
$$\begin{aligned}
 Biped(a) \Rightarrow & \exists l_1, l_2, b \ Leg(l_1) \wedge Leg(l_2) \wedge Body(b) \\
 & \wedge PartOf(l_1, a) \wedge PartOf(l_2, a) \wedge PartOf(b, a) \\
 & \wedge Attached(l_1, b) \wedge Attached(l_2, b) \\
 & \wedge l_1 \neq l_2 \wedge [\forall l_3 \ Leg(l_3) \wedge PartOf(l_3, a) \Rightarrow (l_3 = l_1 \vee l_3 = l_2)]
 \end{aligned}$$

“恰好”

10.2 物理组成

■ PartPartition关系

- 一个对象由它的PartPartition中的各个部分组成，那么从这些部分也得到了某些性质
- 例如，一个复合对象的质量是各个部分的质量的总和。注意，对类别来说不是这种情况，因为类别没有质量
- Part涉及到计算机视觉、心理学、神经学中的层次化表示(hierarchical representation) 的构建。



10.2 物理组成

■ PartPartition关系

- 束 (bunch) 定义具有确定部分但没有特定结构的复合对象也是有用的
- 例如， “袋子里的苹果重两磅。” 我们会倾向于认为这个重量属性归于袋子中苹果的集合，这会产生错误，**因为集合是一个抽象的数学概念，它只有元素却没有重量**。相反，我们需要一个新的概念，我们称之为**束 (bunch)**。例如，如果有苹果Apple1, Apple2, Apple3，那么 BunchOf ({Apple1, Apple2, Apple3})
- 表示了由三个苹果作为部分(**不是元素**)组成的复合对象，可以把束当作一个平常的 (尽管没有结构的) **对象来使用**

10.2 物理组成

- BunchOf(Apples)不同于Apples。前者是由所有苹果组成一个复合对象，后者是所有苹果组成的类别或者集合
- 使用PartOf定义： s 中的每个元素是BunchOf(s)的部分：
 - $\forall x \ x \in s \Rightarrow PartOf(x, BunchOf(s))$

10.2 量度

- 关于世界的科学理论和常识理论中，物体有高度、质量、成本这样的性质，赋予这些属性的值称为量度。
- 比如一条线段的长度，可以称这个长度为3.81cm，**用单位函数和一个数字作为参数来表示长度**，如果线段为L1，那么

$$\text{Length}(L_1) = \text{Inches}(1.5) = \text{Centimeters}(3.81)$$

L1的长度为1.5英寸或者3.81厘米

- 单位之间的转换用某个单位的倍数等于另一个单位来完成：
 - $\text{Centimeters}(2.54 \times d) = \text{Inches}(d)$
 - 可以为磅和千克，秒和天，元和分写出类似的公理。
 - **量度就被对象化了**

10.2 量度

■ 量度可以像下面那样用来描述对象：

- $Diameter(Basketball_12)=Inches(9.5)$
- $ListPrice(Basketball_12)=(19)
- $d \in Days \Rightarrow Duration(d)=Hours(24)$

10.2 量度

- 有些量度没有公认的尺度
 - 习题具有难度，甜点具有美味程度，诗歌有优美程度，而我们无法为这些量赋以数值。
- 对量度来说，最重要的不是其特定的数值，而是**它是可以被排序的**。
 - 尽管量度不是数字，但我们还是可以用诸如“>”之类的定序符号来比较它们。

$$Difficulty(e_1) > Difficulty(e_2)$$

10.2 对象：物体和物质

■ 物质(stuff)和物体 (thing) :

- 现实世界可以视为由**基元对象**（例如原子粒子）和由其构成的**复合对象**组成的。通过在诸如苹果和汽车这类大对象的层次上进行推理，可以克服处理大量基元对象所带来的复杂度。然而，现实中有相当一部分对象似乎不服从明显的**个体化**——它们不能划分成截然分开的对象，给这部分对象一个通用的名称：**物质**。
- 例如，假设我们有一些黄油和一只土豚。可以说有一只土豚，但是说不出“黄油对象”的明确数量，因为一个黄油对象的任何一部分仍是一个黄油对象。这是**物质**和**物体**的最大区别
- 类似英语中的可数名词和不可数名词的区别

10.2 对象：物体和物质

■ 关于物质的表示：

- 一“块”黄油可以作为一个对象，称它为 $Butter_3$ ，还定义类别 $Butter$
- $Butter$ 类的所有元素将是所有那些我们可以说“它是黄油”的东西，包括 $Butter_3$
- 任何一个黄油对象的部分也是一个黄油对象：
$$x \in Butter \wedge PartOf(y, x) \Rightarrow y \in Butter$$
- 黄油在30摄氏度时熔化：
$$x \in Butter \Rightarrow MeltingPoint(x, Centigrade(30))$$

10.2 对象：物体和物质

■ 固有与非固有属性

- 有一些属性是**固有的**，属于对象的组成物质，当你将物质切成两半的时候，那两半保留了同样的固有属性集——如密度、沸点、口味、颜色、所有权等
- **非固有属性**则不是这样：诸如重量、长度、形状、功能之类的属性在划分时不能保持不变
- 在定义中**只包括了固有属性**的对象类就是物质（不可数名词）；
- 在定义中**包含了任何非固有属性**的类就是物体（可数名词）

- 本体论工程
- 类别和对象
- 事件
- 精神对象和模态逻辑
- 类别的推理系统

10.3 事件

- 事件主要处理跟**时间相关的知识**
- 前面章节中，命题逻辑使用流和后继状态公理来描述发生的事情。**复习：流是世界中随时间变化的部分。**
 - 适用性是有限的：动作是离散的、瞬间的、一次只能发生一个动作
 - 例如给浴盆装水。后继状态公理能够说出，在这个动作之前浴盆是空的，动作完成之后浴盆是满的，但它不能说出在动作期间发生了什么例如，等浴盆装水的时候同时刷牙。
 - 为了处理这种情况，我们引入称为**事件演算**的另一种形式体系，它是基于时间点而不基于情景的

