



计算机科学与技术学院
SCHOOL OF COMPUTER SCIENCE AND TECHNOLOGY

第二讲 知识表示的建模

院系：西南科技大学计算机学院

教师：李波（讲师）

Email: libo@swust.edu.cn



人工智能→知识工程

- 计算智能

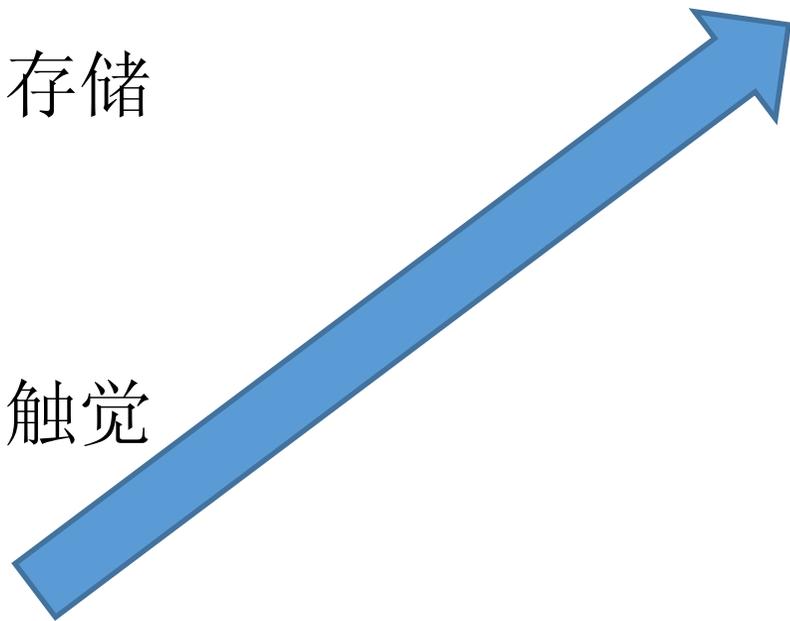
- 计算、记忆、存储

- 感知智能

- 视觉、听觉、触觉

- 认知智能

- 理解、解释



- 知识工程

- 知识表示
- 知识存储
- 知识获取
- 知识推理
- 知识应用

• 什么是知识？（西方）

- Justified（证明）：方法
- True（真理）：客观
- Belief（信仰）：主观

Knowledge is justified true belief.

——柏拉图《泰阿泰德篇》

• 什么是知识？（东方）

- 知物、知行、知言
- 人为中心（主观）
 - 道德修养
 - 人格境界

“格物致知”

——《大学》

“知行合一”

——王阳明

知识特性

- 相对正确性
 - 一定条件、某种环境.....
- 不确定性
 - 中间状态、真假程度、随机性、模糊性.....
- 可表示性
 - 语言、文字、图形、图像.....



- 能识别领域中概念、实体、属性、关系
- 能使用工具对领域知识建模



学习指南



本周（第二周）学习任务

2 知识表示的建模（1/3）

- 课前预习
 - 阅读教材：第2章 知识图谱的表示（2.1,2.2）
- 线上学习
 - 课程视频：第二章 知识图谱的表示
 - 第二章测试
- 课堂教学
 - 知识图谱的表示——概念和历史



下周（第三周）学习任务

2 知识表示的建模（2/3）

- 课前预习
 - 阅读教材：第2章 知识图谱的表示(2.3, 2.4)
- 线上学习（复习）
 - 课程视频：第二章 知识图谱的表示
- 课堂教学
 - 知识图谱的表示——方法
- 课后作业
 - 创建“寝室百事通”知识建模，准备ppt



第四周学习任务

2 知识表示的建模（3/3）

- 课后学习（复习）
 - 阅读教材：第2章 知识图谱的表示
- 线上学习（复习）
 - 课程视频：第二章 知识图谱的表示
- 课堂教学
 - 知识图谱的表示——案例
 - 实验一 知识表示与建模（4学时）
 - “寝室百事通”知识图谱口头报告



Contents

- 01 知识表示的概念
- 02 知识图谱的符号表示
- 03 知识图谱的向量表示
- 04 知识图谱的建模工具

1

知识表示的概念



1.1 什么是知识表示？

- 知识表示（Knowledge Representation）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 = Computational Model of Reality
- KR的五个用途：
 - 客观事物的**机器标识**
 - 一组本体约定和**概念模型**
 - 支持推理的**表示基础**
 - 用于高效计算的**数据结构**
 - 人可理解的**机器语言**

What Is a Knowledge Representation?

来自 ResearchGate | 喜欢 0 阅读量: 78

作者: R Davis, H Shrobe, P Szolovits

摘要: Although knowledge representation is one of the central and in some ways most familiar concepts in AI, the most fundamental question about it--What is it?--has rarely been answered directly. Numerous papers have lobbied for one or another variety of representation, other papers have argued for various properties a representation should have, while still others have focused on properties that are important to the notion of representation in general. In this paper we go back to basics to address the question directly. We believe that the answer can best be understood in terms of five important and distinctly different roles that a representation plays, each of which places different and at times conflicting demands on the properties a representation should have. We argue that keeping in mind all five of these roles provides a usefully broad perspective that sheds light on some longstanding disputes and can invigorate both research and practice in the field. 1 This report describes res...

关键词: HANDWRITTEN NUMERAL RECOGNITION HANDWRITTEN KANJI RECOGNITION LOCAL AFFINE TRANSFORMATION

DOI: doi:10.1016/0167-8655(93)90100-R

被引量: 1229

年份: 1993

来源期刊



引用走势

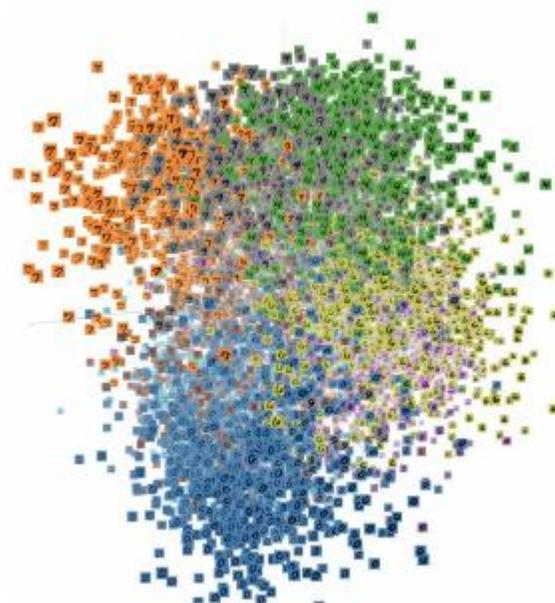
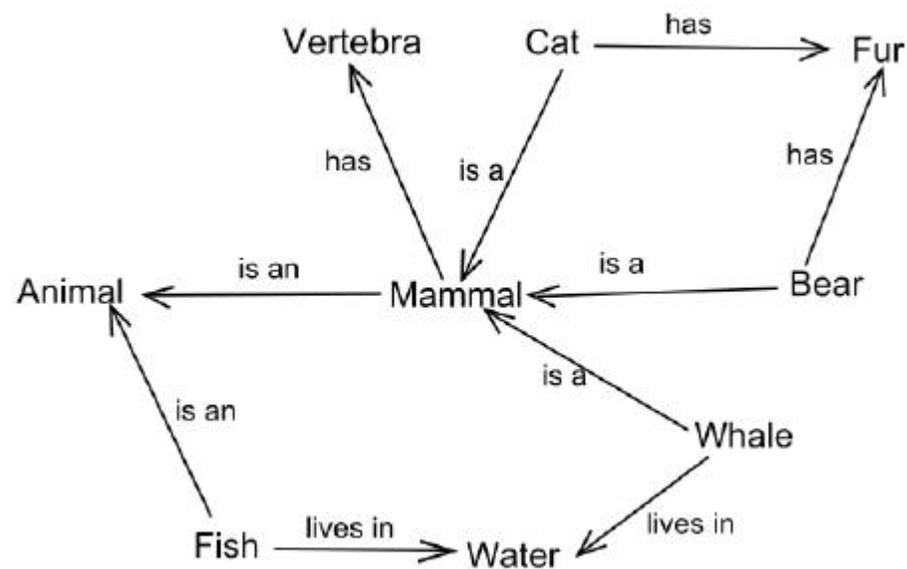
累加量 170 1993年被引量 10



Davis R , Shrobe H , Szolovits P . What Is a Knowledge Representation?[J]. Ai Magazine, 1993, 14(1).

