

# 第二讲 知识图谱的表示

第1节 什么是知识表示

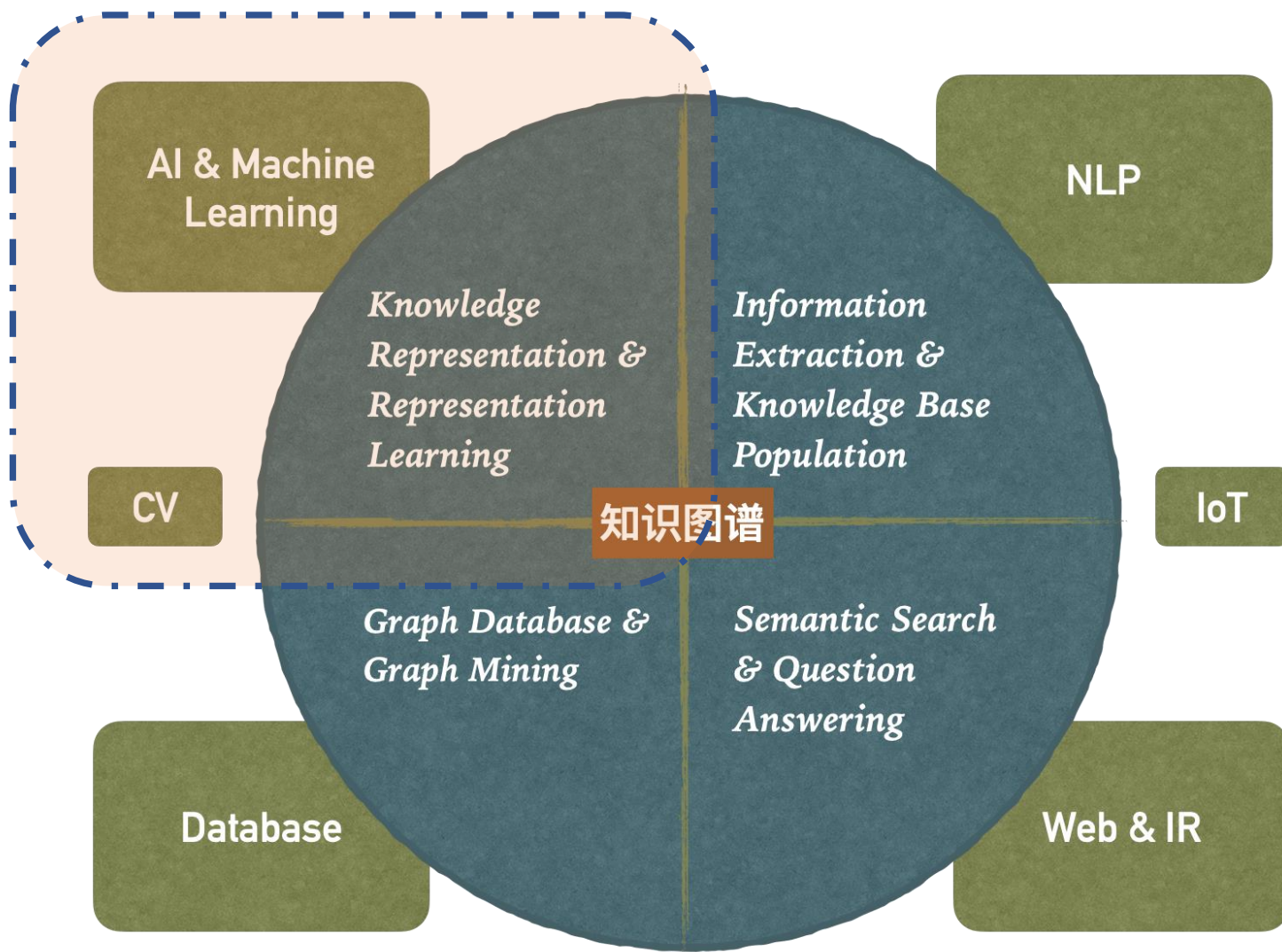
第2节 AI历史长河中的知识表示方法

第3节 知识图谱的符号表示

第4节 知识图谱的向量表示

浙江大学计算机科学与技术学院 陈华钧 教授/博导

# 知识图谱是交叉技术领域



# 第二讲 知识图谱的表示

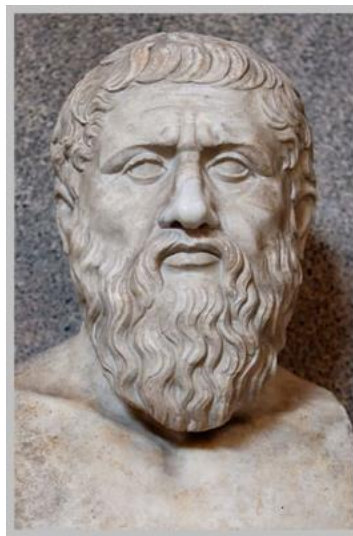
## 第1节 什么是知识表示

浙江大学计算机科学与技术学院 陈华钧 教授/博导

# What is Knowledge (in AI)?

Knowledge is justified true belief.

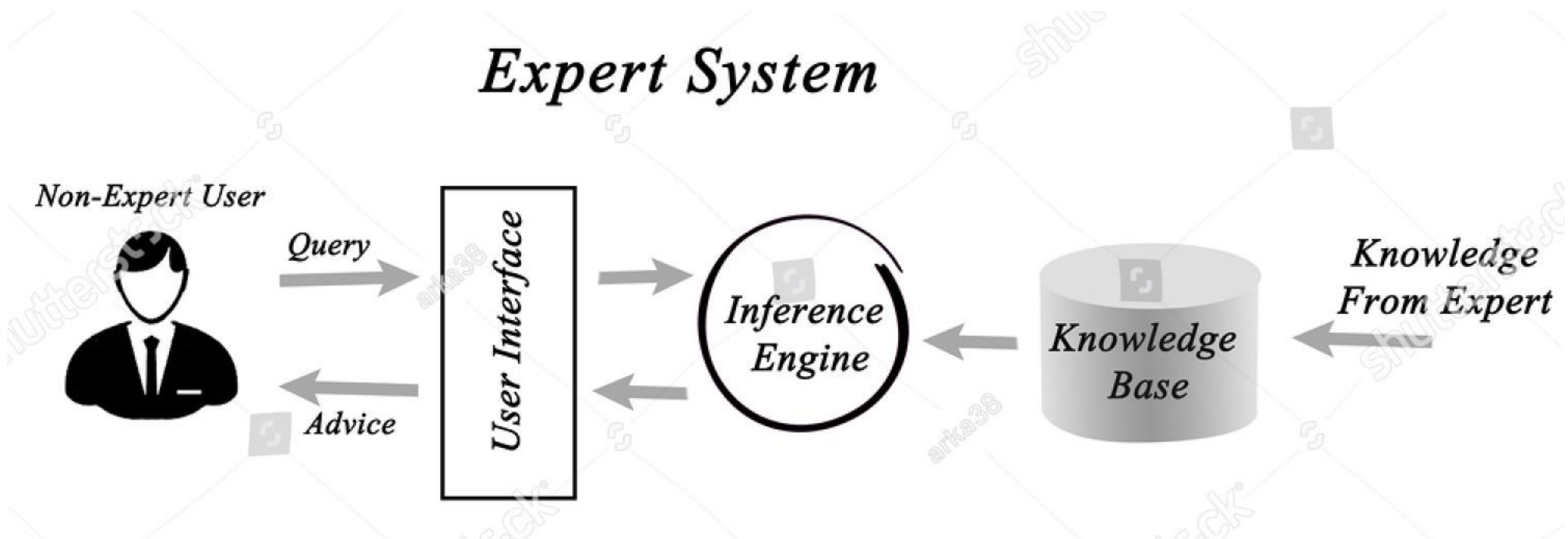
人类的自然语言，以及创作的绘画和音乐、数学语言、物理模型、化学公式等都是人类知识的表示形式和传承方式。具有获取、表示和处理知识的能力是人类心智区别于其它物种心智的最本质特征，也是人脑智能的最本质特征。



柏拉图

Knowledge is a familiarity, awareness, or understanding of someone or something, such as facts, information, descriptions, or skills, which is acquired through experience or education by perceiving, discovering, or learning.

# 知识工程与专家系统



The Good Old Fashioned AI

# What is Knowledge Representation

- 简单而言，知识表示（KR）就是用易于计算机处理的方式来描述人脑的知识的方法。
- KR不是数据格式、不等同于数据结构、也不是编程语言，对于人工智能而言，数据与知识的区别在于KR支持推理。

## What is a Knowledge Representation?

Randall Davis  
MIT AI Lab

Howard Shrobe  
MIT AI Lab and Symbolics, Inc.

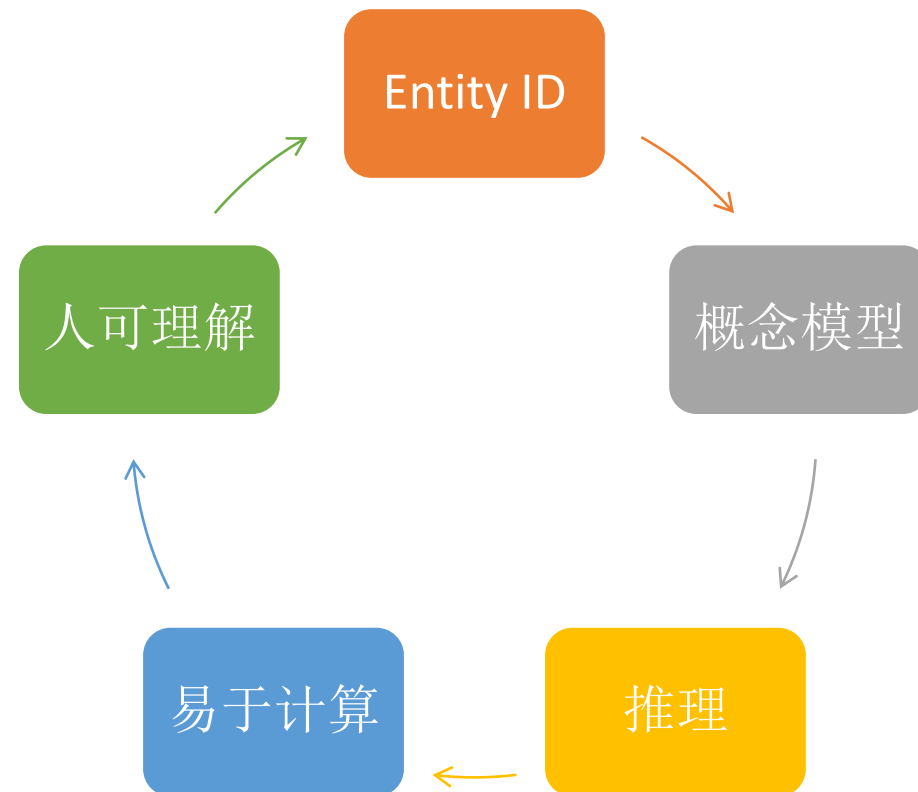
Peter Szolovits  
MIT Lab for Computer Science