10.3 事件

- 事件演算的对象是事件、流和时间点
 - 流At(Shankar, Berkeley)是一个对象，指Shankar在Berkeley这个事实，但这个事实是否成立，它本身并没有说出任何信息，他的出现通常伴随着时间
 - 为了声称一个流在某些时间点成立，我们使用像 $T(At(Shankar, Berkeley), t)$ 中一样的谓词 T
 - 事件是事件类别的实例。Shankar从SanFrancisco飞到Washington D.C.的事件 E_1 描述为

$E_1 \in Flyings \wedge Flyer(E_1, Shankar) \wedge Origin(E_1, SF) \wedge Destination(E_1, DC)$

Flying是一个事件类， E_1 是一个事件

10.3 事件

- 使用Happens(E_1, i)来表示事件 E_1 发生在时间区间*i*，用Extent(E_1)=*i*的函数形式也表示同样的意思。
- 用时间的(start, end)对表示时间区间；也就是说， $i=(t_1, t_2)$ 是开始于 t_1 结束于 t_2 的时间区间。
- 一个事件演算版本的完整的谓词集为

$T(f, t_1, t_2)$	流 f 在 t_1 和 t_2 之间的所有时刻为真
$Happens(e, t_1, t_2)$	事件 e 从 t_1 开始，于 t_2 结束
$Initiates(e, f, t)$	事件 e 导致流 f 在时刻 t 为真
$Terminates(e, f, t)$	事件 e 导致流 f 在时刻 t 不再为真
$Initiated(f, t_1, t_2)$	流 f 在 t_1 和 t_2 之间的某时刻开始为真
$Terminated(f, t_1, t_2)$	流 f 在 t_1 和 t_2 之间的某时刻停止为真
$t_1 < t_2$	时刻 t_1 出现在时刻 t_2 之前

10.3 事件

- 可以将飞行事件的效果描述为
 - $E = Flyings(a, here, there) \wedge Happens(E, t_1, t_2) \Rightarrow Terminates(E, At(a, here), t_1) \wedge Initiates(E, At(a, there), t_2)$

事件E使得[a在here]这个流在t1时刻不再为真，
且使得[a在there]这个流在t2时刻开始为真

10.3 事件

- 假设一个事件在时刻 t_1 和 t_3 之间发生，而在该时段内的某时刻 t_2 ，该事件通过启动流（使其为真）或终止流（使其为假）改变了流 f 的值。则在**未来的某个时刻 t_4** ，**如果没有其他事件改变过流**（不论是启动流还是终止流），**流的值保持不变。**
- 公理可以表示成
 - $Happens(e, t_1, t_3) \wedge Initiates(e, f, t_2) \wedge \neg Terminated(f, t_2, t_4) \wedge t_1 \leq t_2 \leq t_3 \leq t_4 \Rightarrow T(f, t_2, t_4)$
 - $Happens(e, t_1, t_3) \wedge Terminates(e, f, t_2) \wedge \neg Initiated(f, t_2, t_4) \wedge t_1 \leq t_2 \leq t_3 \leq t_4 \Rightarrow \neg T(f, t_2, t_4)$

10.3 事件

- *Terminated*和*Initiated*定义为:
- $\text{Terminated}(f, t_1, t_5)$ 流 f 在 t_1 和 t_2 之间某时刻停止为真
 - $\text{Terminated}(f, t_1, t_5) \Leftrightarrow \exists e, t_2, t_3, t_4 \text{ Happens}(e, t_2, t_4) \wedge \text{Terminates}(e, f, t_3) \wedge t_1 \leq t_2 \leq t_3 \leq t_4 \leq t_5$
- $\text{Initiated}(f, t_1, t_5)$ 流 f 在 t_1 和 t_2 之间的某时刻开始为真
 - $\text{Initiated}(f, t_1, t_5) \Leftrightarrow \exists e, t_2, t_3, t_4 \text{ Happens}(e, t_2, t_4) \wedge \text{Initiates}(e, f, t_3) \wedge t_1 \leq t_2 \leq t_3 \leq t_4 \leq t_5$

10.3 时间区间

- 时间的性质
- 时间间隔包含瞬间和延续间隔
 - $\text{Partition}(\{\text{Moments}, \text{ExtendedIntervals}\}, \text{Intervals})$
 - 瞬间的持续时间为0
 - $i \in \text{Moments} \Leftrightarrow \text{Duration}(i) = \text{Seconds}(0)$

10.3 时间区间

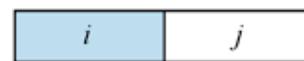
- 选择用秒来设定时间尺度，制定格林尼治标准时1900年1月1日午夜瞬间为时间0
- 函数 *Begin*和*End* 输出一个间隔的最早瞬间和最晚瞬间，函数*Time*则将瞬间转换为时间尺度上的点。函 *Duration*给出开始时间和结束时间之间的差值
 - $\text{Interval}(i) \Rightarrow \text{Duration}(i) = (\text{Time}(\text{End}(i)) - \text{Time}(\text{Begin}(i)))$
 - $\text{Time}(\text{Begin}(\text{AD}1990)) = \text{Seconds}(0)$
 - $\text{Time}(\text{Begin}(\text{AD}2001)) = \text{Seconds}(3,187,324,800)$
 - $\text{Time}(\text{End}(\text{AD}2001)) = \text{Seconds}(3,218,860,800)$
 - $\text{Duration}(\text{AD}2021) = \text{Seconds}(31,536,000)$

10.3 时间区间

间隔关系的完整集合（时间间隔的谓词）

Meet

$Meet(i, j)$



Before

$Before(i, j)$

$After(j, i)$



After

During

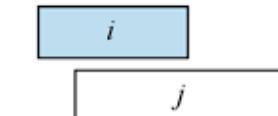
$During(i, j)$



Overlap

Starts

$Overlap(i, j)$



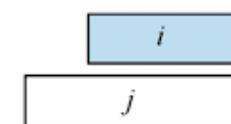
Finishes

Equals

$Starts(i, j)$



$Finishes(i, j)$



$Equals(i, j)$

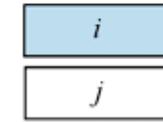


图10-2 时间间隔的谓词

10.3 时间区间

■ 例子

- 英国女王伊丽莎白二世的统治紧接着乔治六世
- $Meets(ReignOf(GeorgeVI), ReignOf(ElizabethII))$
- 猫王埃尔维斯主宰乐坛的时间与20世纪50年代重叠
- $Overlap(Fifties, ReignOf(Elvis))$
- $Begin(Fifties) = Begin(AD1950)$
- $End(Fifties) = End(AD1959)$