1.2 知识表示方式

- 符号表示（离散、显式）
 - 谓词逻辑（Predicate Logic）
 - 语义网络（Semantic Net）
 - 产生式规则（Production Rule）
 - 框架系统（Framework）
- 向量（数值）表示（连续、隐式）
 - 离散表示
 - 分布式表示



自然语言：句子中的每个词

知识图谱：每个实体和关系

图像视频：视觉中的每个对象

1.2 知识表示方式

- 符号表示（离散、显式）
 - 谓词逻辑（Predicate Logic）

- 谓词
- 连接符
- 量词

举例：三段论逻辑

- 谓词

Human(x): x是人

Die(x): x会死

s: 苏格拉底

- 前提

$(\forall x) (\text{Human}(x) \rightarrow \text{Die}(x)), \text{Human}(s)$

- 结论

$(\forall x) (\text{Human}(x) \rightarrow \text{Die}(x)) \wedge \text{Human}(s) \Rightarrow \text{Die}(s)$

1.2 知识表示方式

- 符号表示（离散、显式）
 - 语义网络（Sementic Net）

基于人类联想记忆模型；

结构化的知识表示方法，显示表示事物属性以及事物间的各种语义联想。

举例：词典知识库WordNet

普林斯顿大学认知实验室开发；

用于词义消歧；

定义了名词、动词、形容词和副词之间的语义关系；



WordNet Search - 3.1
- [WordNet home page](#) - [Glossary](#) - [Help](#)

Word to search for:

Display Options:
Key: "S:" = Show Synset (semantic) relations, "W:" = Show Word (lexical) relations
Display options for sense: (gloss) "an example sentence"

Noun

- **S: (n) hamburger, beefburger, burger** (a sandwich consisting of a fried cake of minced beef served on a bun, often with other ingredients)
 - [direct hyponym](#) / [full hyponym](#)
 - **S: (n) cheeseburger** (a hamburger with melted cheese on it)
 - [part meronym](#)
 - **S: (n) ground beef, hamburger** (beef that has been ground)
 - [direct hypernym](#) / [inherited hypernym](#) / [sister term](#)
 - **S: (n) sandwich** (two (or more) slices of bread with a filling between them)
 - **S: (n) snack food** (food for light meals or for eating between meals)
 - **S: (n) dish** (a particular item of prepared food) "she prepared a special dish for dinner"



1.2 知识表示方式

- 符号表示（离散、显式）
 - 产生式规则（Production Rule）

$P \rightarrow Q$ 或

IF P THEN Q CF=[0,1]

其中：

P：前提，Q：结论

CF：置信度

举例：病情诊断专家系统

IF 病人发烧 AND 咳嗽

THEN 病人得了病毒性感冒，置信度为CF=0.9



1.2 知识表示方式

- 符号表示（离散、显式）
 - 框架系统（Framework）

人们对现实世界中事物的认识是以一种类似**框架的结构**存储在记忆中的，当人面临新的情景时，会从记忆中找到一个**合适的框架**，并根据实际情况对这一框架的细节进行**加工、修改和补充**，形成对新情景的认识并存入人脑中。

```
<框架名>  
<槽名 1>: : <侧面 11>: <值 111, 值 112, ..., 值 11k1>  
...  
    <侧面 1m>: <值 1m1, 值 1m2, ..., 值 1mkm>  
<槽名 2>: <侧面 21>: <值 211, 值 212, ..., 值 21k1>  
...  
    <侧面 2m>: <值 2m1, 值 2m2, ..., 值 2mkm>  
<约束>: <约束 1>  
...  
    <约束 m>
```

一般框架表示

```
框架名: <哲学家>  
类属: <人物>  
工作: 范围: (教学, 科研)  
      缺省: 哲学  
性别: (男, 女)  
类型: (<唯物主义哲学家>, <唯心主义哲学家>)
```

哲学家框架举例



1.2 知识表示方式

- 向量（数值）表示（连续、隐式）
 - 离散表示

问题：不能表示词语间的上下文语义关系

	Rome	Paris	Italy	France	...	Word k
Rome	1	0	0	0	0	0
Paris	0	1	0	0	0	0
Italy	0	0	1	0	0	0
France	0	0	0	1	0	0

k = 词典大小

One-hot词向量

	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

词袋向量

	i	want	to	eat	chinese	food	lunch	spend
i	0.002	0.33	0	0.0036	0	0	0	0.00079
want	0.0022	0	0.66	0.0011	0.0065	0.0065	0.0054	0.0011
to	0.00083	0	0.0017	0.28	0.00083	0	0.0025	0.087
eat	0	0	0.0027	0	0.021	0.0027	0.056	0
chinese	0.0063	0	0	0	0	0.52	0.0063	0
food	0.014	0	0.014	0	0.00092	0.0037	0	0
lunch	0.0059	0	0	0	0	0.0029	0	0
spend	0.0036	0	0.0036	0	0	0	0	0

2-gram向量

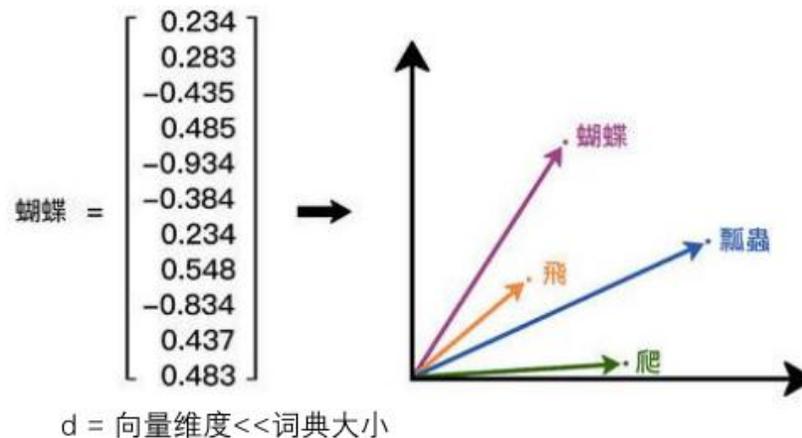
1.2 知识表示方式

- 向量（数值）表示（连续、隐式）
 - 分布式表示

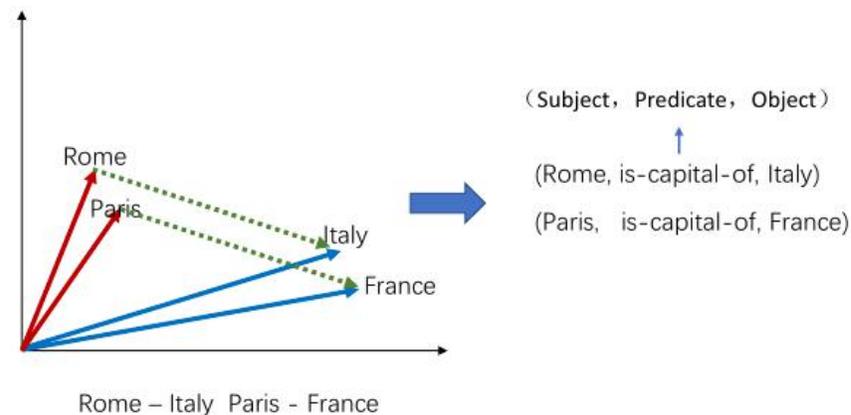
又叫嵌入（Embedding）

分布式假设：两个语义相似的词通常具有类似的上下文。

通过基于语言模型的知识表示学习来生成



词嵌入的例子



知识图谱嵌入的例子

2

知识图谱的符号表示



2.1 符号表示的建模流程



数据源

领域知识

.....

本体模型

UML模型

.....

三元组模型 (RDF, RDFS, OWL)

图模型 (属性图)

关系表模型

原生图模型

知识图谱符号表示的建模流程

概念

- 本体(ontology)是指领域概念体系的形式化规范定义
 - 领域概念体系：领域术语定义和术语间关系
 - 形式化规范定义
 - 概念或类
 - 实体或实例
 - 关系
 - 属性
- 知识库=本体+（大量）实例
- 知识图谱≈知识库

建模流程

- 1) 确定本体的领域和范围
- 2) 考虑重新使用现有本体
- 3) 列举重要术语
- 4) 定义类和类的层次结构
- 5) 定义类的属性和关系
- 6) 定义属性的限制
- 7) 创建实例

建模流程

- 1) **确定本体的领域和范围**
- 2) 考虑重新使用现有本体
- 3) 列举重要术语
- 4) 定义类和类的层次结构
- 5) 定义类的属性和关系
- 6) 定义属性的限制
- 7) 创建实例

- 本体涉及的领域是什么？

古希腊哲学家领域

- 如何使用本体？

用于搜索三大古希腊哲学家信息的应用

- 本体表达的信息应解答什么样的问题？

三大古希腊哲学家的相关信息和相互关系

- 谁将使用和维护本体？

使用者：互联网用户；维护者：软件开发公司

建模流程

- 1) 确定本体的领域和范围
- 2) **考虑重新使用现有本体**
- 3) 列举重要术语
- 4) 定义类和类的层次结构
- 5) 定义类的属性和关系
- 6) 定义属性的限制
- 7) 创建实例

- 改进或扩展现有本体资源

- 本体库

DAML: <http://www.daml.org/ontologies/>

schema: <https://schema.org/>

cnSchema: <http://cnschema-vocabulary.openkg.cn/>

.....

