

# 人工智能：知识表示和推理 III

饶洋辉

计算机学院,

中山大学

[raoyangh@mail.sysu.edu.cn](mailto:raoyangh@mail.sysu.edu.cn)

<http://cse.sysu.edu.cn/node/2471>

课件来源：中山大学王甲海教授；浙江大学吴飞教授等

# 知识表示和推理

- 1 谓词逻辑
- 2 归结推理
- 3 知识图谱

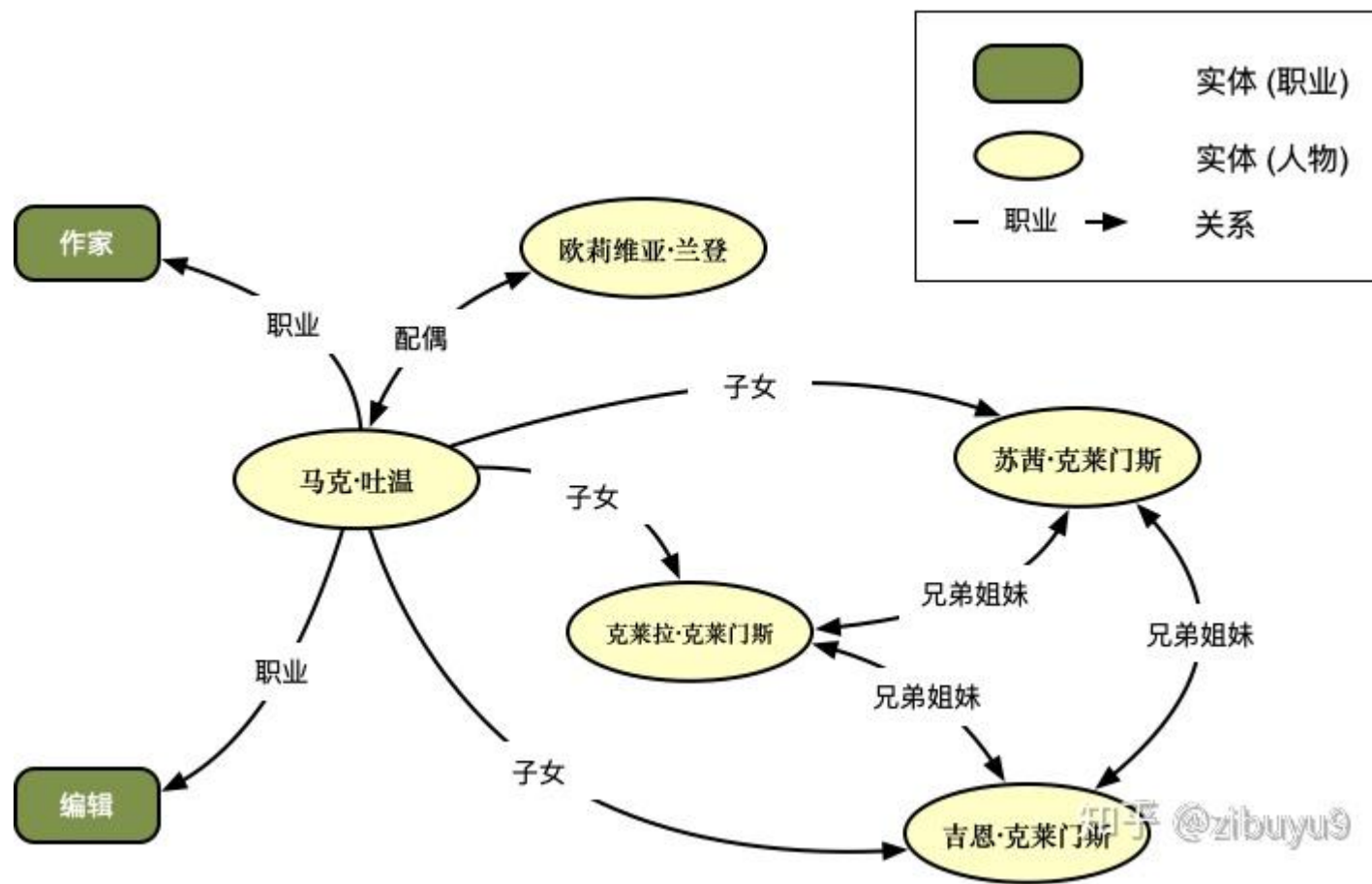
# 知识图谱及其表示

- 由于互联网内容的大规模、异质多元、组织结构松散的特点，给人们有效获取信息和知识提出了挑战。
- 谷歌于2012年5月16日首先发布了知识图谱（Knowledge Graph）。
- 知识图谱是一种互联网环境下的知识表示方法。
- 知识图谱的目的是为了提高搜索引擎的能力，改善用户的搜索质量以及搜索体验。
- 谷歌、百度和搜狗等搜索引擎公司构建的知识图谱，分别称为知识图谱、知心和知立方。