- 本体论工程
- 类别和对象
- 事件
- 精神对象和模态逻辑
- 类别的推理系统

10.4 精神对象和模态逻辑

- 除了客观世界，也需要描述主观世界
- 关于自身所具有的知识和推导过程的知识对控制推断非常有用。
 - 甲问道“1764 的平方根是多少”，而乙回答“我不知道”。如果甲坚持道“再想一想”，乙应当意识到通过进一步思考其实可以回答这个问题。
 - 但是，如果甲提出的问题是“总统现在是坐着还是站着？”那么乙应当明白再怎么思考都不太会有作用
- 知道其他智能体所具有的知识也很重要，乙应当意识到总统自己肯定知道答案。

10.4 精神对象和模态逻辑

- 智能体大脑中的**精神对象**和操纵这些精神对象的**精神过程的模型**非常有用。
- 精神对象是指某些人脑子中的知识
 - *Believes* (相信) , *Knows* (知道) , *Wants* (想要) , and *Informs* (通知)

10.4 精神对象和模态逻辑

- Lois知道超人会飞
 - $\text{Knows}(\text{Lois}, \text{CanFly}(\text{Superman}))$
- 如果超人是Clark为真，那么就会得出Lois知道Clark会飞的结论。但是这是错误的，因为Lois不知道Clark是超人
 - $(\text{Superman} = \text{Clark}) \wedge \text{Knows}(\text{Lois}, \text{CanFly}(\text{Superman})) \models \text{Knows}(\text{Lois}, \text{CanFly}(\text{Clark})) \times$
- 等值推理失效

10.4 精神对象和模态逻辑

■ 指代透明性

- $2+2=4$ 且 $4 < 5$ ，希望智能体知道 $2+2 < 5$ 。第一条中的4和第二条中的4是相等的，可以直接推导。

■ 指代不透明性

- 涉及精神的“believe”，“know”，需要指代不透明性，等值推理不能直接应用，因为并非所有的智能体都知道哪些项是指代同一个对象的。例如 $Superman = Clark$ 不能直接替换

10.4 精神对象和模态逻辑

- 模态逻辑：模态逻辑含有以语句（而非项）为参数的模态算子
 - A知道P，记为 $K_A P$ ，其中K是知识的模态算子
 - 使用两个参数，一个是智能体（以下标表示），另一个是语句。
- 模态逻辑公理

$$(K_a P \wedge K_a (P \Rightarrow Q)) \Rightarrow K_a Q$$

A知道P，且A知道P能推出Q，则A知道Q

$$K_a P \Rightarrow K_a (K_a P)$$

A知道自己知道P

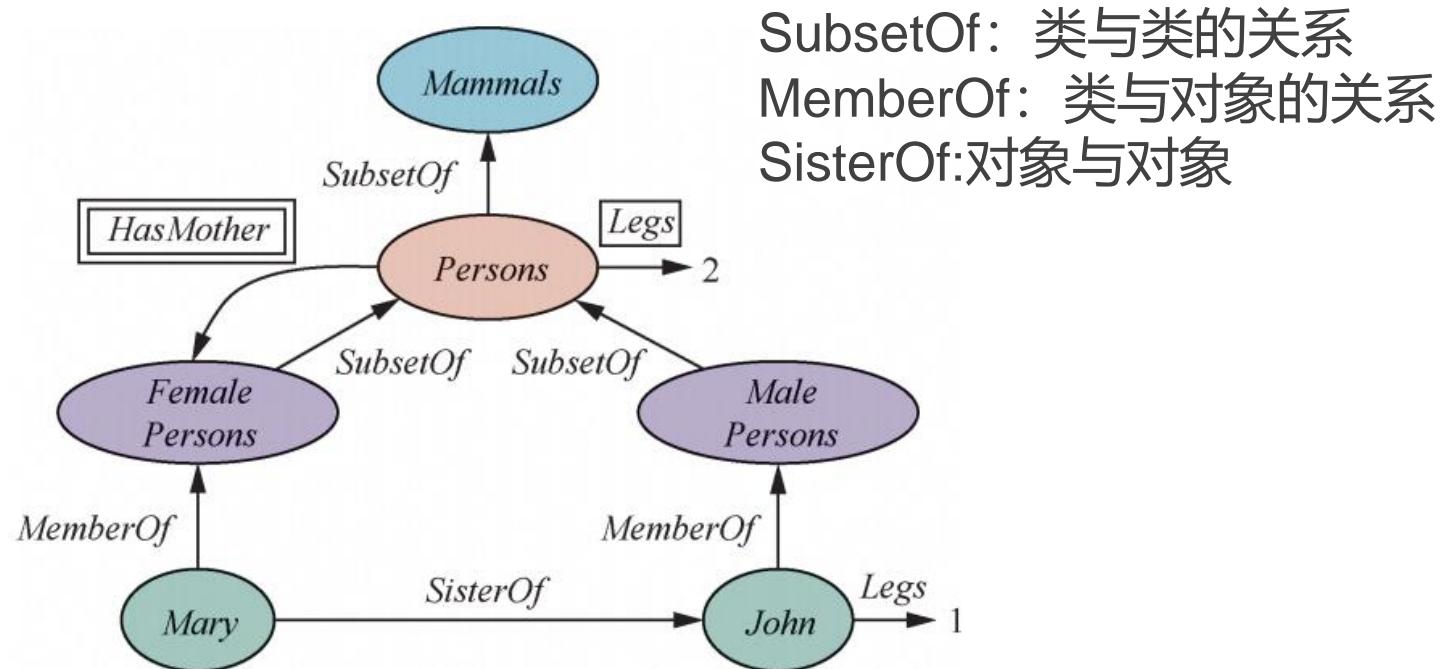
- 本体论工程
- 类别和对象
- 事件
- 精神对象和模态逻辑
- 类别的推理系统