- 商业本体

unspsc: <https://www.unspsc.org/>

.....

建模流程

- 1) 确定本体的领域和范围
- 2) 考虑重新使用现有本体
- 3) **列举重要术语**
- 4) 定义类和类的层次结构
- 5) 定义类的属性和关系
- 6) 定义属性的限制
- 7) 创建实例

- 列出相关领域的所有术语

术语在线: <https://www.termonline.cn/>

东奥术语平台: <http://owgt.blcu.edu.cn/>

联合国粮农组织术语库: <https://agrovoc.fao.org/>

.....

- 从数据中获取

建模流程

- 1) 确定本体的领域和范围
- 2) 考虑重新使用现有本体
- 3) 列举重要术语
- 4) **定义类和类的层次结构**
- 5) 定义类的属性和关系
- 6) 定义属性的限制
- 7) 创建实例

- 自顶向下（依赖领域专家）
父类->子类
- 自底向上（基于数据）
子类->父类

建模流程

- 1) 确定本体的领域和范围
 - 2) 考虑重新使用现有本体
 - 3) 列举重要术语
 - 4) 定义类和类的层次结构
 - 5) **定义类的属性和关系**
 - 6) 定义属性的限制
 - 7) 创建实例
- 属性
 - 内在属性
 - 外在属性
 - 关系
 - 类之间

建模流程

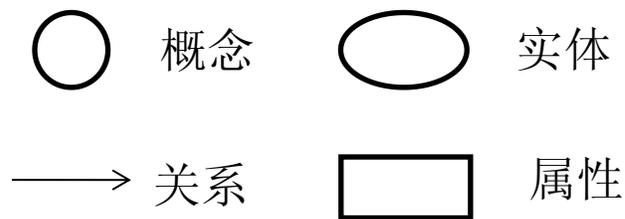
- 1) 确定本体的领域和范围
 - 2) 考虑重新使用现有本体
 - 3) 列举重要术语
 - 4) 定义类和类的层次结构
 - 5) 定义类的属性和关系
 - 6) **定义属性的限制**
 - 7) 创建实例
- 基数
 - 值类型
 - 定义域和值域

建模流程

- 1) 确定本体的领域和范围
 - 2) 考虑重新使用现有本体
 - 3) 列举重要术语
 - 4) 定义类和类的层次结构
 - 5) 定义类的属性和关系
 - 6) 定义属性的限制
 - 7) **创建实例**
- 选择类
 - 创建实例
 - 填充属性

• 课堂练习

从以下文本中构建金山寺知识图谱



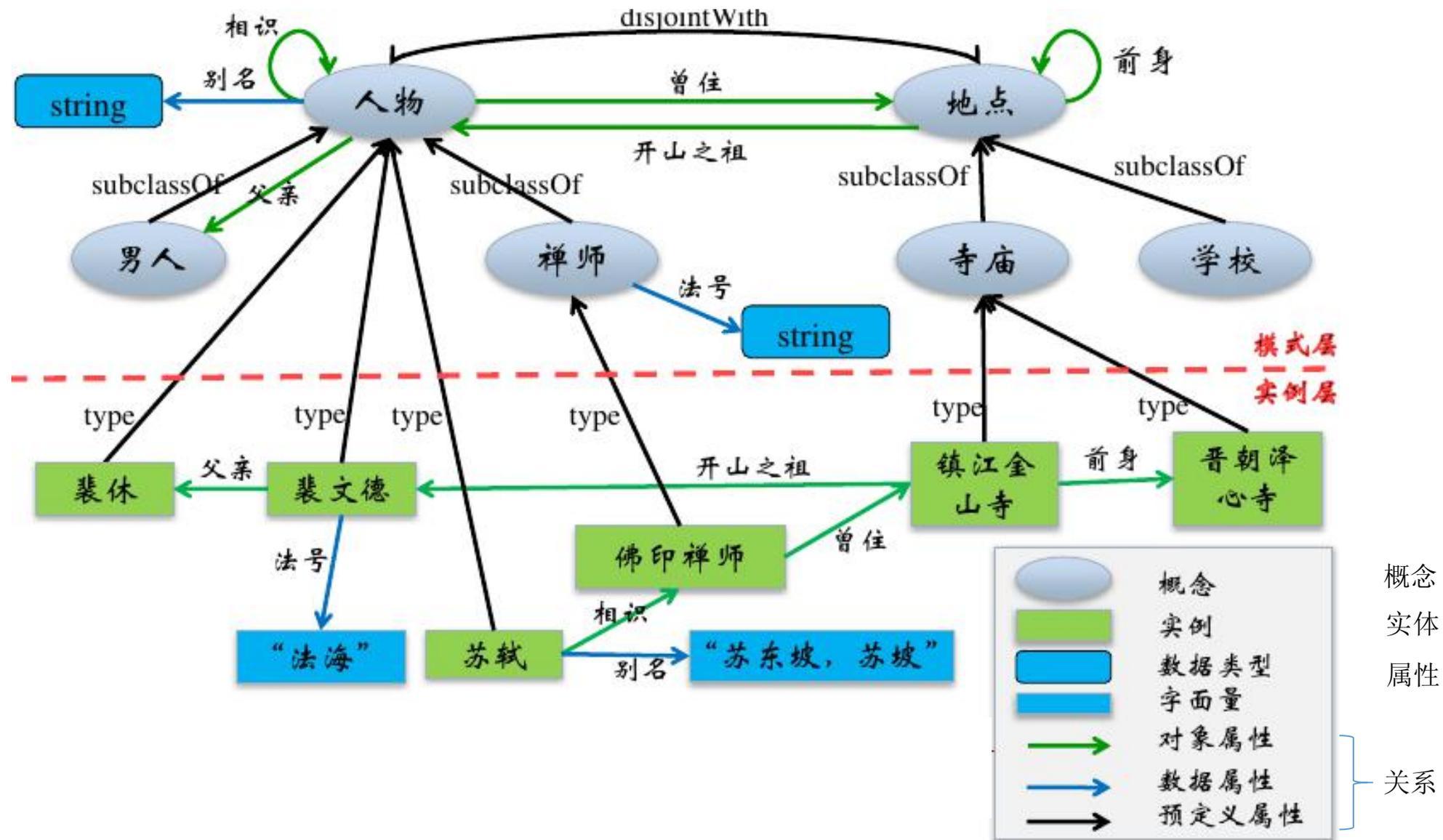
金山寺，又名江天禅寺，始建于东晋，原名泽心寺，自唐代起，法海禅师重修，唐宣宗赐名为金山寺。

金山寺是禅宗的寺院，历来龙象辈出，唐朝有著名的法海和尚主持，宋朝有佛印禅师和苏东坡交往的典故。

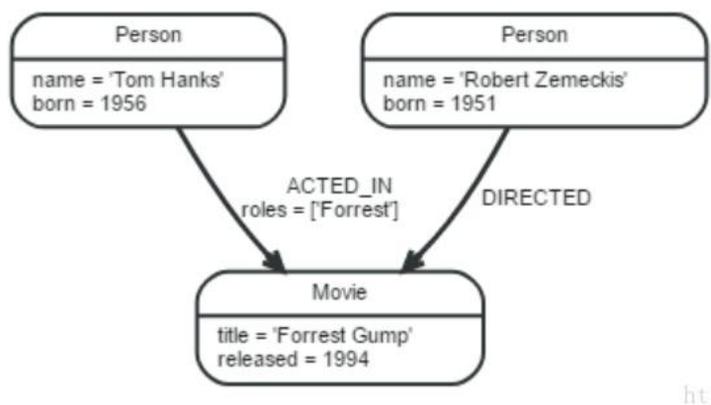
不过如今世人对金山寺的了解，无非是白娘子水漫金山寺。其实法海和尚是一位真正的高僧大德。我们稍微介绍一下法海禅师：

法海禅师是唐代名相裴休之子，俗名裴文德。唐宣宗大中三年，时任宰相的裴休捐建密印寺，时皇子得恶疾，看尽名医均不奏效，裴休送自己的儿子代皇子出家以消业障之疾，密印寺主持汾山灵佑禅师亲赐法号“法海”，人称法海禅师。

2.2 本体建模



2.3 符号表示模型



图模型

(Tom Hanks, born, 1956)
(Robert Zemeckis, born, 1951)
(Tom Hanks, roles, Forrest)
(Tom Hanks, acted_in, Forrest Gump)
(Robertz Zemeckis, directed, Forrest Gump)
(Forrest Gump, released, 1994)