# 知识图谱及其表示

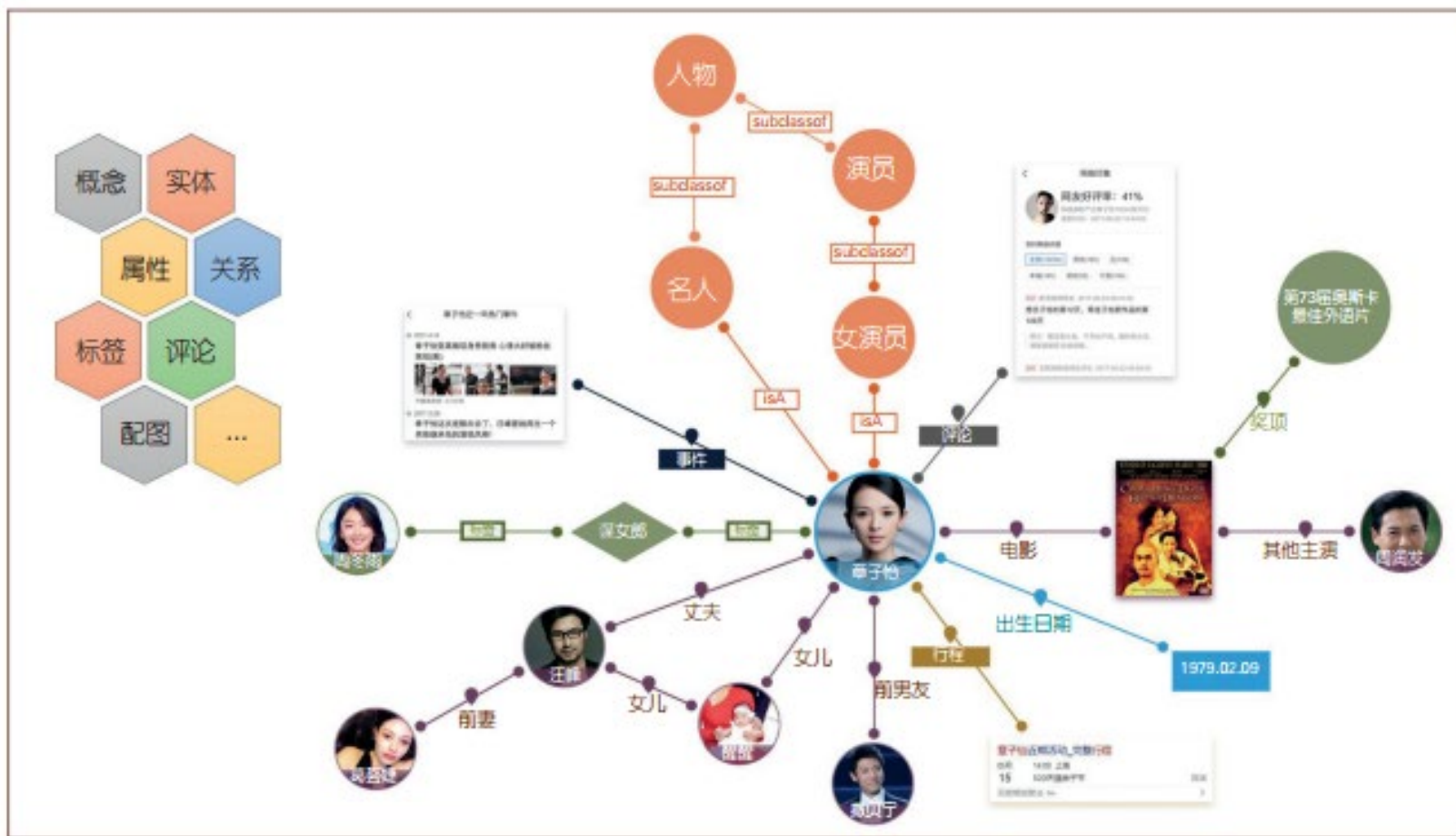
- 知识图谱的概念于2013年以后开始在学术界和工业界普及，并在智能问答、医疗、反欺诈等应用中发挥了重要作用。
- 关于知识图谱的2种观点：
  - 知识图谱是一个实体-关系的有向图。
  - 知识图谱包含更抽象的概念之间的关系，例如，谷歌和必应、雅虎一起推出了Schema.org，用来提供一个覆盖广泛主题（包括人物、地点、事件等）的模式（schema）。