10.5 类别的推理系统

- **类别是大规模知识表示系统最重要的构造模块，面向类别的组织和推理，专门设计了两类系统：**
- **语义网络为知识库可视化提供图形的帮助，并在类别隶属关系基础上推断对象的属性。**
- **描述逻辑为构建和组合类别定义提供形式语言，并为判定类别之间的子集和超集关系提供有效算法**

10.5 语义网络

直接连接不行，
类别没有母亲，
对象才有



SubsetOf: 类与类的关系
 MemberOf: 类与对象的关系
 SisterOf: 对象与对象

图 10-4 具有 4 个对象 (*John*、*Mary*、1 和 2) 和 4 个类别的语义网络。关系使用带标签的连线表示

双线框连接 $\forall x \ x \in Persons \Rightarrow [\forall y HasMother(x, y) \Rightarrow y \in FemalePersons]$

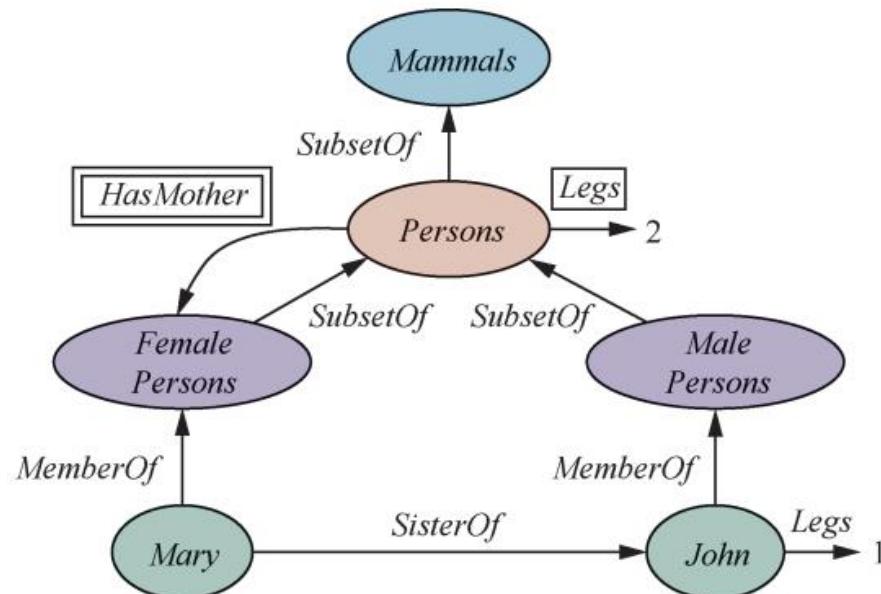
单线框连接：用于断言类别的每个成员的具有的属性

$$\forall x \ x \in Persons \Rightarrow Legs(x, 2)$$

10.5 语义网络

■ 继承

- 要找出Mary有多少条腿，继承算法跟踪从Mary到她所属类别的**连接MemberOf**，接着跟踪通向上一层的**连接SubsetOf**，直到找到一个带有线框连接Legs的类别——这个例子中，就是Persons类别



与逻辑定理证明相比，这种推理机制的简单性和高效性已经成为语义网络的主要吸引力之一

10.5 语义网络

■ 多重继承

- 当一个对象能够属于不止一个类别或一个类别能够是不止一个其他类别的子类时，继承变得复杂了：这称为**多重继承**
- 继承算法可能找到两个和更多个相互冲突的值来回答查询，多重继承在一些面向对象程序设计（OOP）语言中被禁止，比如Java，这些语言在类层次中使用继承。

10.5 语义网络

- 语义网络的连线只能表示二元关系。语句

$Fly(Shankar, \text{纽约}, \text{新德里}, Yesterday)$ 无法直接从语义网络中断言。可以通过将命题自身物化为属于合适的事件类别的事件类，从而达到n元断言的效果

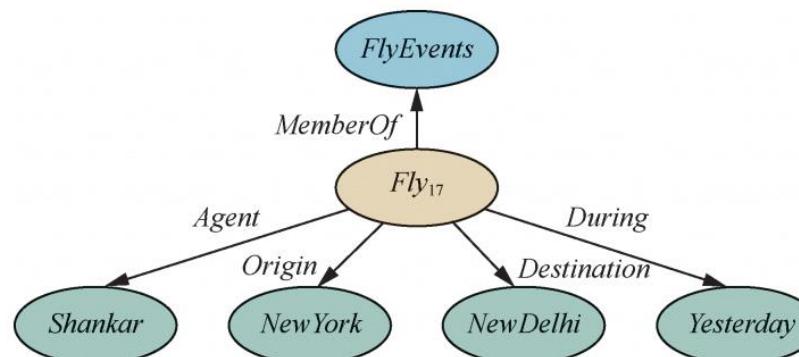


图 10-5 一个语义网络的一部分，展示了对逻辑断言 $Fly(Shankar, New York, New Delhi, Yesterday)$ 的表示

10.5 描述逻辑

- **描述逻辑是被设计成描述类别的定义和属性的形式化语言**
- **描述逻辑的主要推理的是**
 - **包容**: 通过比较定义, 检查一个类别是否是另一个类别的子集
 - **分类**: 检查一个对象是否属于一个类
 - **类别定义一致性**: 成为该类别的条件在逻辑上是否可满足 (某些系统支持)

10.5 描述逻辑

- CLASSIC语言是一种典型的描述逻辑。

- 例如，单身汉是未婚的成年男性
 - CLASSIC语言：

$bachelor = And(Unmarried, Adult, Male)$

- 一阶逻辑：

$Bachelor(x) \Leftrightarrow Unmarried(x) \wedge Adult(x) \wedge Male(x)$

10.5 描述逻辑

■ CLASSIC中子集的描述语法。

```

概念 → Thing | 概念名
| And(概念, ...)
| All(角色名, 概念)
| AtLeast(整数, 角色名)
| AtMost(整数, 角色名)
| Fills(角色名, 个体名, ...)
| SameAs(路径, 路径)
| OneOf(个体名, ...)

路径 → [角色名, ...]

概念名 → 成人 | 女人 | 男人 | ...
角色名 → 配偶 | 女儿 | 儿子 | ...