R. Davis, H. Shrobe, and P. Szolovits. What is a Knowledge Representation? AI Magazine, 14(1):17–33, 1993.

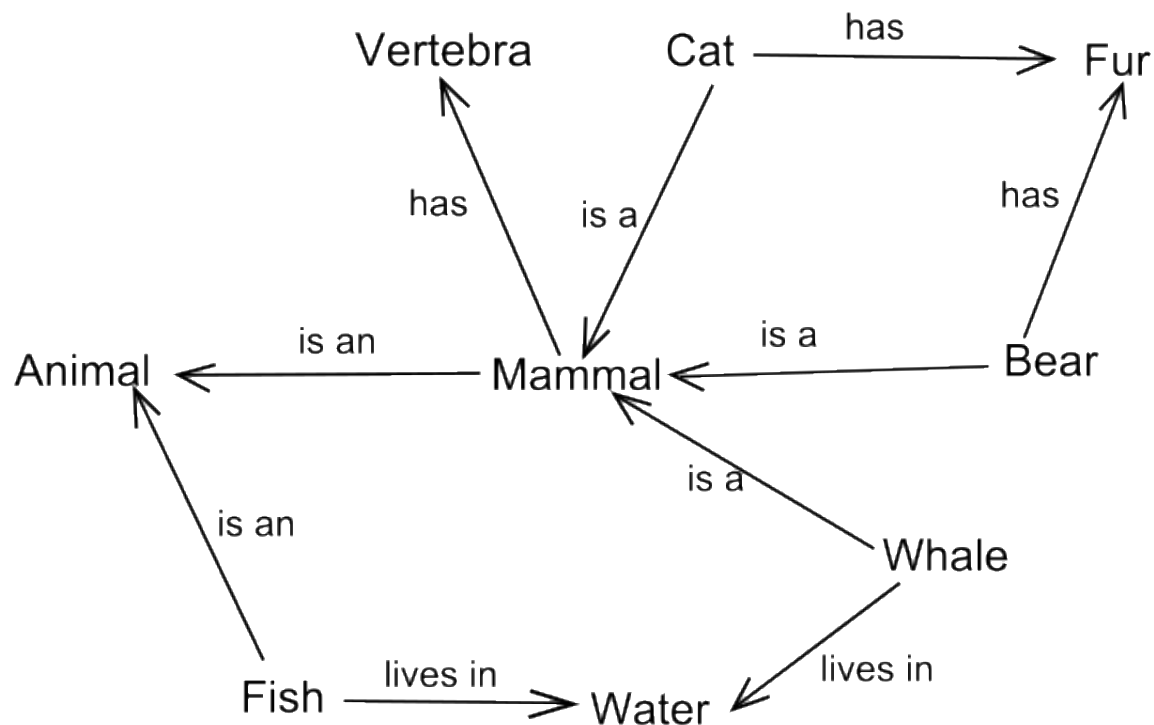
KR = Computational Model of Reality

# What is Knowledge Representation



R. Davis, H. Shrobe, and P. Szolovits. What is a Knowledge Representation? AI Magazine, 14(1):17–33, 1993.

# 知识的符号表示



- 一阶谓词逻辑 (First-Order Logic)
- 语义网络 (Semantic Net)
- 产生式规则 (Production Rule)
- 框架系统 (Framework)
- 描述逻辑 (Description Logic)
- 逻辑程序 (Logic Programming)

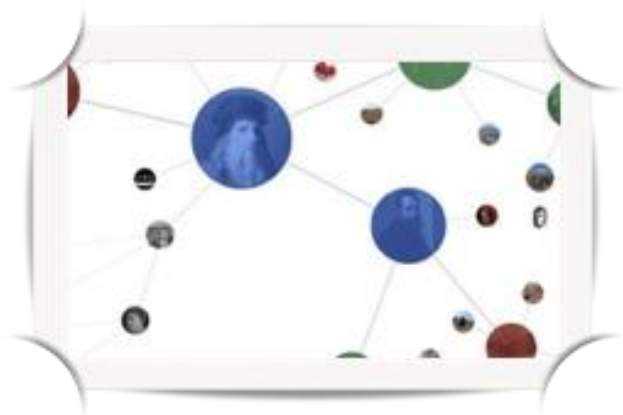


# 从符号表示到向量表示

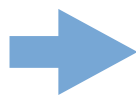
Knowledge Representation: 知识表示

基于离散符号的知识表示

RDF, OWL, 各种Rule Language等



显式知识、强逻辑约束、易于解释、推理不易扩展

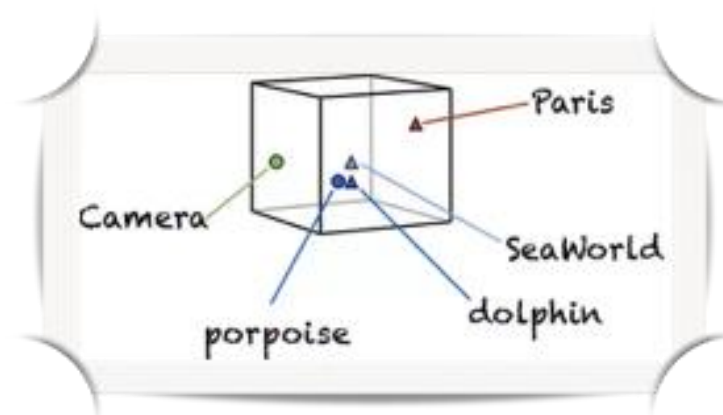


VS

Knowledge Representation Learning: 知识表示学习

基于连续向量的知识表示

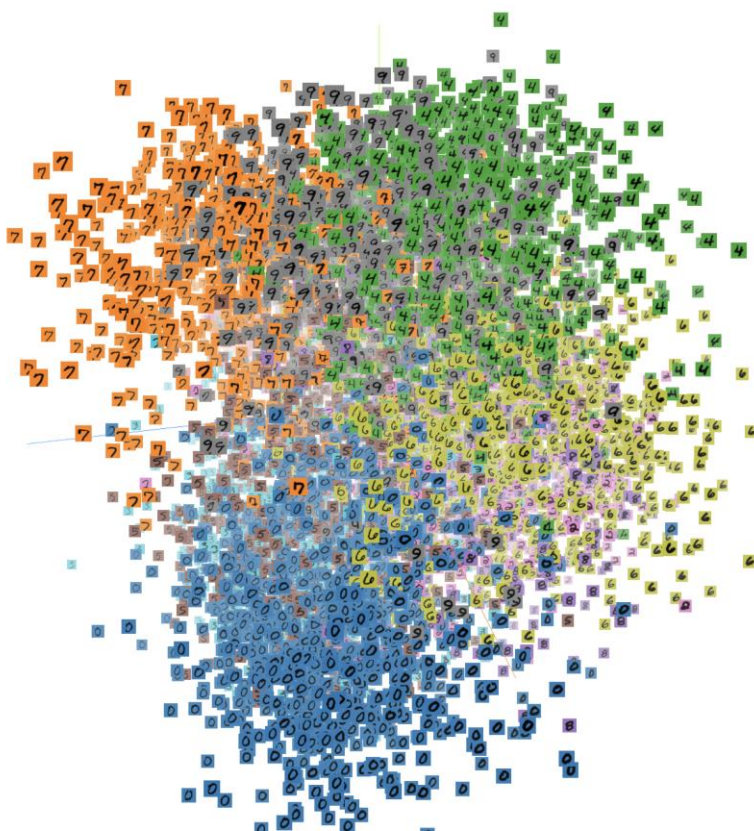
Tensor, 各种Embedding, 神经网络表示等



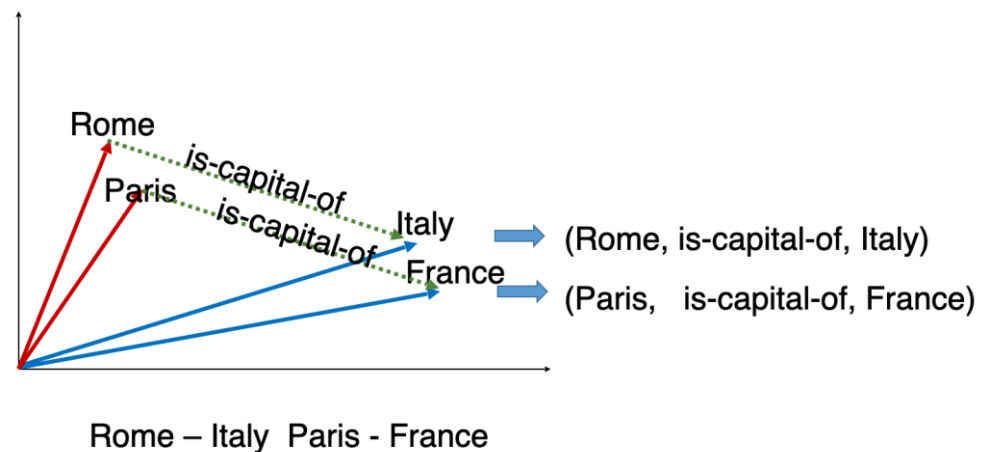
隐式知识、弱逻辑约束、不易解释、对接神经网络

# 知识的向量表示：促进多种模态的融合

Embeddings: Distributed Vector Representation



- 自然语言：为句子中的每个词学一个向量表示
- 知识图谱：为每个实体和关系学习一个向量表示
- 图像视频：为视觉中的每个对象学习一个向量表示



# 小结：Aetherial Symbols—超凡的符号



## 符号表示 + 神经网络

- The essence of intelligence is reasoning, a good starting point for understanding reasoning is formal logic.
  - 智能的精华是怎样实现推理，推理的关键是形式化逻辑。
- Learning will only be solved once we know what needs to be learned, The idea that almost everything is learned is absurd.
  - “学习”需要我们清楚的知道要学习的是什么，那些认为任何东西都可以学习出来的想法是愚蠢的。
- There are no pixels or symbol strings in the head. All we have in our heads is big activity vectors that cause more big activity vectors.
  - 大脑中并没有符号和图像的存储，而是一些Big Activity Vectors之间相互发生作用。

[HTTP://OpenKG.CN](http://OpenKG.CN)



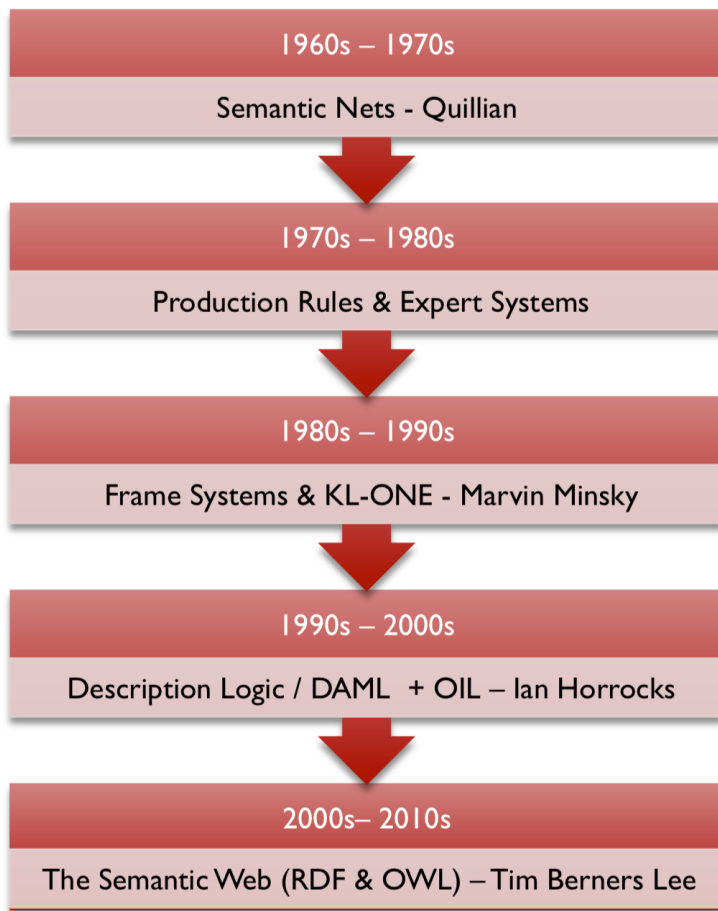
谢谢大家！

# 第二讲 知识图谱的表示

## 第2节 人工智能历史发展长河中的知识表示

浙江大学计算机科学与技术学院 陈华钧 教授/博导

# 古老的知识表示研究



- 一阶谓词逻辑 (First-Order Logic)
  - Horn Logic
  - Description Logic
- 语义网络 (Semantic Net)
- 产生式规则 (Production Rule)
- 框架系统 (Framework)
- 描述逻辑 (Description Logic)