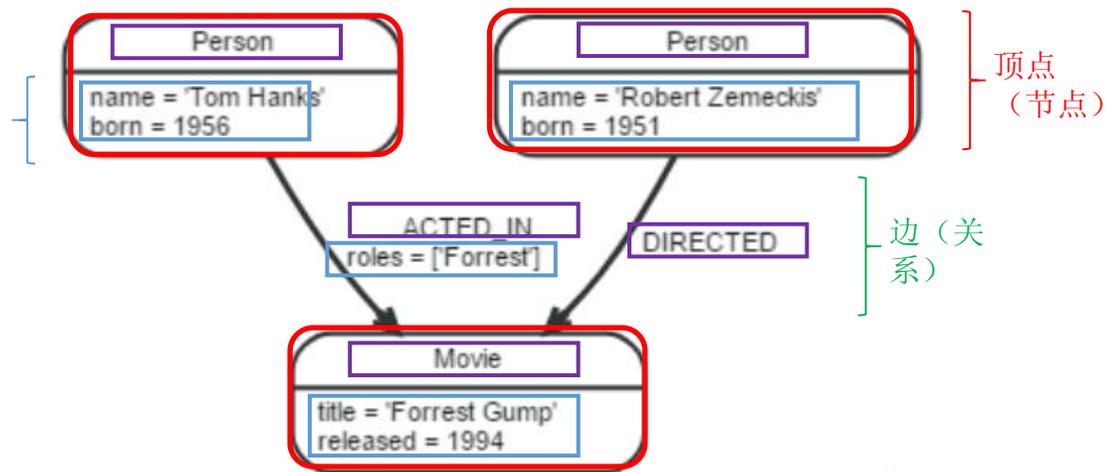
三元组模型



图模型标准：属性图

- 属性图（Property Graph）：图数据库 **Neo4J** 的 **图结构表示模型**，工业界使用广泛
- 优点
 - 表达方式灵活
 - 图结构优化，查询效率高
- 缺点
 - 缺乏工业标准支持
 - 不关注深层语义表达

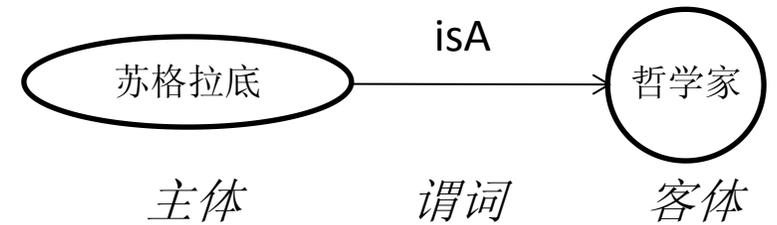
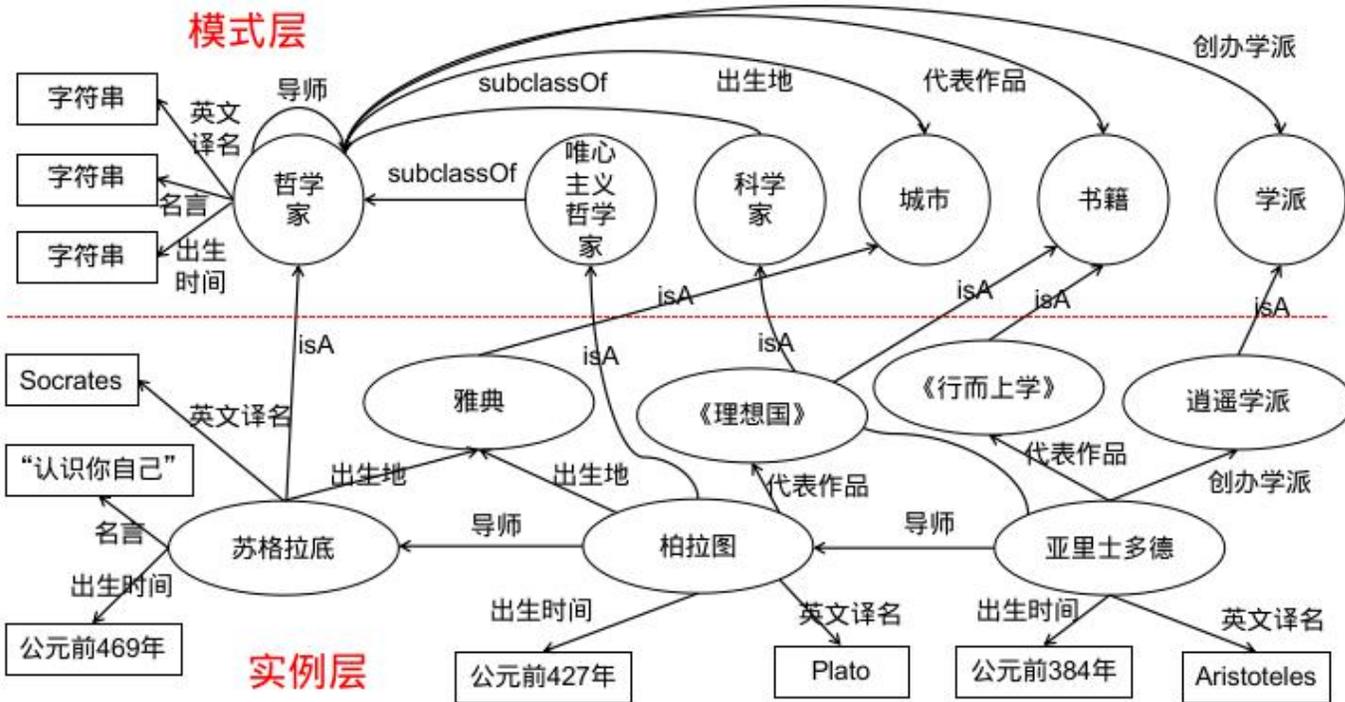
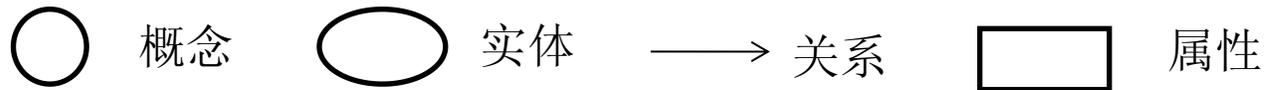
标签
属性



htt

2.3 符号表示模型

- 三元组模型: Subject Predicate Object (主体-谓词-客体 或 主语-谓语-宾语)



(苏格拉底, isA, 哲学家)