```

图 10-6 CLASSIC 语言的一个子集中的描述语法

■ 举例：一个男人的集合，这样的男人至少有三个儿子且至多有两个女儿，儿子都失业了并与医生结了婚，女儿都是物理系或数学系的教授，用下述语句表示：

$$\begin{aligned}
& \text{And}(\text{Man}, \text{Atleast}(3, \text{Son}), \text{AtMost}(2, \text{Daughter}), \\
& \text{All}(\text{Son}, \text{And}(\text{Unemployed}, \text{Married}, \text{All}(\text{Spouse}, \text{Doctor}))), \\
& \text{All}(\text{Daughter}, \text{And}(\text{Professor}, \text{Fills}(\text{Department}, \text{Physics}, \text{Math})))
\end{aligned}$$

10.6 用缺省信息推理

- 技巧：常识推理似乎总是“跳到结论”。

□ 例如

轮子

自动

□ 在没

如果

且车

■ 这违
论)

有 4 个

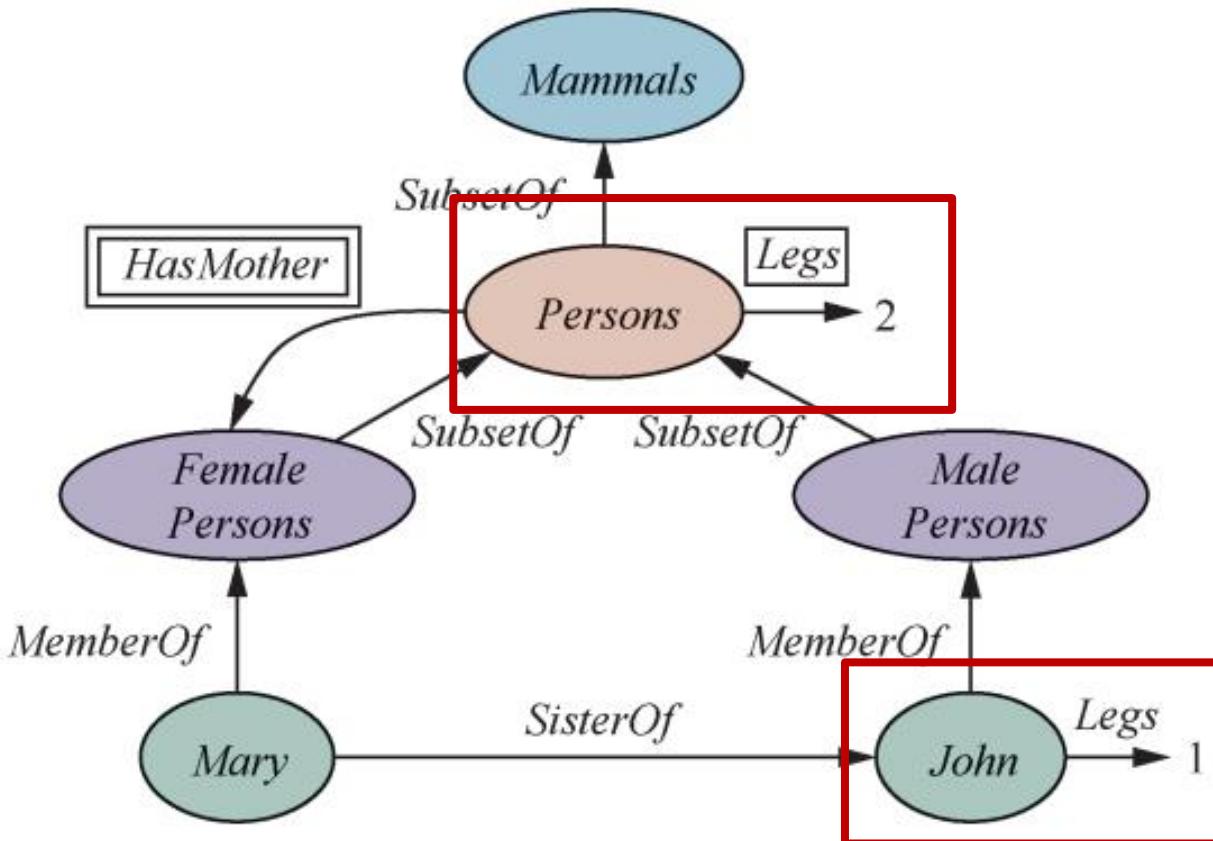
E 新证据

能性。

达成的。

轮胎，而

前的结



10.6 用缺省信息推理

■ 用限定和缺省逻辑

■ 限定逻辑：

- 确定一个“不正常 (Abnormal) ”的谓词，也就是，对所有对象都为假，除了那些我们已知为真的对象。
- 断言缺省规则，鸟会飞：

$$Bird(x) \wedge \neg Abnormal_1(x) \Rightarrow Flies(x)$$

- 如果我们说 $Abnormal(x)$ 被限定了，则推理器就有权假设 $\neg Abnormal(x)$ ，除非已知为真。这就使结论 $Flies(eagle)$ 可以从前提 $Bird(eagle)$ 中得出，但如果被断言 $Abnormal(eagle)$ ，结论便不再成立。

10.6 用缺省信息推理

■ 缺省逻辑：

- $P: J_1, \dots, J_n / C$
- P 称为先决条件， C 是结论， J_i 是论证
- 如果其中任意一个论证可以被证明为假，则无法得出结论
- J_i 和 C 中的所有变量都必须同时在 P 中

$Bird(x) : Flies(x) / Flies(x)$

如果 $Bird(x)$ 为真，且知识库中没有明确说 x 不会飞，即 $\neg Flies(x)$ ，则
可以得出缺省结论 $Flies(x)$

10.6 真值维护系统

■ 信念修正

- 许多知识表示系统得出的推断只是具有缺省状态，而非完全确定
- 在有新信息的情况下，许多推断出的事实最终证明是错误的，必须被收回

■ 真值维护系统 (TMS)