# Description Logic

- 描述逻辑是一阶谓词逻辑的可判定子集，主要用于描述本体概念和属性，对于本体知识库的构建提供了便捷的表达形式，是与知识图谱最密切相关的知识表示方法之一。
- 核心表达要素：
  - 概念Concepts—解释为一个领域的子集
    - 例如：学生，已婚者：  $\{x \mid \text{Student}(x)\}$  ,  $\{x \mid \text{Married}(x)\}$
  - 关系Relations—解释为指该领域上的二元关系
    - 例如：示例：朋友，爱人：  $\{\langle x, y \rangle \mid \text{friend}(x, y)\}$  ,  $\{\langle x, y \rangle \mid \text{loves}(x, y)\}$
  - 个体Individuals—一个领域内的实例
    - 例如：小明，小红：  $\{\text{Ming}, \text{Hong}\}$

# Description Logic

- 描述逻辑的知识库  $O := \langle T, A \rangle$ , T即Tbox, A即Abox
  - TBox包含内涵知识, 描述概念的一般性质
    - 定义: 引入概念以及关系的名称, 例如: `Mother, Person, has_child`
    - 包含: 声明包含关系的公理, 例如: `Mother  $\sqsubseteq$   $\exists$  has_child.Person`
  - ABox包含外延知识 (又称断言知识), 描述论域中的特定个体。
    - 概念断言—表示一个对象是否属于某个概念, 例如: `Mother(Helen),`
    - 关系断言—表示两个对象是否满足一定的关系, 例如: `has_child(Helen, Jack)`



# Horn Logic

➤ Horn Logic是一阶谓词逻辑的子集，主要特点是表达形式简单，复杂度低。  
著名的Prolog语言就是基于Horn逻辑设计实现的。

➤ 核心表达要素：

➤ 原子Atoms

➤  $p(t_1, t_2, \dots, t_n)$ ,  $p$  是谓词,  $t_i$  是项（变量或者常量），例子: `has_child(X, Y)`

➤ 规则Rules

➤ 由原子构建:  $H :- B_1, B_2, \dots, B_m$

➤  $H$ 称为Head;  $B_1, B_2, \dots, B_m$ 称为Body, 例子: `has_child(X, Y) :- has_son(X, Y)`

➤ 事实Facts:

➤ 没有体部且没有变量的规则, 例子: `has_son(Helen, Jack) :-`

# 一阶谓词逻辑优缺点比较

## 优点

- 接近自然语言，容易被接受，且易于表示精确知识
- 有严格的形式定义和推理规则，易于精确实现

## 缺点

- 无法表示不确定性知识，难以表示启发性知识及元知识
- 组合爆炸，经常出现事实、规则等的组合爆炸，导致效率低，推理复杂度通常较高
- 是以人的逻辑为主导的表示方式，可能并不适合机器，机器可能有自己的逻辑表示。

# Production Systems: 产生式系统

➤产生式系统是一种更广泛意义的规则系统，专家系统多数是基于产生式系统：

➤Feigenbaum研制的化学分子结构专家系统DENDRAL

➤Shortliffe研制的诊断感染性疾病的专家系统MYCIN

➤产生式系统的核心表达形式：

➤IF P THEN Q     $CF = [0, 1]$ ，其中P是产生式的前提，Q是一组结论或操作，CF (Certainty Factor)为确定性因子，也称置信度。

*IF    本微生物的染色斑是革兰氏阴性  
      本微生物的形状呈杆状  
      病人是中间宿主  
THEN 该微生物是绿脓杆菌，置信度为 $CF=0.6$*

# 产生式系统的优缺点

## 优点

- 自然性：产生式系统采用人类常用的因果关系知识表示形式，既直观、自然，又便于进行推理。
- 模块性：产生式规则形式相同，易于模块化管理。
- 有效性：能表示确定性知识、不确定性知识、启发性知识、过程性知识等。
- 清晰性：产生式有固定的格式，既便于规则设计，又易于对规则库中的知识进行一致性、完整性检测。

## 缺点

- 效率不高：产生式系统求解问题的过程是一个反复进行“匹配—冲突消解—执行”的过程。而规则库一般都比较大会，匹配又十分费时，因此其工作效率不高。
- 在求解复杂问题时容易引起组合爆炸。
- 不能表达结构性知识：产生式系统对具有结构关系的知识无能为力，它不能把具有结构关系的事物间的区别与联系表示出来。

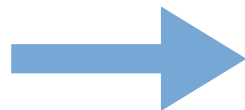
# Frame System: 框架系统

- 框架理论的基本思想：认为人们对现实世界中事物的认识都是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，就从记忆找出一个合适的框架，并根据实际情况对其细节加以修改、补充，从而形成对当前事物的认识。框架系统在很多NLP任务如Dialogue系统中都有广泛的应用。
- 基本表达要素：
  - 框架：是一种描述对象（事物、事件或概念等）属性的数据结构。
  - 一个框架由若干个“槽”（Slot）结构组成，每个槽又可分为若干个“侧面”。
    - 一个槽：用于描述所论对象某一方面的属性；
    - 一个侧面：用于描述相应属性的一个方面。
    - 槽和侧面所具有的属性值分别称为槽值和侧面值。

# Frame System: 框架系统

<框架名>

槽名1:	侧面名1	值1, 值2, ..., 值p1
	侧面名2	值1, 值2, ..., 值p2
	侧面名m1	值1, 值2, ..., 值pm1
槽名n:	侧面名1	值1, 值2, ..., 值r1
约束:	约束条件1	
	约束条件n	



框架名:	小A抢劫杀人案
犯罪意图:	抢劫
犯罪结果:	杀人
被杀者:	小B
知情人:	小C
罪犯:	小A
条件一:	有小C指控小A
条件二:	小A招认

# 框架系统优缺点

## 优点

- 对于知识的描述完整和全面；
- 基于框架的知识库质量高；
- 框架允许数值计算。

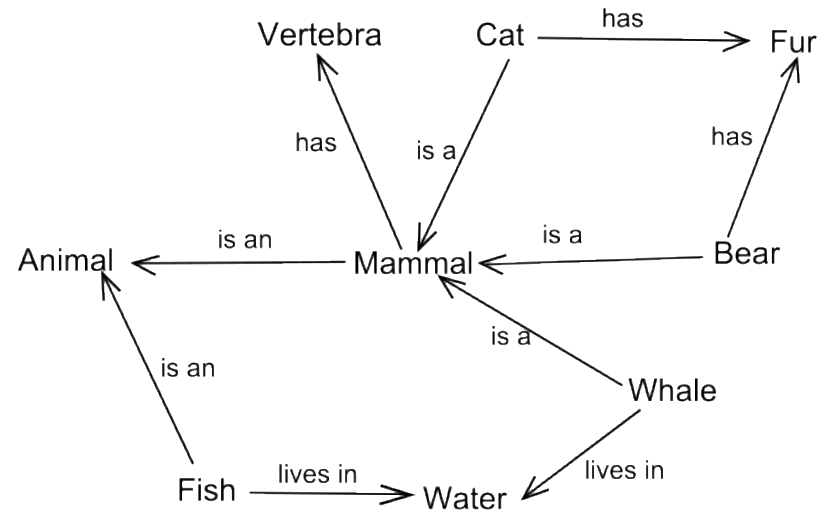
## 缺点

- 框架构建成本高，质量要求高；
- 框架的表达形式不灵活，很难同其它形式的数据集相互关联使用。

# 语义网络 (Semantic Network)

- 1968年J. R. Quillian在其博士论文中最先提出语义网络，把它作为人类联想记忆的一个显式心理学模型，并在他设计的可教式语言理解器TLC (Teachable Language Comprehenden)中用作知识表示方法。

```
(defun *database* ()  
  '((canary (is-a bird)  
           (color yellow)  
           (size small))  
    (penguin (is-a bird)  
             (movement swim))  
    (bird (is-a vertebrate)  
          (has-part wings)  
          (reproduction egg-laying))))
```



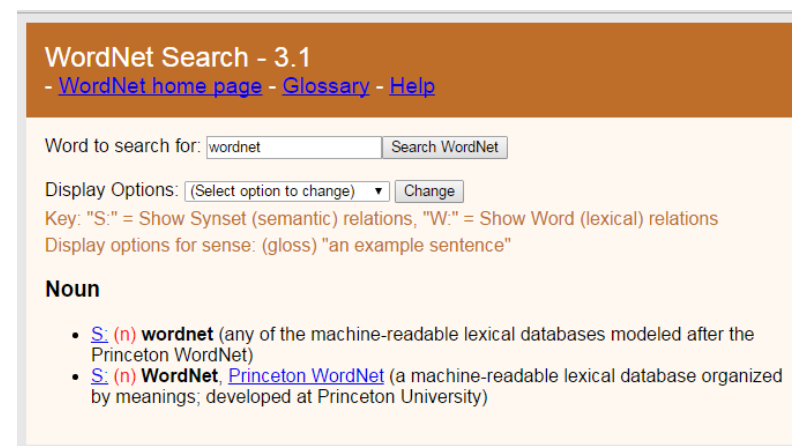


# 语义网络 (Semantic Network) —— WordNet

- WordNet是最著名的词典知识库，主要用于词义消歧。WordNet由普林斯顿大学认识科学实验室从1985年开始开发。
- WordNet主要定义了名词、动词、形容词和副词之间的语义关系。例如名词之间的上下位关系（如：“猫科动物”是“猫”的上位词），动词之间的蕴含关系（如：“打鼾”蕴含着“睡眠”）等。
- WordNet3.0已经包含超过15万个词和20万个语义关系。

## Hamburger

- Hamburger (an inhabitant of Hamburg)
  - direct hypernym:
    - German (a person of German nationality)
  - sister term
    - German (a person of German nationality)
      - East German (a native/inhabitant of the former GDR)
      - Bavarian (a native/inhabitant of Bavaria)
  - derivationally related form
    - Hamburg (a port city in northern Germany on the Elbe River that was founded by Chalemagne in the...)



# 语义网络的优缺点

## 优点

- 结构性：语义网络是一种结构化的知识表示方法，它能把事物的属性以及事物间的各种语义联想显式地表示出来。
- 联想性：最初是作为人类联想记忆模型提出来的。
- 自然性：直观地把事物的属性及其语义联系表示出来，便于理解，自然语言与语义网络的转换比较容易实现，故语义网络表示法在自然语言理解系统中应用最为广泛。

## 缺点

- 非严格性：语义网络没有公认的形式表示体系。一个给定的语义网络所表达的含义完全依赖于处理程序如何对它进行解释。
- 通过推理网络而实现的推理不能保证其正确性。支持全称量词和存在量词的语义网络在逻辑上是不充分的，不能保证不存在二义性。
- 处理上的复杂性：语义网络表示知识的手段多种多样，灵活性高，但由于表示形式不一致使得处理复杂性高，对知识的检索相对复杂。

# 小结

- 知识表示与推理在人工智能的发展历史上一直居于核心位置。
- 传统的知识表示方法主要以符号表示为主，与知识图谱有关的典型方法有：  
描述逻辑、Horn Logic、产生式规则、框架系统、语义网络等等。
- 这些方法各有优缺点，但都有一个共同的缺点是知识的获取过程主要依靠专家和人工，越复杂的知识表示框架知识获取过程越困难。

[HTTP://OpenKG.CN](http://OpenKG.CN)



谢谢大家!