# 知识图谱及其表示



<https://zhuanlan.zhihu.com/p/91762831>

# 知识图谱及其表示



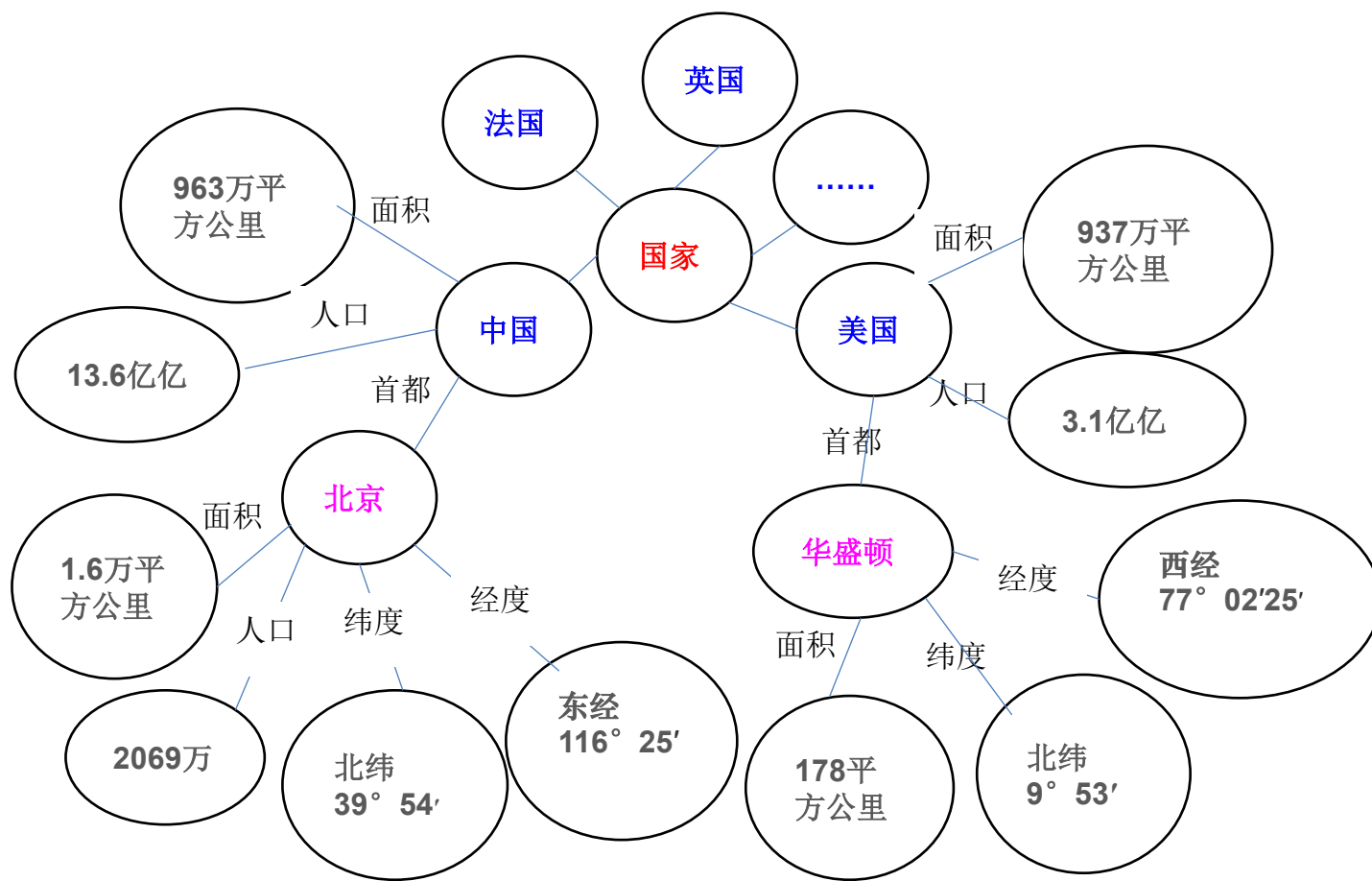
□ 知识图谱本质上是语义网络（semantic network）知识库

# 知识图谱及其表示

- 知识图谱（Knowledge Graph/Vault），又称科学知识图谱，是用各种图形等可视化技术描述知识资源及其载体，挖掘、分析、构建、绘制和显示知识及它们之间的相互联系。
- 知识图谱是由一些相互连接的实体及其属性构成的。
- 三元组是知识图谱的一种通用表示方式：
  - (实体1-关系-实体2)：中国-首都-北京
  - (实体-属性-属性值)：北京-人口-2069万