- 假设知识库KB中含有语句P（可能是缺省结论，也可能只是不正确的断言），想执行TELL(KB, $\neg P$)

- 收回由P得出的所有语句

- 除非某语句还有除P之外的其他论证

- 本体论工程
- 类别和对象
- 事件
- 精神对象和模态逻辑
- 类别的推理系统
- 拓展知识：RAG技术

10.7 RAG技术（目前大模型的问题）

- 大模型只能认识“旧知识”
 - 训练知识的局限性：大模型的知识完全来自预训练语料。训练完成后，它无法主动获取新的政策法规、公司制度、科研进展等信息，因此所有“新知识”对它都是未知的。
 - 现实痛点：真实世界的知识每天都在变化，而让大模型通过重新预训练来“重新读书”代价极高，因此大模型难以在动态场景中保持准确性。
 - 为了解决知识更新难的问题，Retrieval-Augmented Generation(RAG)为模型提供了“随时查阅外部资料”的能力，让大模型不再依赖记忆，而是通过检索实时知识来生成更可靠的答案。

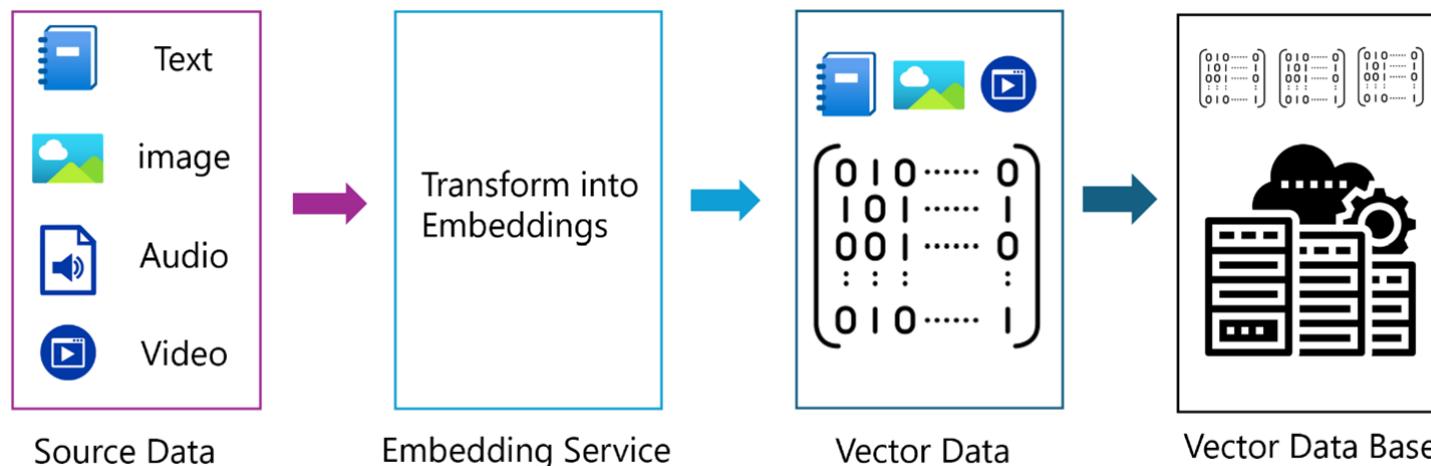
10.7 RAG技术（目前大模型的问题）

- 幻觉：AI的“胡编乱造”
 - 为什么会出现幻觉？当问题超出大模型的训练知识范围时，模型无法基于事实回答。由于生成模型的目标是“保持语言连贯”，它会补全缺失内容、生成看似合理但虚假的信息。
 - 没有RAG时：高风险场景容易出错模型只能依赖自身的“记忆”来推断答案。在医疗、金融、法律需要高可靠性的领域中，这产生严重风险
 - RAG通过“先检索权威资料，再生成答案”的方式，让模型：
 - 以真实信息为依据作答
 - 将回答范围限制在可信文档片段内
 - 显著降低幻觉 / 编造信息的概率

10.7 RAG技术

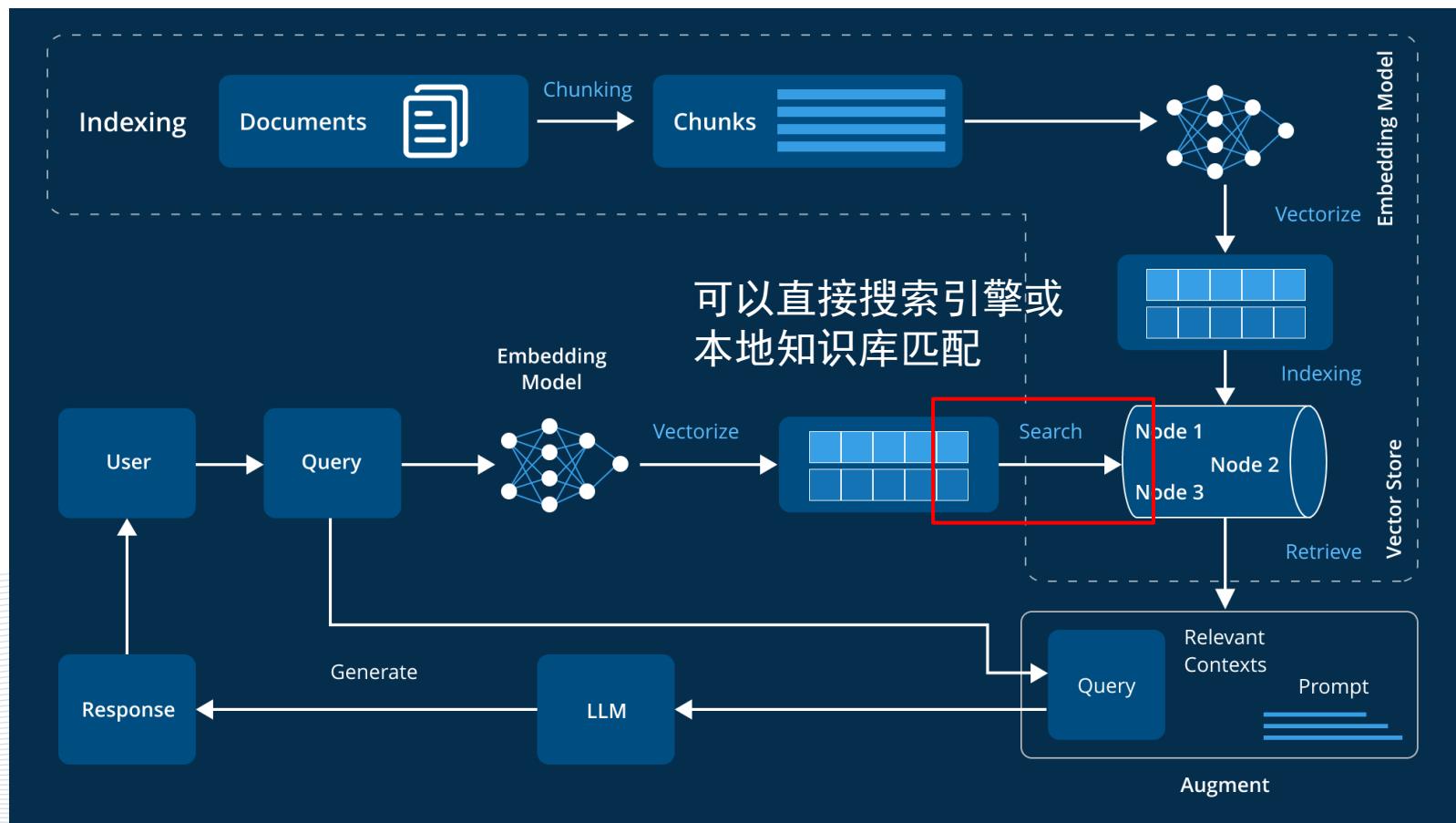
- RAG中向量知识库的示例
 - 每个条目 (entry) = (embedding, raw text chunk, metadata)