# 第二讲 知识图谱的表示

## 第3节 知识图谱的符号表示方法

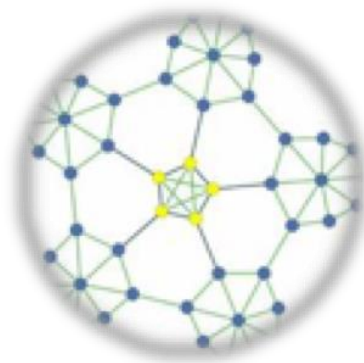
浙江大学计算机科学与技术学院 陈华钧 教授/博导

# 基于图的知识表示与建模

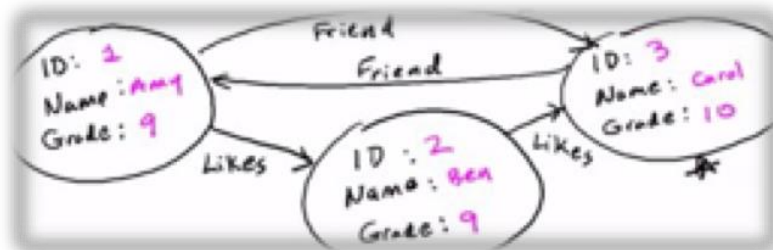
Knowledge Graph is more expressive than pure graph but less complex than formal logic

表达能力逐渐增强

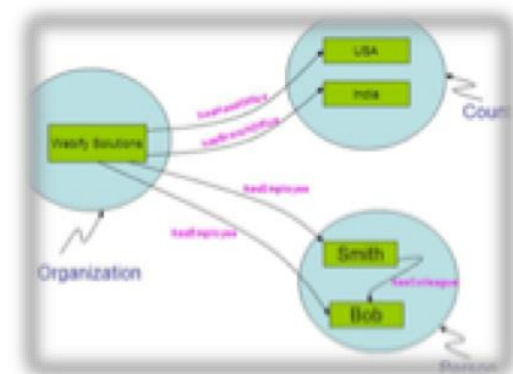
简单图建模



有向标记图  
RDF & Property Graph



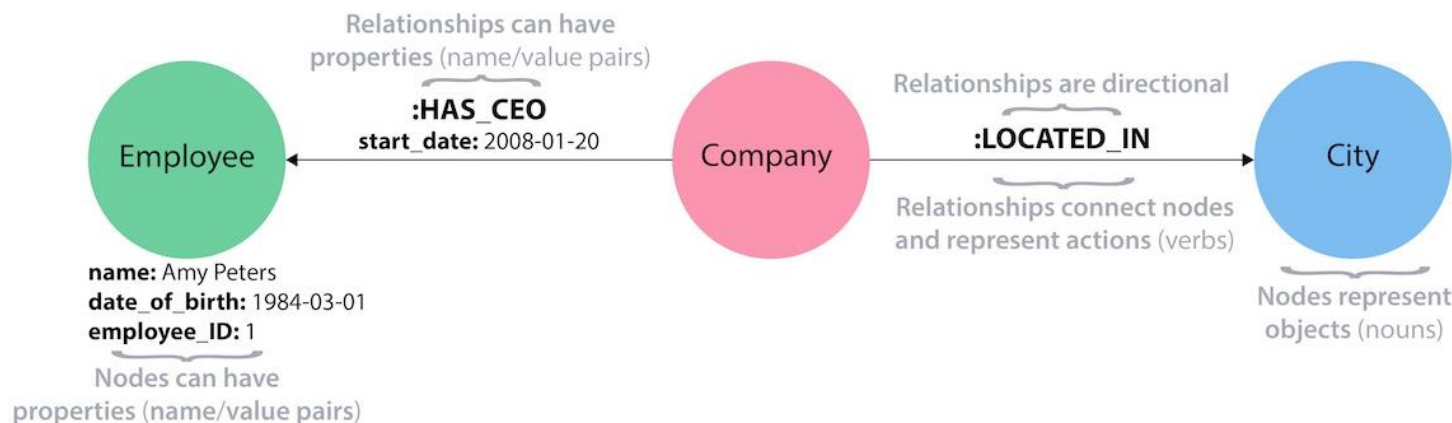
OWL与Ontology



推理能力逐渐变弱

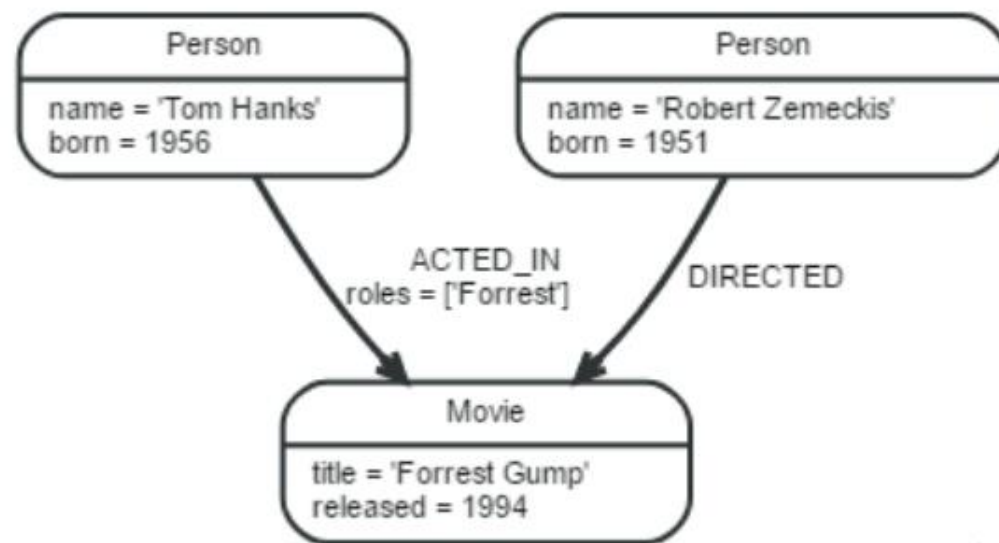
# Property Graph—属性图

- 属性图是图数据库Neo4J实现的图结构表示模型，在工业界有广泛应用。
- 属性图的优点是表达方式非常灵活，例如，它允许为边增加属性，非常便于表示多元关系。
- 属性图的存储充分利用图的结构进行优化，因而在查询计算方面具有较高优势。
- 属性图的缺点是缺乏工业标准规范的支持，由于不关注更深层的语义表达，也不支持符号逻辑推理。



# Property Graph—属性图

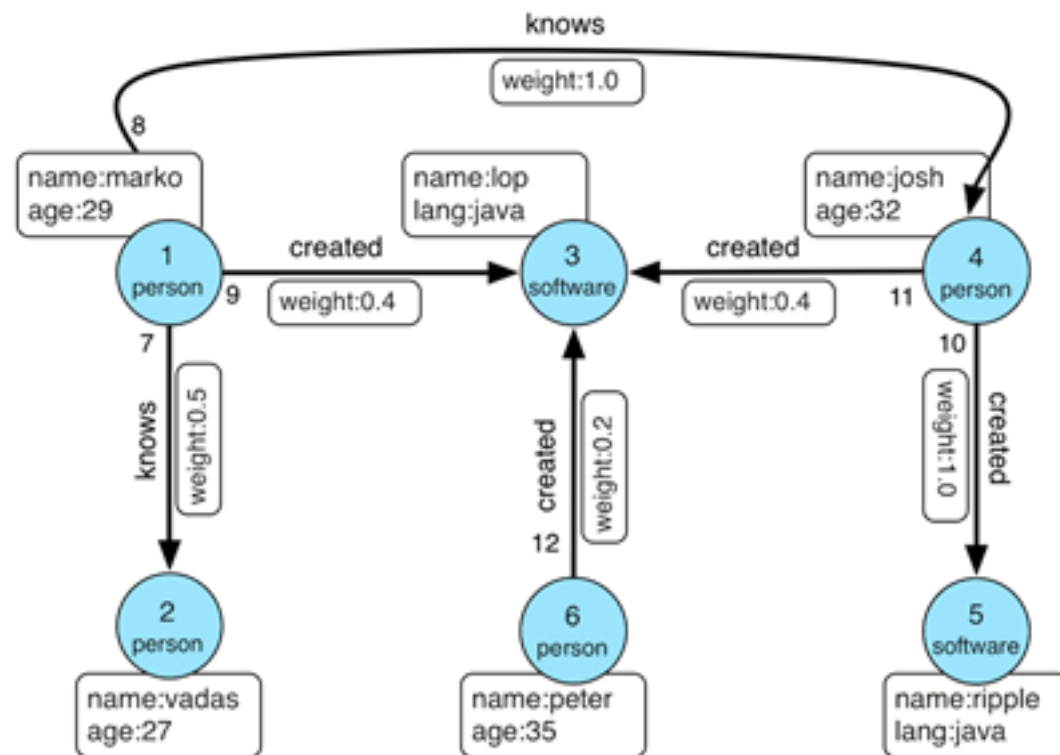
- 在属性图的术语中，属性图是由 顶点（Vertex），边（Edge），标签（Label），关系类型还有属性（Property）组成的有向图。
- 顶点也称为 节点（Node），边也称为 关系（Relationship）。
- 在属性图中，节点和关系是最重要的实体。节点上包含属性，属性可以以任何键值形式存在。





# Property Graph—属性图

- 关系连接节点，每个关系都有一个方向、一个标签、一个开始节点和结束节点。关系的方向的标签使得属性图具有语义化特征。
- 和节点一样，关系也可以有属性，即边属性，可以通过在关系上增加属性给图算法提供有关边的元信息，如创建时间等，此外还可以通过边属性为边增加权重和特性等其他额外语义。

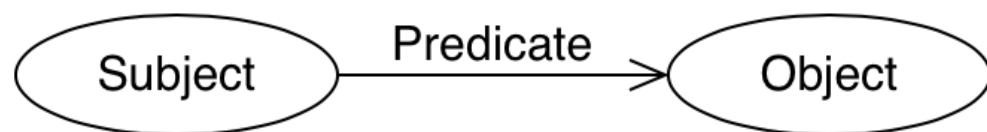


# RDF: Triple-based Assertion model

RDF 代表 Resource Description Framework (资源描述框架)

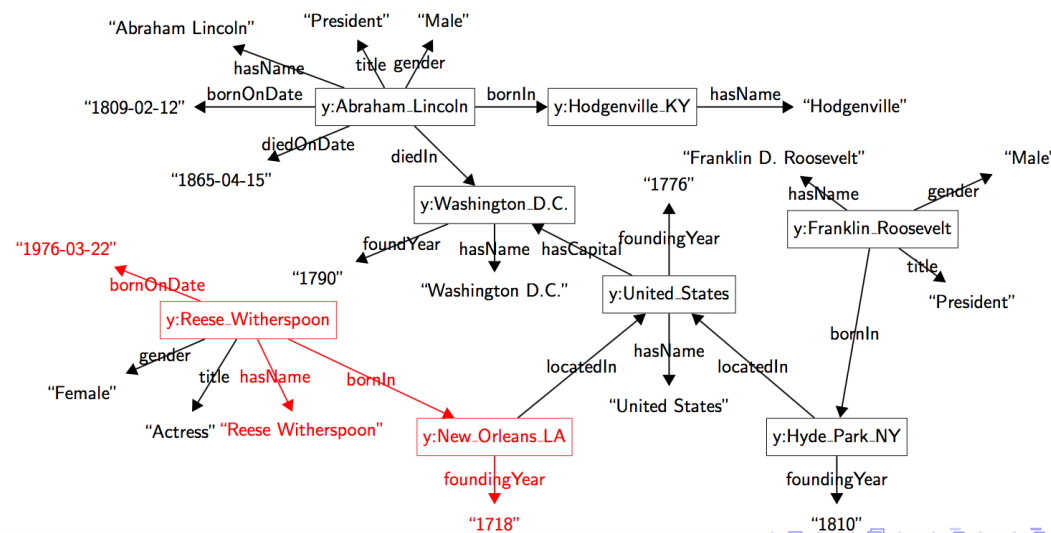
RDF是国际万维网联盟W3C推动的面向Web的语义数据标准

An RDF triple (S,P,O) encodes a statement—a simple **logical expression**, or claim about the world.