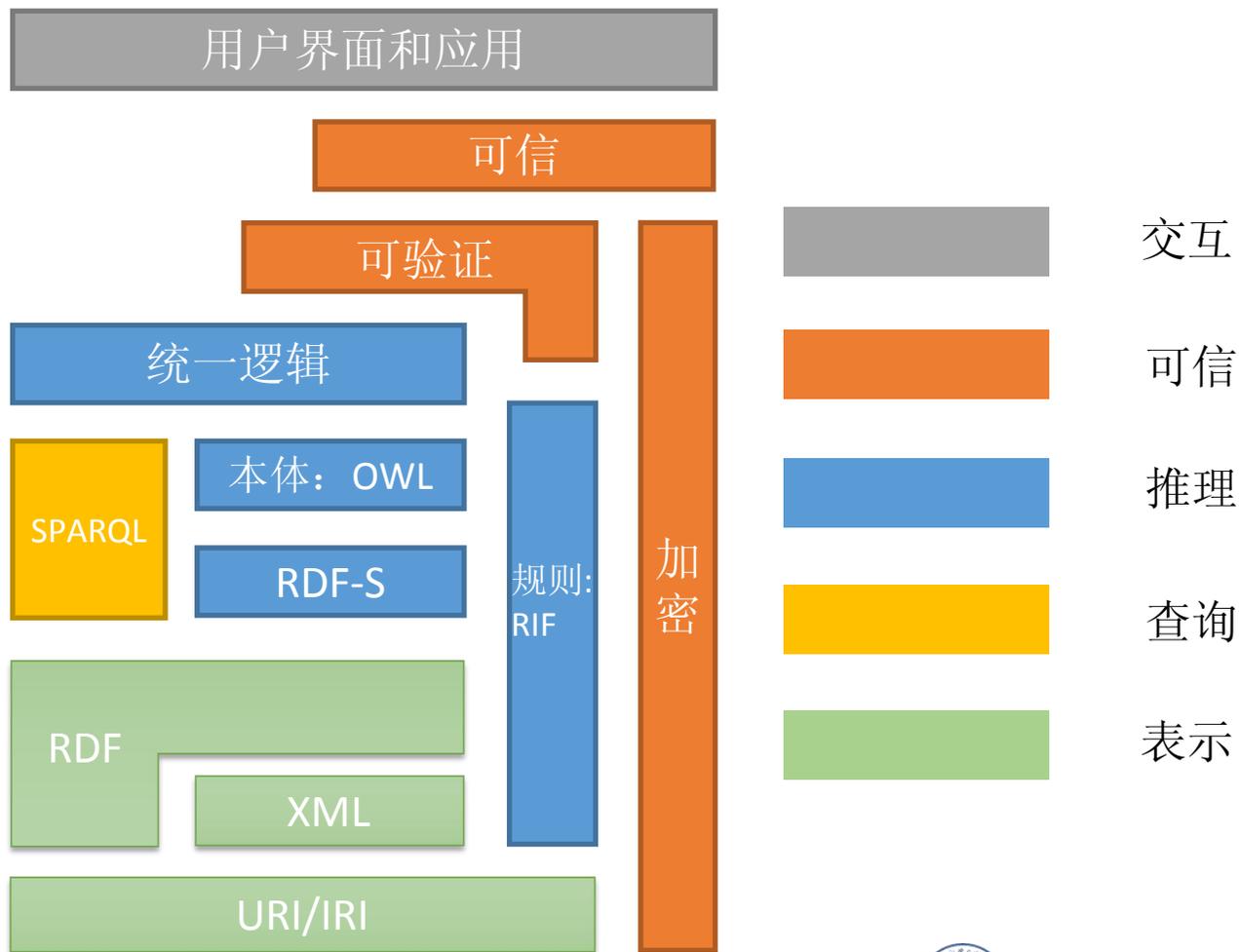
三元组模型

关于古希腊三大哲学家的知识图谱片段

三元组模型标准: **RDF**

- RDF(Resource Description Framework): W3C语义网标准中的知识表示标准
 - **R**esource: 页面、视频、图片等任何具有IRI标识符 (<http://xx.xxx.xx/xxx>)
 - **D**escription: 属性、特征和资源之间的关系
 - **F**ramework: 模型、语言和这些描述的语法

W3C语义网标准框架



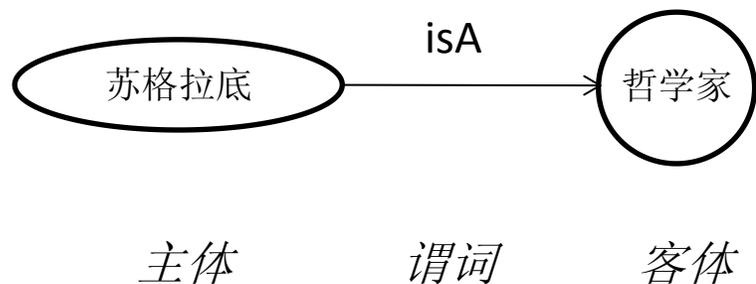
三元组模型标准: **RDF**

- **Resource**: RDF中的资源由IRI标识 (前缀)

(<http://www.perceive.net/schemas/relationship/Sokrates>,
<http://www.w3.org/2001/XMLSchema#isA>,
<http://www.perceive.net/schemas/relationship/philosopher>)

举例: 苏格拉底是哲学家

(苏格拉底, isA, 哲学家)



Prefix	IRI
xsd:	http://www.w3.org/2001/XMLSchema#
somePrefix:	http://www.perceive.net/schemas/relationship

前缀

RDF

(somePrefix:Sokrates, xsd:isA, somePrefix:philosopher)

三元组模型标准: *RDF*

- **Description:** RDF定义的部分词汇

- 实例化词汇

rdf:Statement rdf:subject rdf:predicate rdf:object

- 容器词汇

rdf:List rdf:first rdf:rest rdf:nil

- 其他

rdf:type rdf:langString rdf:Property

举例:

(村上春树 rdf:type 作家) 用来表示村上春树是作家的一个实例。

三元组模型标准: *RDF*

- **Description:** RDF定义的部分词汇

- 实例化词汇

rdf:Statement rdf:subject rdf:predicate rdf:object

- 容器词汇

rdf:List rdf:first rdf:rest rdf:nil

- 其他

rdf:type rdf:langString rdf:Property

举例:

(村上春树 rdf:type 作家) 用来表示村上春树是作家的一个实例。

三元组模型标准: **RDF**

- RDF是数据模型（三元组），不是存储格式（XML、Turtle、N-Triples）

```
<rdf:RDF
  xmlns:ex-schema=http://ex.org/schema#>
  <rdf:Description rdf:about="http://ex.org/ccf_adl">
    <ex-schema:speaker rdf:resource="http://ex.org/haofen"/>
    <ex-schema:theme rdf:resource="http://ex.org/KG"/>
  </rdf:Description>
</rdf:RDF>
```

XML格式 (.xml)

```
@prefix ex: <http://ex.org/> .
@prefix ex-schema: <http://ex.org/schema#>
ex:ccf_adl
ex-schema:speaker ex:haofen;
ex-schema:theme ex:KG.
```

Turtle格式 (.ttl)

```
< http://ex.org/ccf_adl>
  <http://ex.org/schema#speaker>
  <http://ex.org/haofen> .

< http://ex.org/ccf_adl>
  <http://ex.org/schema#theme>
  <http://ex.org/KG> .
```

N-Triples格式 (.nt)

三元组模型标准: **RDFS**

- RDFS(RDF Schema): 在RDF上扩展了术语、概念等的定义方式。
- 用途
 - 元数据交换
 - 推理

rdfs:domain rdfs:range
rdfs:Resource rdfs:Literal
rdfs:Datatype rdfs:Class
rdfs:subClassOf
rdfs:subPropertyOf
rdfs:member rdfs:Container
rdfs:comment rdfs:seeAlso

属性	定义
Contributor	一个负责为资源内容做出贡献的实体(如作者)。
Coverage	资源内容的氛围或作用域
Creator	一个主要负责创建资源内容的实体。
Format	物理或数字的资源表现形式。
Date	在资源生命周期中某事件的日期。
Description	对资源内容的说明。
Identifier	一个对在给定上下文中的资源的确切引用
Language	资源智力内容所用的语言。
Publisher	一个负责使得资源内容可用的实体
Relation	一个对某个相关资源的引用
Rights	有关保留在资源之内和之上的权利的信息
Source	一个对作为目前资源的来源的资源引用。
Subject	一个资源内容的主题
Title	一个给资源起的名称
Type	资源内容的种类或类型。

http://blog.csdn.net/qq_21460525

元数据交换举例:

<http://ex.org/Sokrates>

<rdfs: subClassOf>

<http://purlorg/dc/elements/1.0/Creator>.

RDFS定义的部分词汇

(完整见: <https://www.w3.org/TR/rdf-schema/>)

Dublin Core元数据

用户自定义元数据Sokrates是Dublin Core元数据Creator的子类

三元组模型标准: **RDFS**

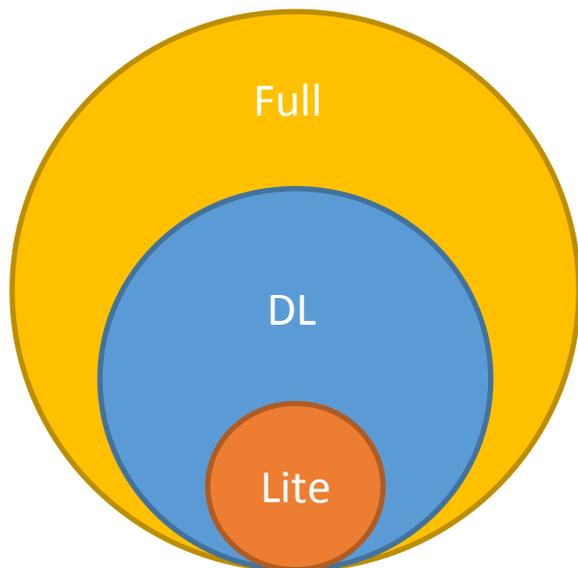
- RDFS(RDF Schema): 在RDF上扩展了术语、概念等的定义方式。
- 用途
 - 元数据交换
 - **推理**