字段	内容
向量 e_i	chunk 的高维 embedding (例如 1024 维)
原文 chunk_i	文本内容本身 (用于给 LLM)
metadata	文档标题、来源、位置、时间戳、分段 ID 等



10.7 RAG技术

- RAG的基本流程
- 文本切块 -> 构建向量库 -> 检索相关知识 -> 增强生成



10.7 RAG技术

RAG效果示例

已思考 (用时 11 秒) >

截至2025年12月9日，2025赛季的F1世界冠军尚未决出，因为2025赛季的比赛还未开始。

不过，我们可以根据当前的信息为您提供背景和展望：

1. 2024赛季冠军已决出：刚刚结束的2024赛季，红牛车手马克斯·维斯塔潘成功卫冕，连续第四次获得车手总冠军。

2. 2025赛季前瞻：

- 赛季时间：2025赛季的赛程已经公布，揭幕战将于2025年3月16日在澳大利亚墨尔本举行，收官战将于11月23日在阿布扎比举行。

- 最大看点：刘易斯·汉密尔顿将从梅赛德斯转会至法拉利车队，这将是2025赛季最受关注的焦点，他与查尔斯·勒克莱尔的搭档备受期待。

冠军热门：

- 马克斯·维斯塔潘（红牛车队）依然是最大的夺冠热门，前提是红牛赛车继续保持竞争力。
- 法拉利（拥有汉密尔顿和勒克莱尔）和迈凯伦（拥有诺里斯和皮亚斯特里）预计将继续对红牛发起强有力地挑战。
- 梅赛德斯在失去汉密尔顿后，将由乔治·拉塞尔带领，其竞争力存在变数。

简单总结：2025年的F1总冠军目前是一个未知的悬念，需要等到2025年赛季结束后才能揭晓。维斯塔潘目标是五连冠，而汉密尔顿在法拉利的新征程则是全球车迷关注的焦点。

无RAG：模型难以回答最近新闻相关问题

f1 2025 总冠军是谁



已思考 (用时 12 秒) >

根据多家权威媒体在2025年12月8日的报道，2025赛季F1车手总冠军是迈凯伦车队的英国车手兰多·诺里斯，这是他职业生涯首次加冕世界冠军 1 3 5。

冠军争夺战关键数据

本次冠军争夺异常激烈，悬念保留到了最后一站阿布扎比大奖赛。以下是收官后的最终积分情况 3 5：

车手	车队	总积分	收官战名次
兰多·诺里斯	迈凯伦	423分	季军
马克斯·维斯塔潘	红牛	421分	冠军
奥斯卡·皮亚斯特里	迈凯伦	410分	亚军

赛季与收官战回顾

- 激烈的冠军悬念：本赛季是近年来竞争最激烈的一季，打破了维斯塔潘过去几年提前夺冠的格局。进入收官战时，诺里斯仅领先维斯塔潘12分，三人（诺里斯、维斯塔潘、皮亚斯特里）都有理论上的夺冠可能 3 4。
- 戏剧性的过程：卫冕冠军维斯塔潘在赛季中段一度落后诺里斯多达104分，但他在后半程发起猛烈追击，将分差大幅缩小，让冠军悬念重燃 7。
- 诺里斯的登顶之路：在阿布扎比，诺里斯只需要登上领奖台（获得前三名）即可确保总冠军 3 5。最终，他顶住压力以第三名完赛，以2分的微弱优势击败维斯塔潘，成功加冕 3 6。

这次夺冠对诺里斯和迈凯伦车队都具有里程碑意义：

- 诺里斯成为F1历史上第35位车手世界冠军 3 5。
- 他终结了维斯塔潘自2021年以来的四连冠，标志着F1进入了新的竞争时代 3 4。

RAG（结合联网搜索）：模型可以根据最新新闻回答问题

本章小结

- 大规模知识表示需要通过**本体论**来组织和结合各种特定领域的知识
- 通用本体论需要涵盖各种广泛的知识，并且原则上应该有能力处理任何问题域
- 介绍了基于**类别**和**事件**演算的上位本体论
 - **类别、子类别、部分、结构化对象、量度、物质、时间、时间与空间、变化和信念等**
- 动作、事件和时间可以用事件演算来表示。这种表示使智能体能够构建动作序列并推断这些动作产生的结果。