( subject (主), predicate (谓), object (宾) )

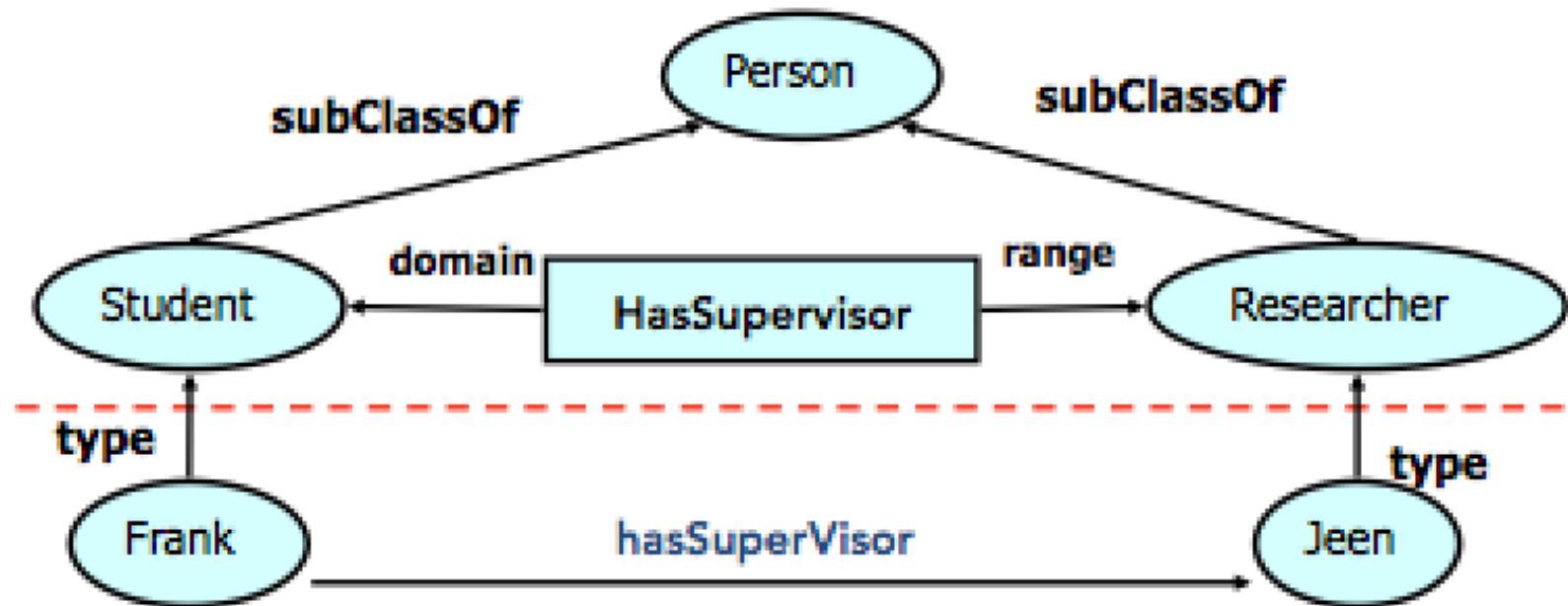
( subject (浙江大学), predicate (位于), object (杭州) )



# RDFS: Simple Vocabulary and Schema

Defines small vocabulary for RDF:

Class, subClassOf, type, Property, subPropertyOf, Domain, Range



# 基于RDFS的简单推理

谷歌 **rdf:type** 人工智能公司



人工智能公司 **rdfs:subclass** 高科技公司



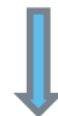
谷歌 **rdf:type** 高科技公司

投资 **rdfs:domain** 投资人

投资 **rdfs:range** 公司



大卫·切瑞顿 投资 谷歌



大卫·切瑞顿 **rdf:type** 投资人

# OWL: Web Ontologies

Ontology in Philosophy

Ontology is the philosophical study of the nature of being, becoming, existence or reality, as well as the basic categories of being and their relations.

---

Merriam-Webster

Ontology in Computer Science and Artificial Intelligence

An ontology is a description (like a formal specification of a program) of the concepts and relationships that can formally exist for an agent or a community of agents.

---

Tom Gruber, Founder of Siri

Web Ontologies

Ontologies based on web standards such as RDFS/OWL. OWL is based on Description Logic, a very very long history of research in Artificial Intelligence.

# OWL extends RDF Schema

**RDF-Schema:** Class, subclass, Property, subProperty.....



**Complex Classes:** intersection, union and complement

**Property Restrictions:** existential quantification, universal quantification, hasValue

**Cardinality Restrictions:** maxQualifiedCardinality, minQualifiedCardinality, qualifiedCardinality

**Property Characteristics:** inverseOf, SymmetricProperty, AsymmetricProperty, propertyDisjointWith, ReflexiveProperty, FunctionalProperty

**Property Chains**

# OWL的表达构件

## 1. 等价性声明

```
exp:运动员 owl:equivalentClass exp:体育选手  
exp:获得 owl:equivalentProperty exp:取得  
exp:运动员A owl:sameIndividualAs exp:小明
```

## 2. 声明属性的传递性

```
exp:ancestor rdf:type owl:TransitiveProperty  
  
exp:小明 exp:ancestor exp:小林;  
exp:小林 exp:ancestor exp:小志  
  
推理得出: exp:小明 exp:ancestor exp:小志 .
```

## 3. 声明两个属性互反

```
exp:ancestor owl:inverseOf exp:descendant  
exp:小明 exp:ancestor exp:小林  
  
推理得出: exp:小林 exp:descendant exp:小明
```

# OWL的表达构件

## 4. 声明属性的函数性

```
exp:hasMother rdf:type owl:FunctionalProperty
```

exp:hasMother 是一个具有函数性的属性，因为每个人只有一个母亲，作为约束作用到知识库

## 5. 声明属性的对称性

```
exp:friend rdf:type owl:SymmetricProperty
```

```
exp:小明 exp:friend exp:小林
```

推理得出：exp:小林 exp:friend exp:小明

## 6. 声明属性的局部约束：全称限定

```
exp:Person owl:allValuesFrom exp:Women
```

```
exp:Person owl:onProperty exp:hasMother
```

exp:hasMother在主语属于exp:Person类的时候，宾语的取值只能来自exp:Women这个类。



# OWL的表达构件

## 7. 声明属性的局部约束：存在限定

```
exp:SemanticWebPaper owl:someValuesFrom exp:AAAI  
exp:SemanticWebPaper owl:onProperty exp:publishedIn
```

exp:publishedIn在主语属于exp:SemanticWebPaper类的时候，宾语的取值部分来自exp:AAAI这个类。上面的三元组相当于：关于语义网的论文部分发表在AAAI上。

## 8. 声明属性的局部约束：基数限定

```
exp:Person owl:cardinality "1" ^^xsd:integer  
exp:Person owl:onProperty exp:hasMother
```

exp:hasMother在主语属于exp:Person类的时候，宾语的取值只能有一个；“1”的数据类型被声明为xsd:integer；  
这是基数约束，本质上属于属性的局部约束。

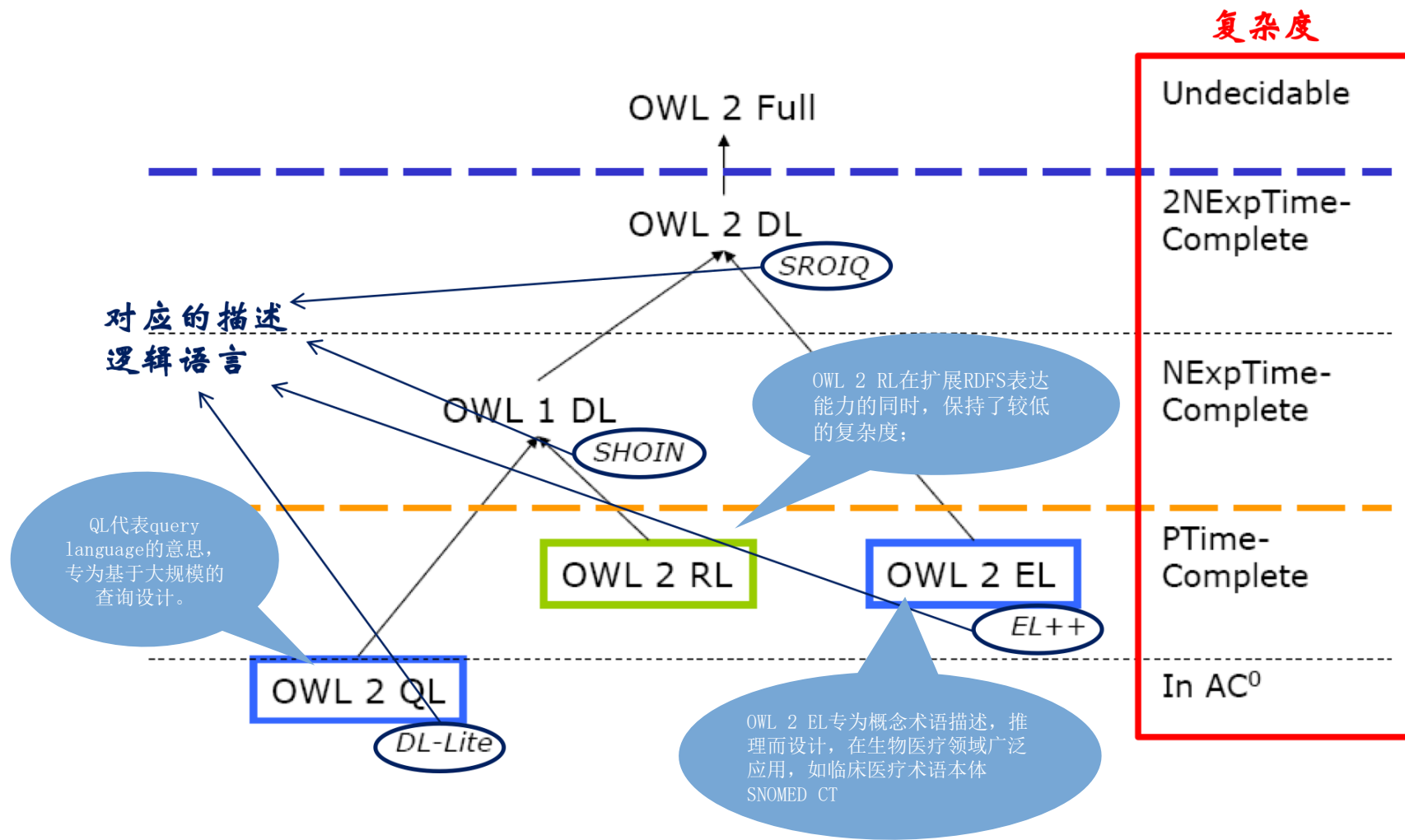
## 9. 声明相交的类

```
exp:Mother owl:intersectionOf _tmp  
_tmp rdf:type rdfs:Collection  
_tmp rdfs:member exp:Person  
_tmp rdfs:member exp:HasChildren
```

\_tmp是临时资源；它是rdfs:Collection类型，是一个容器；它的两个成员是exp:Person，exp:HasChildren；上述三元组说明exp:Mother是exp:Person exp:HasChildren这两个类的交集。