推理举例

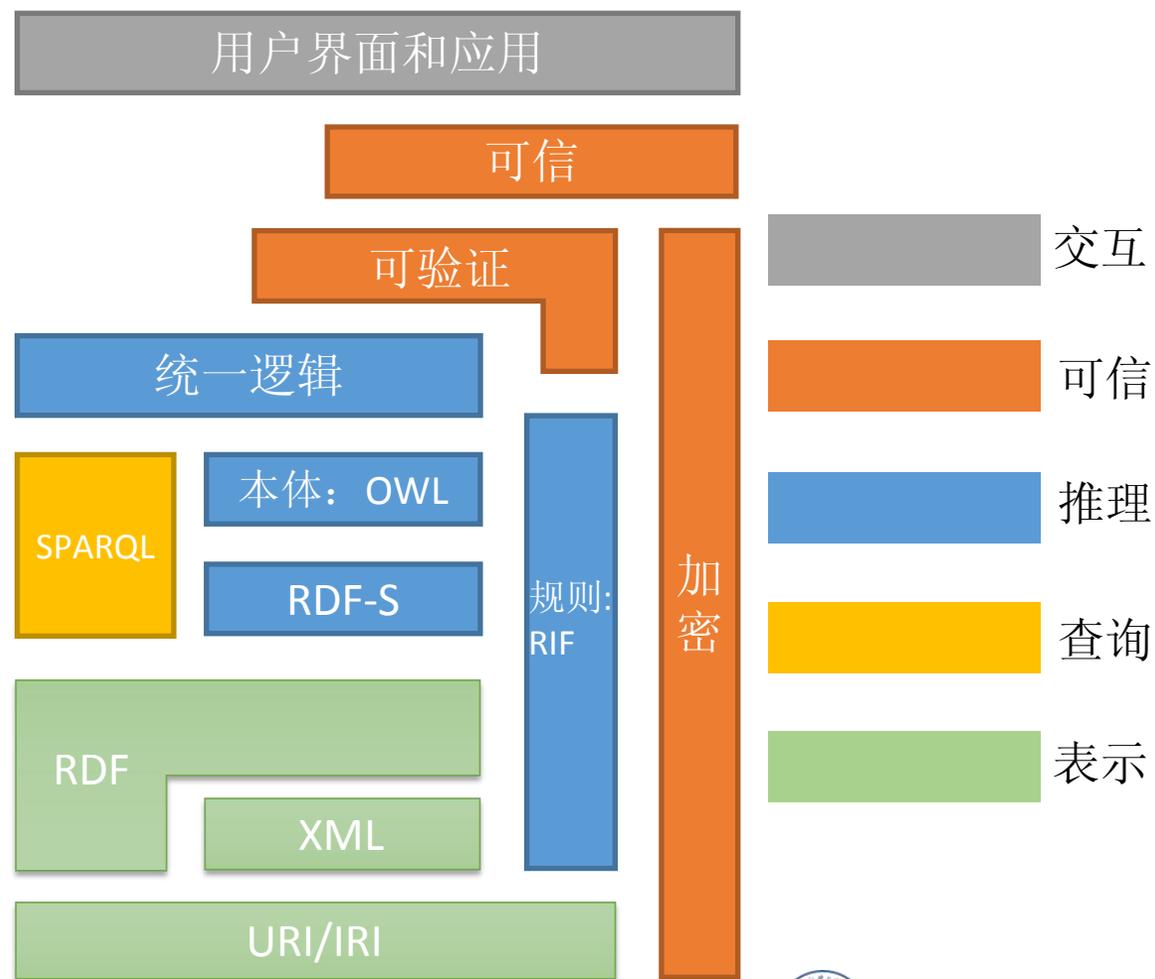
三元组模型标准: OWL

- RDF和RDFS语义表达能力上有缺陷
- OWL扩展了RDFS



OWL的三个子语言

W3C语义网标准框架



2.3 符号表示模型

- OWL的词汇（表达构件）

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

等价申明

exp:hasMother rdf:type owl:FunctionalProperty

属性的函数性

exp:friend rdf:type owl:SymmetricProperty

属性的对称性

exp:ancestor rdf:type owl:TransitiveProperty

属性传递

exp:ancestor owl:inverseOf exp:descendant

属性互反

exp:Person owl:allValuesFrom exp:Women
exp:Person owl:onProperty exp:hasMother

属性的局部约束:全称限定



- OWL的词汇

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

属性的局部约束:存在限定

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

属性的局部约束:基数限定

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

相交的类



3

知识图谱的向量表示



3.1 词的离散向量表示

问题：不能表示词语间的上下文语义关系



词的语义由它被使用的上下文确定

“徒”的语义

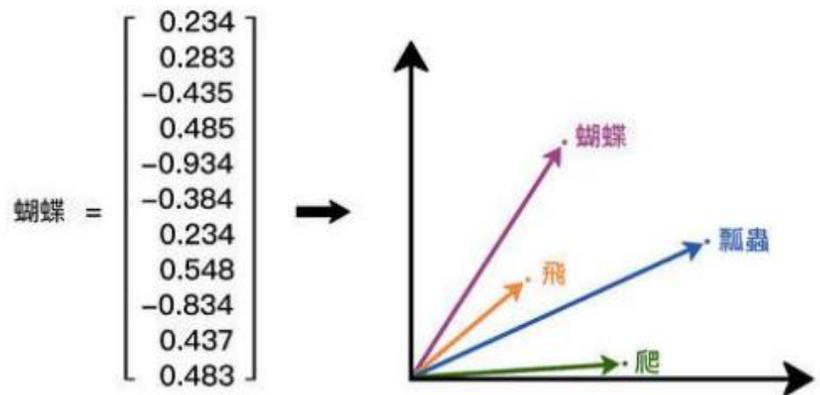
Rome = [1, 0, 0, 0, 0, 0, ..., 0]
Paris = [0, 1, 0, 0, 0, 0, ..., 0]
Italy = [0, 0, 1, 0, 0, 0, ..., 0]
France = [0, 0, 0, 1, 0, 0, ..., 0]

任意两两之间相似度为0

One-hot Encoding

3.2 词的分布式向量表示

如何产生分布式向量表示？



d = 向量维度 << 词典大小

词嵌入的例子

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]

词嵌入的语义相似性1

“蝴蝶”和“飞”比较接近

“瓢虫”和“爬”比较接近

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]

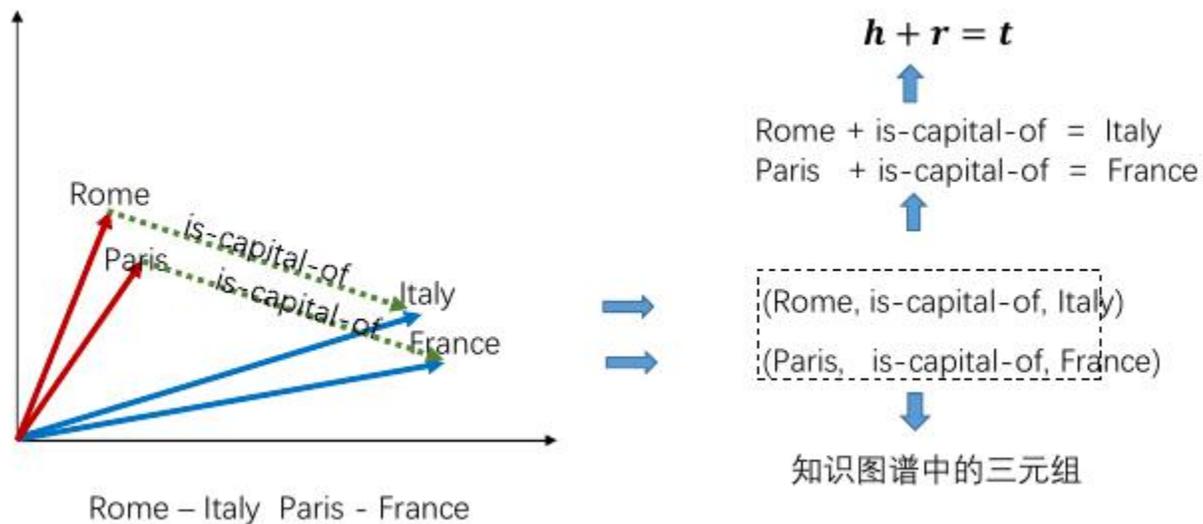
France = [0.33, 0.78, 0.66, ..., 0.97]

词嵌入的语义相似性2

3.3 向量表示学习模型

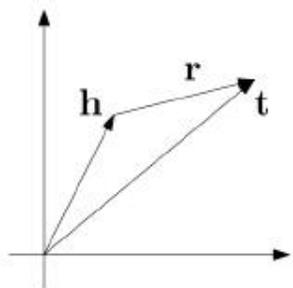
TransE模型

- 动机（假设）：
 - 如果三元组 $\langle h, r, t \rangle$ 代表的事实客观存在，那么一定满足 $h+r=t$
 - 其中， h 是主语向量， r 是关系谓词向量， t 是客体或宾语向量。



TransE模型

- 求解
 - 输入：测试集（三元组(h,r,t)集合）
 - 求解：有监督机器学习（最优化问题）
 - 输出：(h,r,t)的向量表达



对每个三元组(h,r,t)的优化目标:

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

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

负样本的构造：随机替换h或t

Algorithm 1 Learning TransE

input Training set $S = \{(h, \ell, t)\}$, entities and rel. sets E and L , margin γ , embeddings dim. k .