本章小结

- 类似语义网络和描述逻辑的专用表示系统已经被设计用于组织类别层次结构。
- 继承是重要的推断形式，它使得对象的性质可以从其类别从属关系中被推导出来。
- 限定 (Abnormal) 和缺省 ($P:J/C$) 逻辑旨在总体上刻画缺省推理。

课程作业

- 将下面的描述逻辑表达式转换为一阶逻辑：
 - 一个男人的集合，这样的男人至少有三个儿子且至多有两个女儿，儿子都失业了并与医生结了婚，女儿都是物理系或数学系的教授

$$\begin{aligned} & And(Man, Atleast(3, Son), AtMost(2, Daughter)), \\ & All(Son, And(Uemployed, Married, All(Spouse, Doctor))), \\ & All(Daughter, And(Professor, Fills(Department, Physics, Math))) \end{aligned}$$

历史与趣闻

- 布里格斯 (Briggs, 1985) 声称知识表示研究始于印度人在公元前 1000 年对经典梵语语法的理论化。西方哲学家则将在这一问题上的研究追溯到公元前 300 年亚里士多德的《形而上学》 (Metaphysics, 本意为 “在《物理学》著作之后”)。各个领域发展出的术语可以被看作一种知识表示。
- **全面分类法**的发明可追溯到古代。亚里士多德高度强调了分类和类别化方法。他的著作《工具论》 (Organon) 是由其学生在他死后汇编的逻辑学研究成果，在其中的《范畴篇》 (Categories) 里，他试图构建我们今日所说的上层本体论。他还引入了属和种的概念，用于底层分类。我们现今的生物分类系统，以及对 “双名法” （用术语层面的属和种进行分类）的使用，是由瑞典生物学家卡罗勒斯·林尼厄斯 (Carolus Linnaeus, 1701—1778 发明的)。

历史与趣闻

- 对大规模本体论的兴趣正在增长，就像在 Handbook on Ontologies (Staab, 2004) 中记载的一样。OpenCYC 项目发布了一个有 150 000 个概念的本体论，如 “OLED 屏幕” 和 “iPhone” 这样的具体概念是一种 “蜂窝电话”，而 “蜂窝电话” 则是一种 “消费电子产品” “电话” “无线通信设备” 等其他概念。NextKB 项目将 CYC 和包括 FrameNet、WordNet 在内的其他资源拓展为包括近300 万条事实的知识库，并且随之提供了推理机 FIRE (Forbus et al., 2010) 。
- DBpedia 项目从维基百科提取结构化数据，特别是它也从 Infoboxes 提取数据——Infoboxes是与许多维基百科文章配套的属性-值对 (Wu and Weld, 2008; Bizer et al., 2007) 。截至2015年，DBpediea 单是对英语就包含了关于 400 万个对象的 4 亿条事实，对所有 110 种语言则有 15 亿条事实 (Lehmann et al., 2015) 。

历史与趣闻

- 事件演算是由科瓦尔斯基和塞戈特 (Kowalski and Segot, 1986) 为处理连续时间而引入的。出于相同的目的，詹姆斯·艾伦 (James Allen) 引入了时间间隔，他认为对推理延续和并发事件来说，时间间隔比情景更自然。除事件和情景演算外，还有流演算 (Thielscher, 1999)，它物化了构成状态的事实。
- 物质的本体状态问题由来已久。柏拉图提出了物质是与对象完全不同的抽象实体，他会使用 MadeOf (Butter3, Butter) 而非 Butter3 ∈ Butter。这就可以形成物质的层次，其中无盐黄油是比黄油更为确切的物质。本章采取的立场为物质是对象的类别，这是由理查德·蒙太古 (Richard Montague) (Montague, 1973)

历史与趣闻

- 事件演算是由科瓦尔斯基和塞戈特 (Kowalski and Segot, 1986) 为处理连续时间而引入的。出于相同的目的，詹姆斯·艾伦 (James Allen) 引入了时间间隔，他认为对推理延续和并发事件来说，时间间隔比情景更自然。除事件和情景演算外，还有流演算 (Thielscher, 1999)，它物化了构成状态的事实。
- 物质的本体状态问题由来已久。柏拉图提出了物质是与对象完全不同的抽象实体，他会使用 MadeOf (Butter3, Butter) 而非 Butter3 ∈ Butter。这就可以形成物质的层次，其中无盐黄油是比黄油更为确切的物质。本章采取的立场为物质是对象的类别，这是由理查德·蒙太古 (Richard Montague) (Montague, 1973)

历史与趣闻

- 希腊哲学家波菲利 (Porphyry, 234—305) 在评论了亚里士多德的《范畴论》时，描绘了可能算得上是最早的语义网络。查尔斯·皮尔斯 (Peirce, 1909) 使用现代逻辑将存在图发展为最早的语义网络形式体系。罗斯·奎利恩 (Ross Quillian) (Quillian, 1961) 在对人类记忆和语言处理的兴趣驱动下，最先研究了人工智能中的语义网络。马文·明斯基 (Minsky, 1975) 的著名论文呈现了一种称为框架的语义网络，一个框架是一个对象或类别的表示，并带有属性和与其他对象或类别的关系。
- 处理**非单调推断**的 3 种主要形式体系，即限定 (McCarthy, 1980) 、缺省逻辑 (Reiter, 1980) 和模态非单调逻辑 (McDermott and Doyle, 1980) ，都是在 AI Journal 的同期专刊中提出的。

THANKS

Q & A

10.3 时间区间

- 引入*Date*函数，使用6个参数（小时，分钟，秒，日，月，年）并返回一个时间点：
 - $Time(Begin(AD2001)) = Date(0,0,0,1,Jan,2001)$
 - $Date(0,20,21,24,1,1995) = Seconds(3,000,000,000)$

10.3 流和对象

- 从物理对象是一块空间-时间片段的意义上来讲，**物理对象可以被看做泛化事件**
 - 例如，USA可以看作是一个事件，它先作为13个州的联合体开始于1776年，今天它还在前进中，作为50个州的联合体
- 我们可以用**状态流(fluents)**来描述USA的变化的属性
 - 例如Population (USA)
 - USA每4年或8年发生一次改变（除非发生小意外）的一个属性是它的总统