# OWL语言家族

每一种子语言是前述语义表达构件的一类集合，并有相应的复杂度分析



# 小节：属性图 vs RDF vs OWL

- 属性图是工业界最常见的图谱建模方法，属性图数据库充分利用图结构特点做了性能优化，实用度高，但不支持符号推理。
- RDF是W3C推动的语义数据交换标准与规范，有更严格的语义逻辑基础，支持推理，并兼容更复杂的本体表示语言OWL。
- 在三元组无法满足语义表示需要时，OWL作为一种完备的本体语言，提供了更多可供选用的语义表达构件。
- 描述逻辑可以为知识图谱的表示与建模提供理论基础。描述逻辑之于知识图谱，好比关系代数之于关系数据库。

[HTTP://OpenKG.CN](http://OpenKG.CN)



谢谢大家！

# 第二讲 知识图谱的表示

## 第4节 知识图谱的向量表示方法

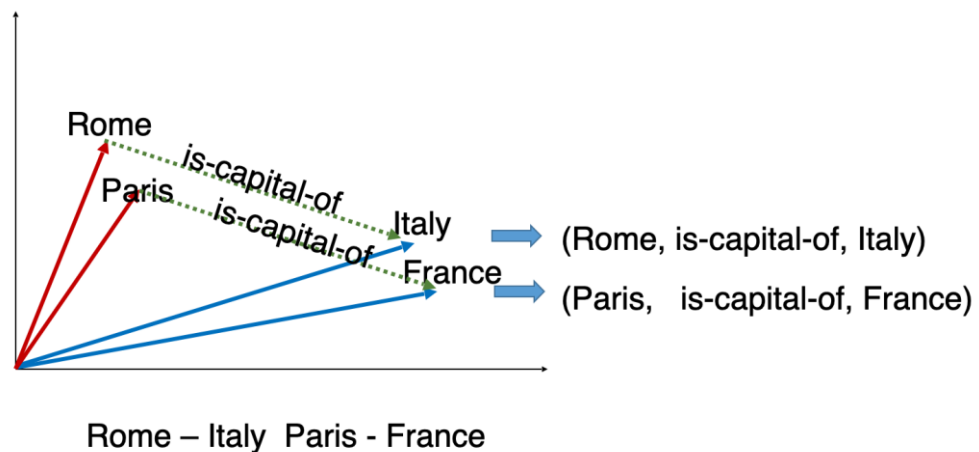
浙江大学计算机科学与技术学院 陈华钧 教授/博导

# 知识的向量表示

Embeddings: Distributed Vector Representation



- 自然语言：为句子中的每个词学一个向量表示
- 知识图谱：为每个实体和关系学习一个向量表示
- 图像视频：为视觉中的每个对象学习一个向量表示



# 从词的向量表示讲起

## One-hot Encoding

Rome Paris Italy France ... Word k

Rome = [1, 0, 0, 0, 0, 0, ..., 0]

Paris = [0, 1, 0, 0, 0, 0, ..., 0]

Italy = [0, 0, 1, 0, 0, 0, ..., 0]

France = [0, 0, 0, 1, 0, 0, ..., 0]

k = 词典大小

## Bag-of-words

Rome Paris Italy France ... Word k

doc\_1 = [32, 14, 1, 0, ..., 6]

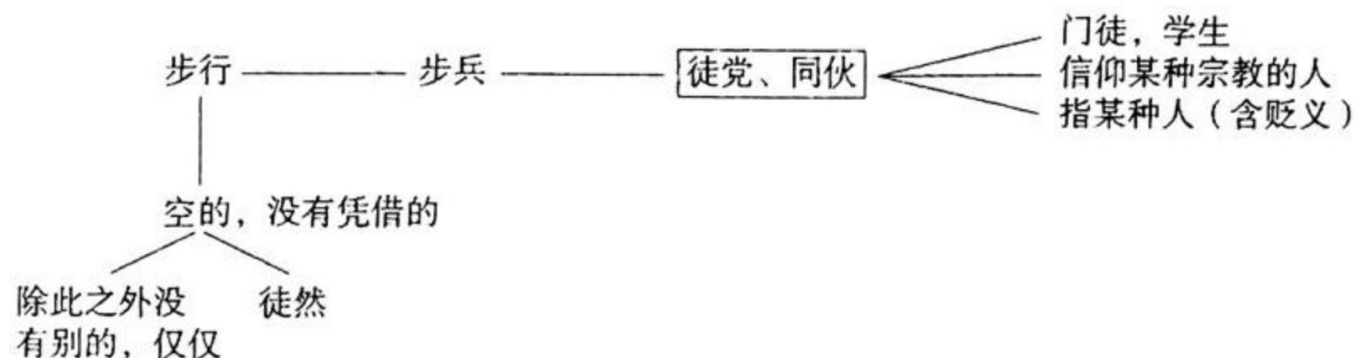
doc\_2 = [2, 12, 0, 28, ..., 12]

... ..

doc\_N = [13, 0, 6, 2, ..., 0]

# 词的分布式向量表示

词的语义由什么决定？



Distributional semantics:

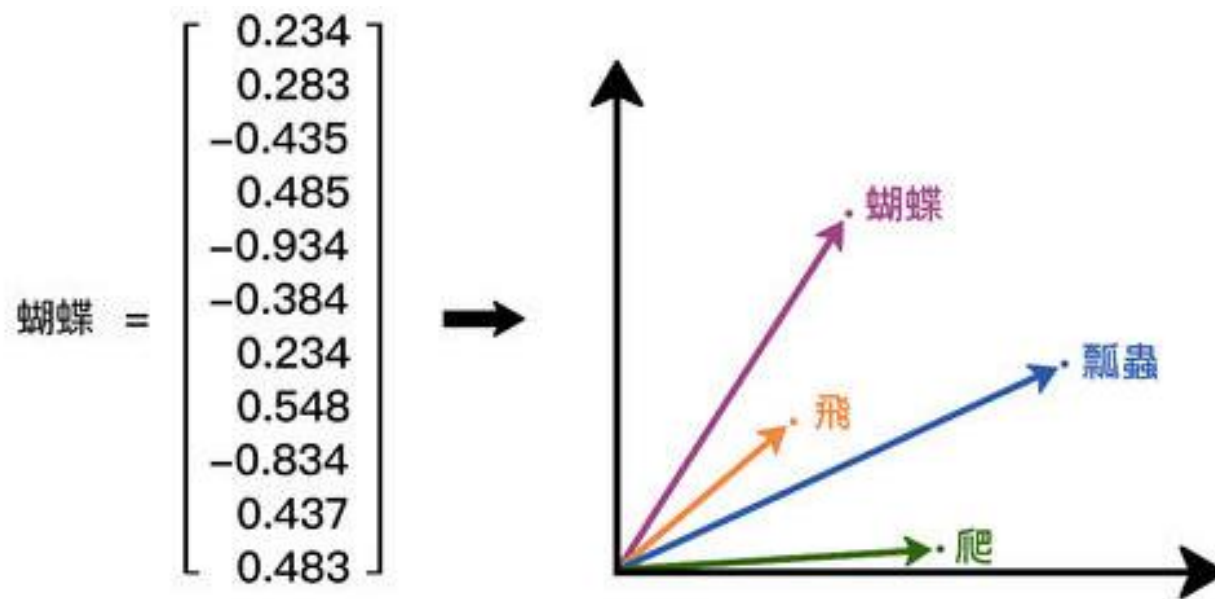
"You shall know a word by the company it keeps." — J.R. Firth (1957)

词是符号化的，词的语义由它被使用的上下文确定



# 从词的向量表示讲起

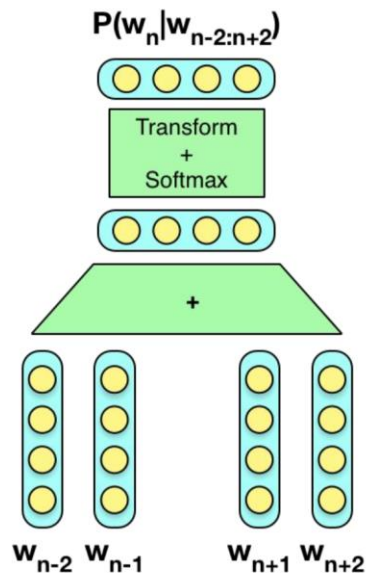
Word Embedding/词嵌入  
=  
Word Vectors/词向量  
=  
Distributed Representations/词的分布式表示



$d$  = 向量维度  $\ll$  词典大小

# 词向量学习模型举例：

## CBoW



Embed context words. Add them.

Project back to vocabulary size. Softmax.

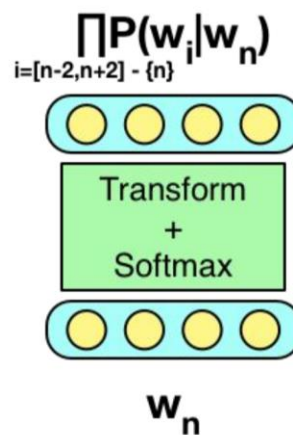
$$\text{softmax}(\mathbf{l})_i = \frac{e^{l_i}}{\sum_j e^{l_j}}$$

$$\begin{aligned} P(t_i | \text{context}(t_i)) &= \text{softmax} \left( \sum_{t_j \in \text{context}(t_i)} \text{onehot}_{t_j}^\top \mathbf{E} W_v \right) \\ &= \text{softmax} \left( \left( \sum_{t_j \in \text{context}(t_i)} \text{onehot}_{t_j}^\top \mathbf{E} \right) W_v \right) \end{aligned}$$

Minimize Negative Log Likelihood:

$$L_{data} = - \sum_{t_i \in data} \log P(t_i | \text{context}(t_i))$$

## Skip-gram



Target word predicts context words.

Embed target word.

Project into vocabulary. Softmax.

$$P(t_j | t_i) = \text{softmax}(\text{onehot}_{t_i}^\top \mathbf{E} W_v)$$

Learn to estimate likelihood of context words.

$$\begin{aligned} -\log P(\text{context}(t_i) | t_i) &= -\log \prod_{t_j \in \text{context}(t_i)} P(t_j | t_i) \\ &= - \sum_{t_j \in \text{context}(t_i)} \log P(t_j | t_i) \end{aligned}$$

# 从词的表示讲起

Rome = [0.91, 0.83, 0.17, ..., 0.41]

Paris = [0.92, 0.82, 0.17, ..., 0.98]

Italy = [0.32, 0.77, 0.67, ..., 0.42]

France = [0.33, 0.78, 0.66, ..., 0.97]

Rome = [0.91, 0.83, 0.17, ..., 0.41]

Paris = [0.92, 0.82, 0.17, ..., 0.98]

Italy = [0.32, 0.77, 0.67, ..., 0.42]

France = [0.33, 0.78, 0.66, ..., 0.97]

# 从词的表示讲起

Rome = [0.91, 0.83, 0.17, ..., 0.41]

Paris = [0.92, 0.82, 0.17, ..., 0.98]

Italy = [0.32, 0.77, 0.67, ..., 0.42]

France = [0.33, 0.78, 0.66, ..., 0.97]

Rome = [0.91, 0.83, 0.17, ..., 0.41]

Paris = [0.92, 0.82, 0.17, ..., 0.98]

Italy = [0.32, 0.77, 0.67, ..., 0.42]