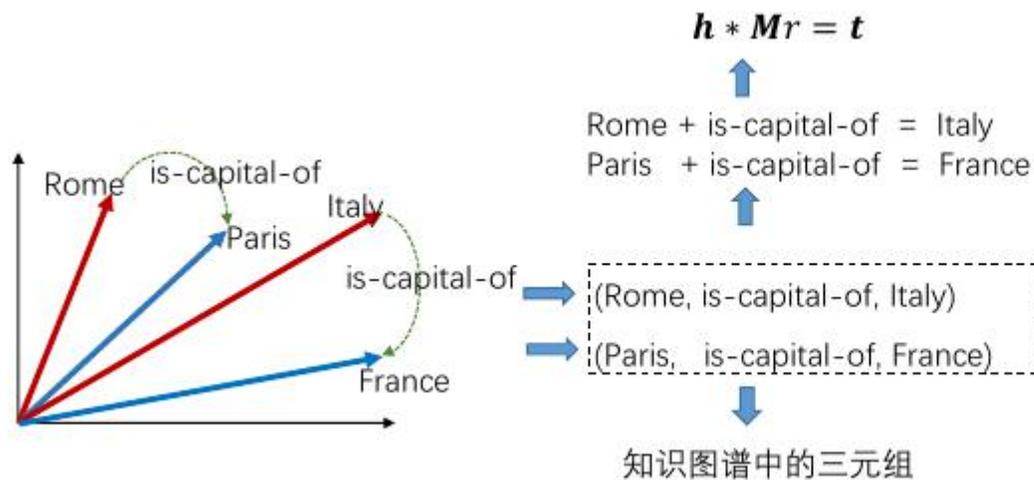
- 1: **initialize** $\ell \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each $\ell \in L$
- 2: $\ell \leftarrow \ell / \|\ell\|$ for each $\ell \in L$
- 3: $e \leftarrow \text{uniform}(-\frac{6}{\sqrt{k}}, \frac{6}{\sqrt{k}})$ for each entity $e \in E$
- 4: **loop**
- 5: $e \leftarrow e / \|e\|$ for each entity $e \in E$
- 6: $S_{batch} \leftarrow \text{sample}(S, b)$ // sample a minibatch of size b
- 7: $T_{batch} \leftarrow \emptyset$ // initialize the set of pairs of triplets
- 8: **for** $(h, \ell, t) \in S_{batch}$ **do**
- 9: $(h', \ell, t') \leftarrow \text{sample}(S'_{(h,\ell,t)})$ // sample a corrupted triplet
- 10: $T_{batch} \leftarrow T_{batch} \cup \{(h, \ell, t), (h', \ell, t')\}$
- 11: **end for**
- 12: Update embeddings w.r.t. $\sum_{((h,\ell,t),(h',\ell,t')) \in T_{batch}} \nabla[\gamma + d(\mathbf{h} + \ell, \mathbf{t}) - d(\mathbf{h}' + \ell, \mathbf{t}')]_+$
- 13: **end loop**



DisMult模型

求解过程同TransE模型

- 动机（假设）：
 - 如果三元组 $\langle h, r, t \rangle$ 代表的事实客观存在，那么一定满足 $h * M_r = t$
 - 其中， h 是主语向量， r 是关系谓词**矩阵**， t 是客体或宾语向量。



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

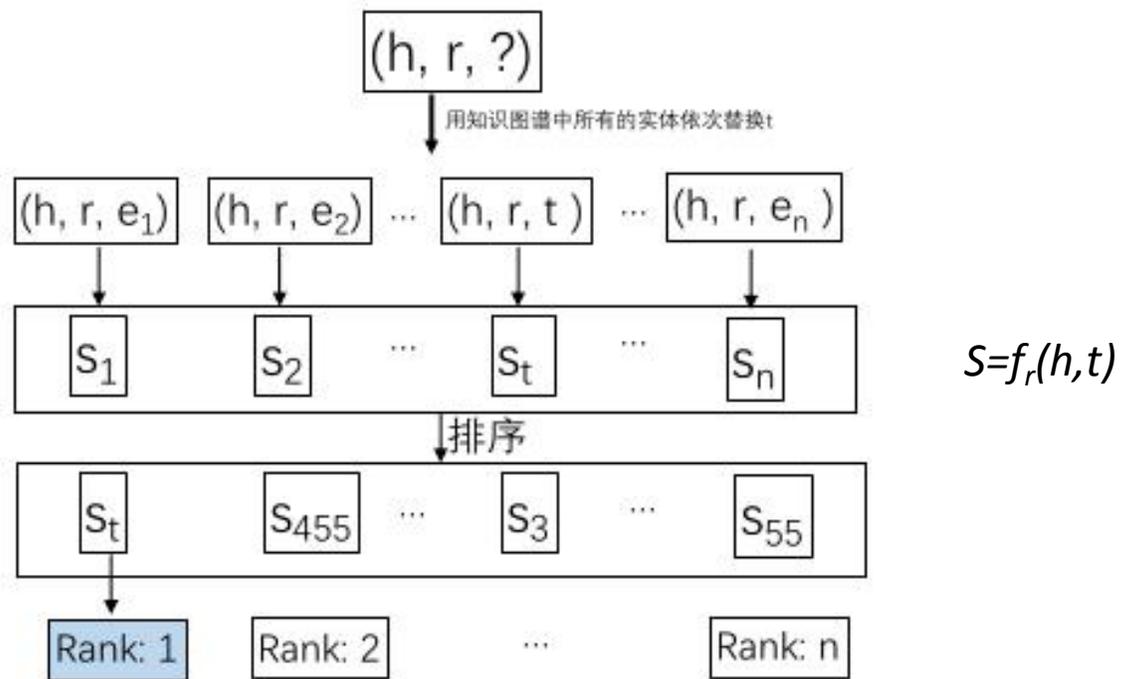
线性变换假设：

$$h M_r = t$$

目标函数：

$$f_r(h, t) = h M_r t$$

向量表示的评估

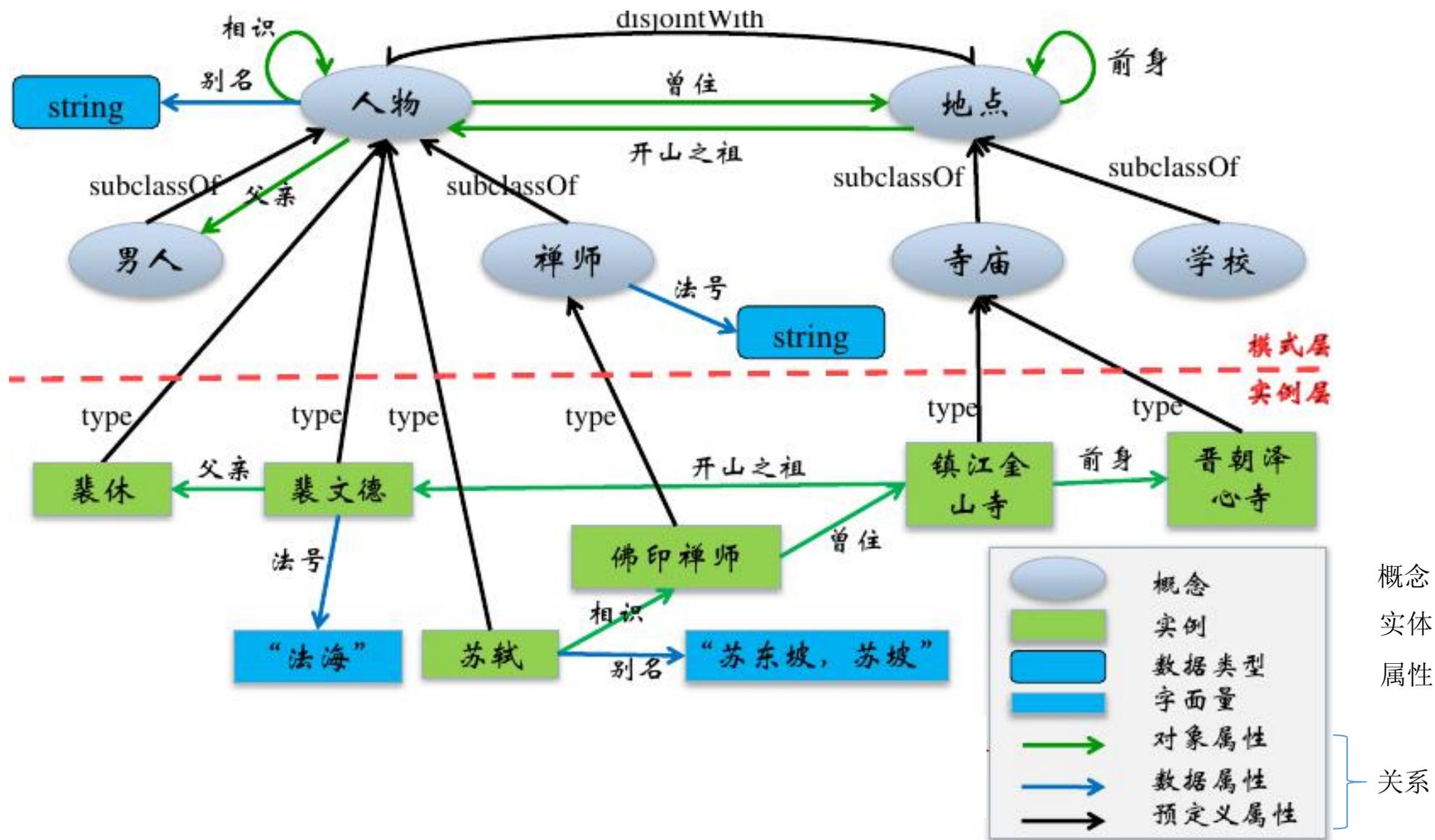


4

知识图谱的建模工具



4.1 知识建模举例



4.2 Protégé简介

- Protégé 软件是斯坦福大学医学院生物信息研究中心基于Java语言开发的**本体编辑和本体开发工具**,也是基于知识的编辑器,属于开放源代码软件。
- 这个软件主要用于语义网中本体的构建,是语义网中本体构建的核心开发工具,现在的最新版本为5.5.0版本