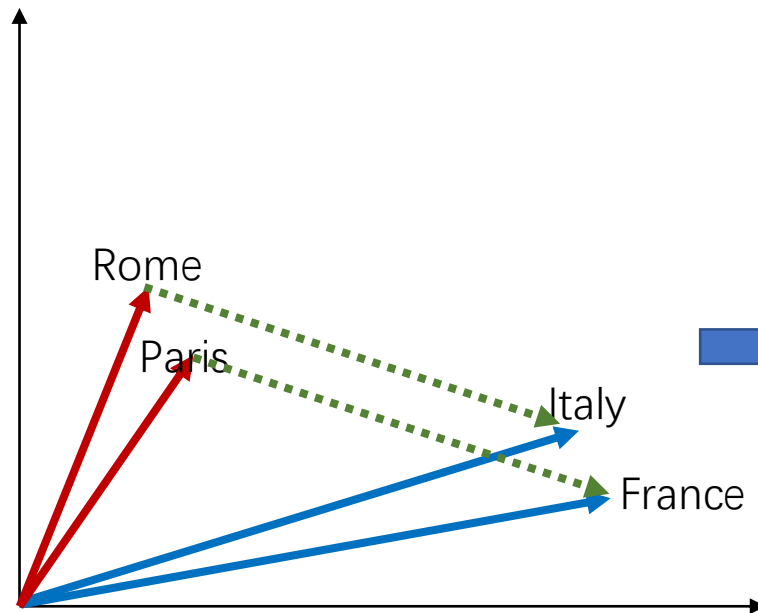
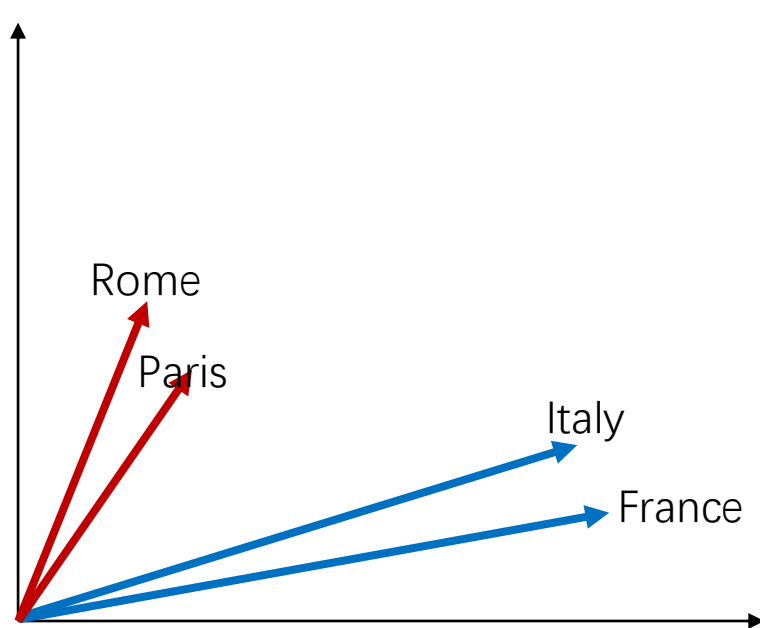
France = [0.33, 0.78, 0.66, ..., 0.97]

# 从词的表示讲起

## ➤ One-hot Encoding

Rome	=	[1, 0, 0, 0, 0, 0, ..., 0]	} 任意两两之间相似度为0
Paris	=	[0, 1, 0, 0, 0, 0, ..., 0]	
Italy	=	[0, 0, 1, 0, 0, 0, ..., 0]	
France	=	[0, 0, 0, 1, 0, 0, ..., 0]	

# 词嵌入 $\rightarrow$ 知识图谱嵌入



Rome – Italy   Paris – France



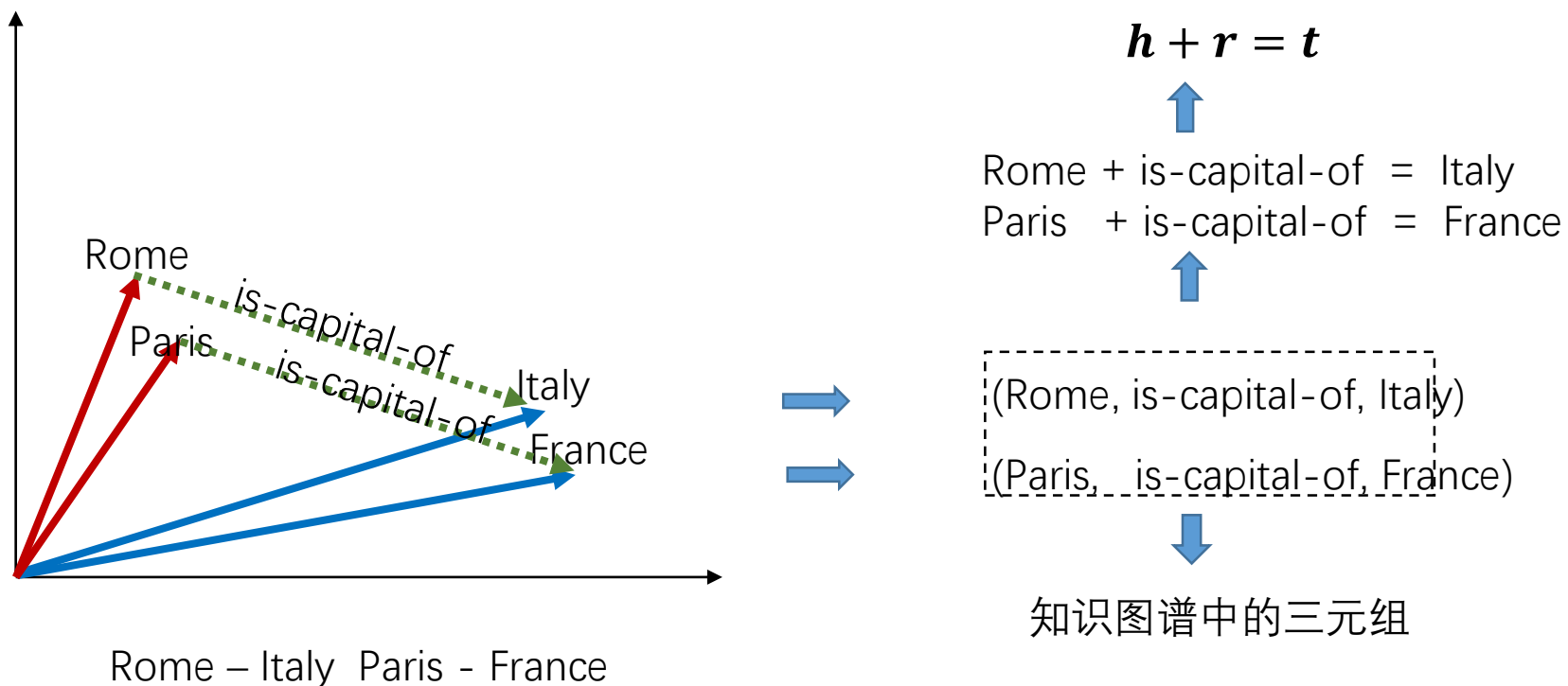
(Subject, Predicate, Object)



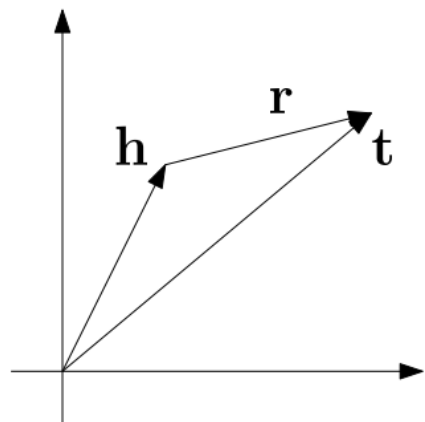
(Rome, is-capital-of, Italy)

(Paris, is-capital-of, France)

# 知识图谱嵌入模型：TransE



# 知识图谱嵌入模型：TransE



对每个三元组(h,r,t)的优化目标:

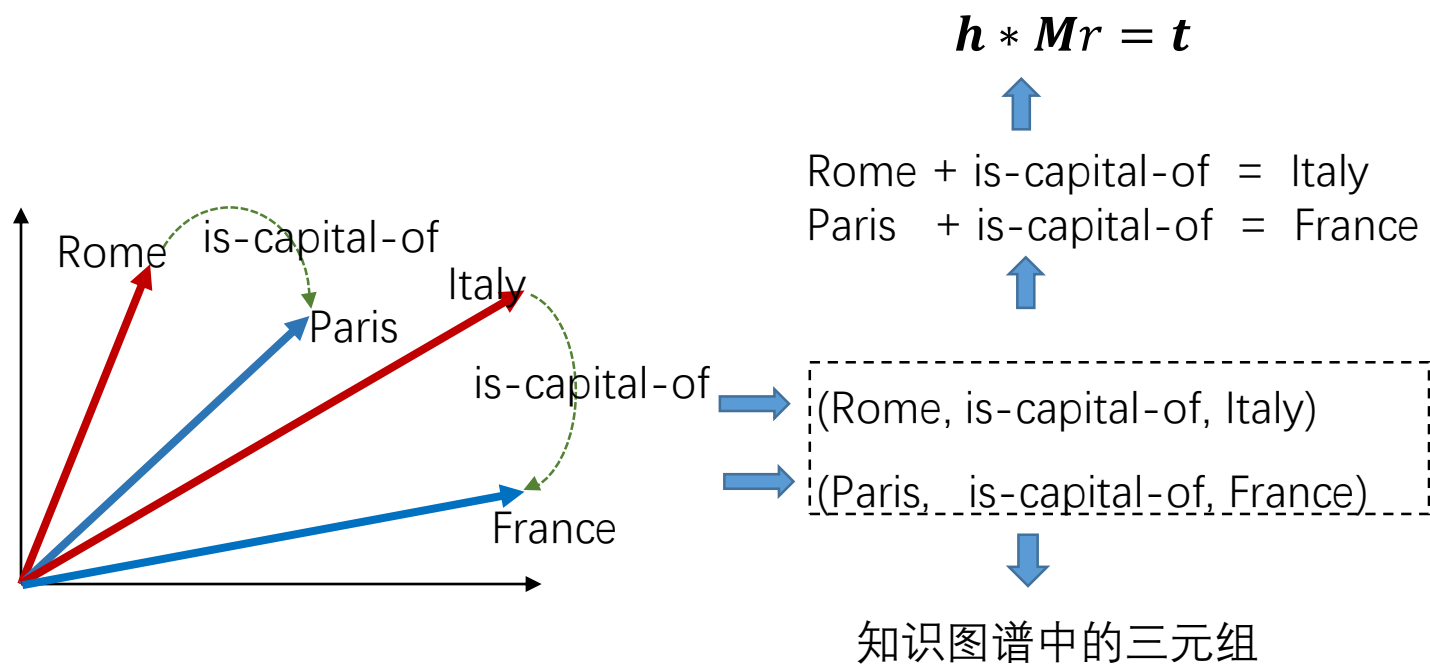
每个三元组的目标函数:  $f_r(h, t) = \|h + r - t\|_{L_1/L_2}$

$$L = \sum_{(h,r,t) \in S} \sum_{(h',r,t') \in S'} \max \left( 0, \underset{\substack{\uparrow \\ \text{(h,r,t)的score}}}{f_r(h, t)} + \underset{\substack{\uparrow \\ \text{Margin}}}{\gamma} - \underset{\substack{\uparrow \\ \text{(h,r,t) 的负样本的score}}}{f_r(h', t')} \right)$$

负样本的构造: 随机替换h或t



# 知识图谱嵌入模型：DistMult



$$\begin{pmatrix} \text{yellow} & \text{yellow} & \text{yellow} \end{pmatrix} \begin{pmatrix} \text{red} & \text{gray} & \text{gray} \\ \text{gray} & \text{red} & \text{gray} \\ \text{gray} & \text{gray} & \text{red} \end{pmatrix} \begin{pmatrix} \text{green} \\ \text{green} \\ \text{green} \end{pmatrix}$$

线性变换假设：

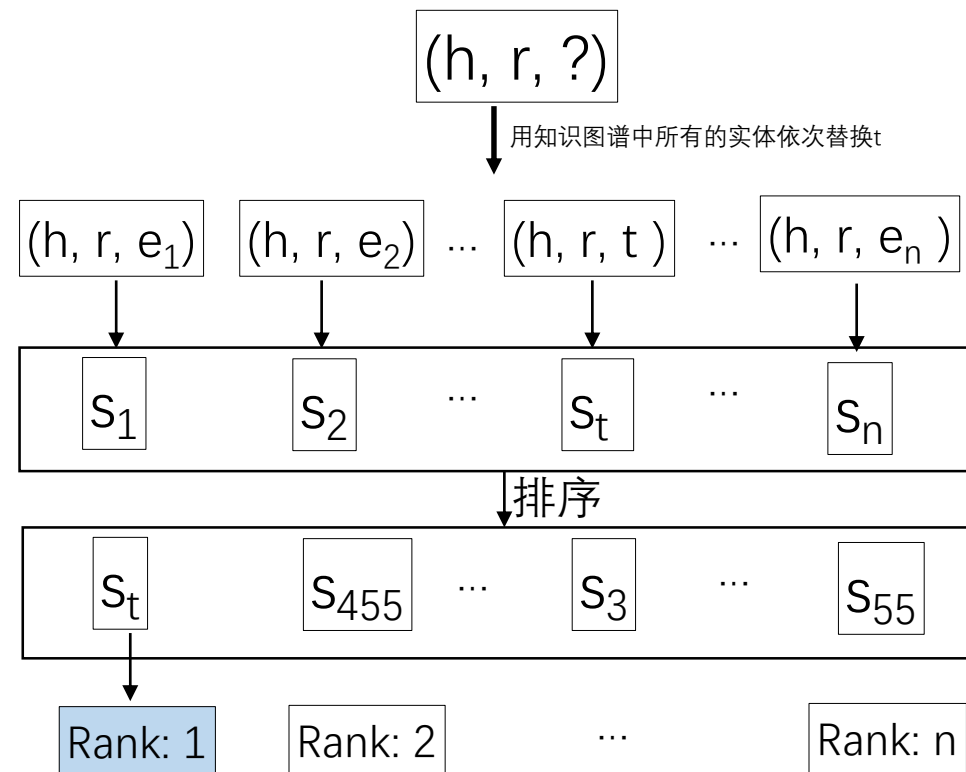
$$\mathbf{h} \mathbf{M}_r = \mathbf{t}$$

目标函数：

$$f_r(h, t) = \mathbf{h} \mathbf{M}_r \mathbf{t}$$

# 知识图谱嵌入模型：推理问题

- 测试三元组(  $h, r, t$  )
- 尾实体预测(  $h, r, ?$  )
- 头实体预测(  $?, r, t$  )



# 小结：玲琅满目的知识图谱嵌入模型

类别	建模								模型学习 (训练)				特征			
	模型	实体嵌入	关系嵌入	辅助因子	打分函数	约束条件	时间复杂度	空间复杂度	建模原理	假设	负例产生	L	优化求解	适用 KG 类型及原因	⊖	弊端及原因
平移模型	TransE	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$-\ h + r - t\ _{L2}$	$\ e\ =1, e=h, t$	$O(dn_e)$	$O((n_r + n_e)d)$	$\Delta$ 原理+Norm	CWA	unif	①	SGD	1:1	✓	多实体/关系重合( Norm)
	TransH	$e \in \mathcal{E}$	$r \in \mathcal{R}$	$w_r \in \mathcal{E}$	$-\left\  (h - w_r^T h w_r) + r - ((t - w_r^T t w_r)) \right\ _{L2}$	$\ o\  \leq 1, o=h, t, w_r$ $\ w_r^T r\  \ r\  \leq \varepsilon$	$O(2dn_e)$	$O(n_e d_e + 2n_r d_e)$	$\Delta$ 原理+ Norm +超平面转化	CWA	unif bern	①	SGD	1:1 多关系(关系映射)	✓	多实体重合( Norm)
	TransR	$e \in \mathcal{E}$	$r \in \mathcal{R}$	$M_r \in \mathcal{E}^{k \times d}$	$-\ M_r h + r - M_r t\ _{L2}$	$\ o\  \leq 1, o=h, t, r$ $\ M_r e\  \leq 1, e=h, t$	$O(2d_e d_r n_e)$	$O(n_e d_e + (d_r + 1)n_r d_e)$	$\Delta$ 原理+ Norm +普通线性变换(共享)	CWA	unif bern	①	SGD	1:1 多关系(关系映射)	✓	多实体重合( Norm) 复杂度高
	TransD	$e \in \mathcal{E}$	$r \in \mathcal{R}$	$\begin{matrix} h_r \in \mathcal{E}^k \\ t_r \in \mathcal{E}^k \\ r_r \in \mathcal{E}^k \end{matrix}$	$-\left\  (r_r \times h_r^T + I)h + r - (r_r \times t_r^T + I)t \right\ _{L2}$	$\ o\  \leq 1, o=h, t, r$ $\ (r_r \times e_r^T + I)e\  \leq 1, e=h, t$	$O(2d_r n_e)$	$O(2(n_e d_r + n_r d_r))$	$\Delta$ 原理+ Norm +低秩线性变换(独立)	CWA	unif bern	①	Ada delta	1:1 均匀多关系(关系映射) n:m (实体映射)	✓	实体密集型代价大(每个实体一个变换向量) 关系灵活性受限(低秩)
	Trans parse(s)	$e \in \mathcal{E}$	$r \in \mathcal{R}$	$M_r(\theta_r) \in \mathcal{E}^{k \times d}$	$-\ M_r(\theta_r)h + r - M_r(\theta_r)t\ _{L2}$	$\ o\  \leq 1, o=h, t, r$ $\ M_r(\theta_r)e\  \leq 1, e=h, t$	$O((1-\theta)d_r d_r n_e)$ ( $0 < \theta < 1$ )	$O(n_e d_r + (1-\theta) \times (d_r + 1)n_r d_r)$	$\Delta$ 原理+ Norm +稀疏线性变换(共享)	CWA	unif bern	①	SGD	1:1 低异构多关系(关系映射)	✓	显著异构(稀疏度) 扇出度较大
	Trans parse(d)	$e \in \mathcal{E}$	$r \in \mathcal{R}$	$\begin{matrix} M_{h_r}(\theta_r^h) \in \mathcal{E}^{k \times d} \\ M_{t_r}(\theta_r^t) \in \mathcal{E}^{k \times d} \end{matrix}$	$-\ M_{h_r}(\theta_r^h)h + r - M_{t_r}(\theta_r^t)t\ _{L2}$	$\ o\  \leq 1, o=h, t, r$ $\ M_{h_r}(\theta_r^h)e\  \leq 1, e=h, t$	$O(2(1-\theta)d_r d_r n_e)$ ( $0 < \theta < 1$ )	$O(n_e d_r + 2(1-\theta) \times (d_r + 1)n_r d_r)$	$\Delta$ 原理+ Norm +稀疏线性变换(独立)	CWA	unif bern	①	SGD	低异构多关系(关系映射) 1:1/扇出度较小(区分头尾)	✓	显著异构(稀疏度) 扇出度较大
组合模型	RESCAL	$e \in \mathcal{E}$	$M_r \in \mathcal{E}^{d \times d}$	——	$h^T M_r t$	$\ o\  \leq 1, o=h, t, r$ $\ M_r\ _F \leq 1$	$O(pq d_r (1+q) + q^2 (3d_r + q + pq))$	$O(n_e d + n_r d^2)$	普通线性组合+dot	OWA	unif	②	ALS	1:1 扇出度较小(dot)	✓	大规模、多关系 (关系建模为矩阵)
	LFM	$e \in \mathcal{E}$	$r_h, r_t \in \mathcal{R}$	——	$h^T \sum_i \pi_i^T u_i v_i^T t \approx (h^T r_h)(r_t^T t)$	$\ o\ =1, o=h, t, r_h, r_t$	$O((d^2 + d)n_e)$	$O(n_e d + n_r d^2)$	秩一线性组合+dot	CWA	unif	①	SGD	1:1/扇出度较小(dot)	✓	低秩约束
	DistMult	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$h^T \text{diag}(r) t = t^T \text{diag}(r) h$	$\ e\ =1, e=h, t$ $\ r\  \leq 1$	$O(2d_r n_e)$	$O(n_e d + n_r d^2)$	对角线性组合+dot	CWA	unif	③	Ada Grad	1:1/对称关系( diag ) 1:1/扇出度较小(dot)	✓	适用于对称关系
神经网络模型	HolE	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$r^T (h * t) = r^T \text{circ}(h) t = r^T M_r t$	$\ o\  \leq 1, o=h, t, r$	$O(d \log d n_e)$	$O((n_r + n_e)d)$	( holographic 映射) 循环线性组合+dot	CWA	unif	①	SGD	兼顾 RESCAL 和 DistMult 的优势	✓	复杂度高
	SME(l)	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$(M_1 h + M_2 r + b_1)^T (M_3 h + M_4 r + b_2)$	$\ e\ =1, e=h, t$	$O(4dn_e n_r)$	神经网络+线性+共享权重矩阵	CWA	unif	①	SGD	1:1/头尾对等	✗	参数复杂, 需要大量训练数据	
	SME(b)	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$(M_1 h \circ M_2 r + b_1)^T (M_3 h \circ M_4 r + b_2)$	$\ e\ =1, e=h, t$	$O(n_e d + n_r d + (4dn_e + 1)n_r)$		CWA	unif	①	SGD				
	NTN	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$r^T \tan(h^T \tilde{M}_1 t + M_2 h + M_3 t + b_1)$	$\ o\  \leq 1, o=h, t, r, b_r$ $\ \tilde{M}_1\ _F \leq 1$ $\ M_2\ _F \leq 1, e=h, t$	$O(((d^2 + d)n_e + (2d + 1)n_r)n_e)$	$O(n_e d + n_r n_e) \times (d^2 \times 2d + 2)$	双线性张量积	CWA	unif	①	LBF GS	兼顾 SLM 和 LFM]的优势	✗	每个关系需要 $O(d^2 k)$ 个参数
SLM	$e \in \mathcal{E}$	$r \in \mathcal{R}$	——	$r^T \tanh(M_2 h + M_3 t)$	$\ o\  \leq 1, o=h, t, r$ $\ M_2\ _F \leq 1, e=h, t$	$O((2dn_e + n_r)n_e)$	$O(n_e d + 2n_r n_e (1 + d))$	单层神经网络+线性+区分头尾	CWA	unif	①	SGD	1:1/头尾不对等	✗	实体、关系弱联系(单层神经网络的非线性运算)	

# 总结（1/2）

- 知识表示是传统符号人工智能研究的核心，知识表示的方法在早期语义网的发展过程主要用来为知识图谱的概念建模提供理论基础；
- 现实的知识图谱项目由于规模化构建的需要，常常降低表示的逻辑严格性，目前较为常见的知识图谱实践包括RDF图模型和属性图模型；
- 尽管很多知识图谱并没有应用复杂的知识表示框架，Schema工程对于知识图谱的构建仍然是基础性和必要性的工作，高质量的知识图谱构建通常从Schema设计开始。

# 总结（2/2）

- 在知识图谱的深度利用中，如复杂语义的表达、规则引擎的构建、推理的实现，会对更有丰富表达能力的知识表示方法有更多的需求。
- 图模型是更加接近于人脑认知和自然语言的数据模型，RDF作为一种知识图谱表示框架的参考标准，向上对接OWL等更丰富的语义表示和推理能力，向下对接简化后的属性图数据库以及图计算引擎，仍然是最值得重视的知识图谱表示框架。
- 知识（图谱）的表示学习是符号表示与神经网络相结合比较自然且有前景的方向。知识的向量表示有利于刻画那些隐含不明确的知识，同时基于神经网络和表示学习实现的推理一定程度上可以解决传统符号推理所面临的鲁棒性不高不容易扩展等众多问题。

[HTTP://OpenKG.CN](http://OpenKG.CN)



谢谢大家！