*Protégé*特点

- Protégé 是一组自由开源的工具软件,用于构建域模型与基于知识的本体化应用程序。
- Protégé 提供了大量的知识模型架构与动作,用于创建、可视化、操纵各种表现形式的本体。
- 可以通过用户定制实现域-友好(领域相关)的支持,用于创建知识模型并填充数据。
- Protégé 可以通过两种方式进行扩展:插件和基于java的API。
- 相比与其他的本体构建工具而言,Protégé 最大的好处在于支持中文,在插件上,用OntoGraf可实现中文关系的显示。



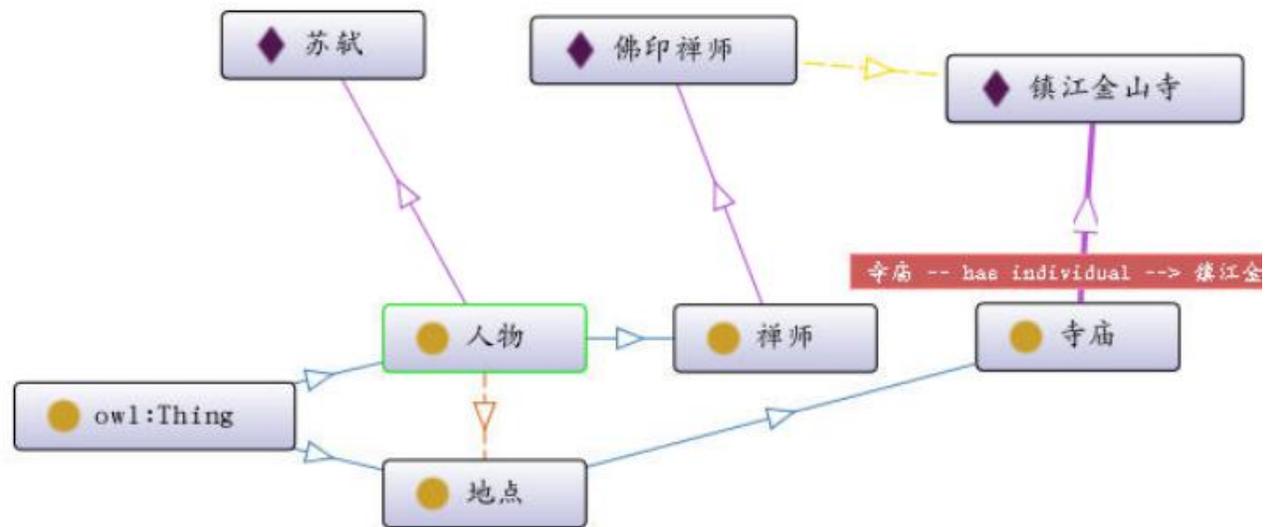
*Protégé*用途

- **类建模** (Class modeling): Protégé 提供了一个图形化用户界面来建模类(领域概念)和它们的属性及关系。
- **实例编辑** (Instance editing):从这些类中, Protégé 自动产生交互式的形式,使用户或领域专家进行有效实例编辑成为可能。
- **模型处理** (Model processing):Protégé 有一些插件库,可以定义语义、解答询问以及定义逻辑行为。
- **模型交换** (Model exchange):最终的模型 (类和实例)能以各种各样的格式被装载和保存,包括XML、UML和资源描述框架RDF。



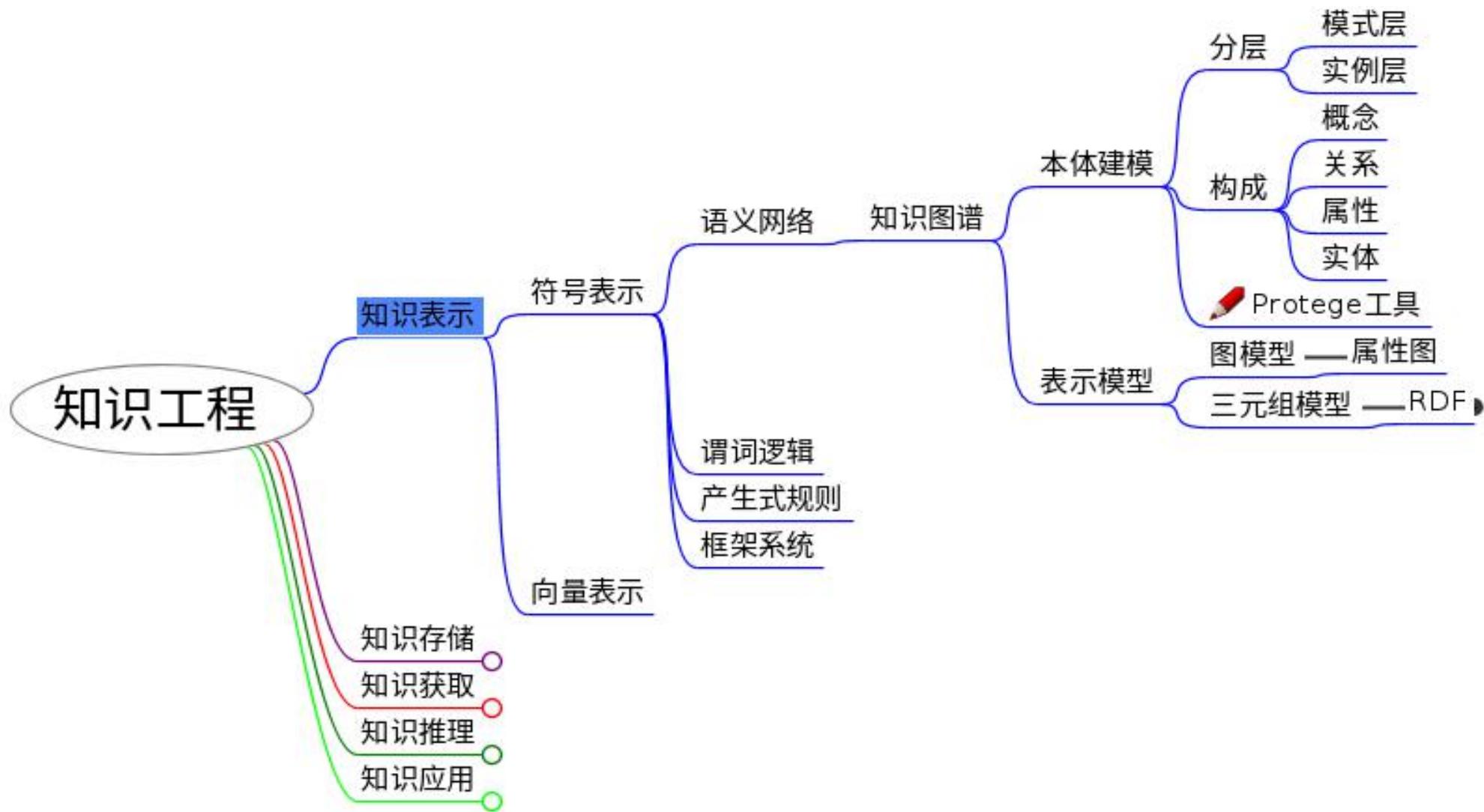
4.3 创建流程

- 创建新的本体
- 构建子类（人物、地点、人物->禅师、地点->寺庙）
- 构建类之间的关系（人物 owl:disjointWith 地点）
- 建立对象属性（曾住）
- 建立数据属性（法号 xsd:string）
- 建立实例（佛印禅师、镇江金山寺）
- 建立实例之间的关系（佛印禅师 曾住 镇江金山寺）
- 保存本体
- 可视化（OntoGraf）
- 推理（Hermit, Start reasoner, Description, 解释）



小结





THANKS !