# 知识图谱及其表示

- 知识图谱也可被看作是一张图，图中的节点表示实体或概念，而图中的边则由属性或关系构成。

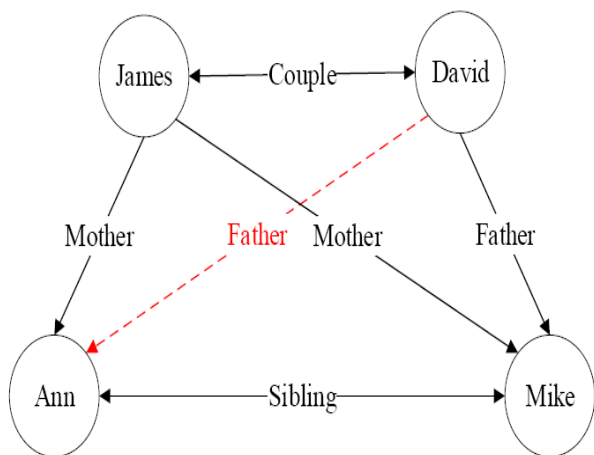




# 知识图谱推理

- **知识图谱推理包括基于符号的推理和基于统计的推理等**
  - **基于符号的推理**一般基于经典逻辑（一阶谓词逻辑或者命题逻辑）或者经典逻辑的变异（比如说缺省逻辑）。基于符号的推理可以从一个已有的知识图谱推理出新的实体间关系；同时可以对知识图谱进行逻辑的冲突检测。
  - **基于统计的推理**一般基于机器学习方法，通过统计规律从知识图谱中学习到新的实体间关系；并且对新学到的关系进行评分，去掉那些可能错误的关系。

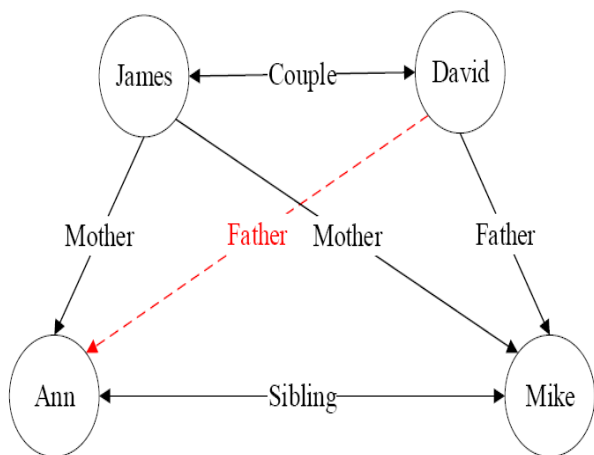
# 基于符号的知识图谱推理



一个简单的家庭关系知识图谱

- 知识图谱中任意两个相连节点及其连接边（关系）可以表示成一个三元组（triplet），即  $\langle \text{left\_node}, \text{relation}, \text{right\_node} \rangle$ ，如  $\langle \text{David}, \text{Father}, \text{Mike} \rangle$ 。
- 主要任务：关系推理（链路预测）。从图中已有关系可推知David和Ann具有父女关系，但这一关系在初始图(无红线)中并不存在，是需要推理的目标。

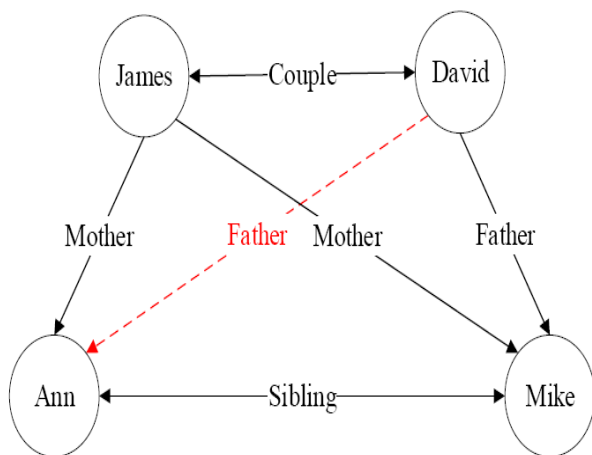
# 基于符号的知识图谱推理



一个简单的家庭关系知识图谱

- 可利用一阶谓词来表达刻画知识图谱中节点之间存在的关系，如图中形如 $\langle \text{James}, \text{Couple}, \text{David} \rangle$ 的关系可用一阶逻辑的形式来描述，即 $\text{Couple}(\text{James}, \text{David})$ 。
- $\text{Couple}(x, y)$ 是一阶谓词， $\text{Couple}$ 是图中实体之间具有的关系， $x$ 和 $y$ 是谓词变量。
- 问题：如何从知识图谱中推理得到  $\text{Father}(\text{David}, \text{Ann})$

# 基于符号的知识图谱推理



一个简单的家庭关系知识图谱

- 可利用一阶谓词来表达刻画知识图谱中节点之间存在的关系，如图中形如<James, Couple, David>的关系可用一阶逻辑的形式来描述，即*Couple*(James, David)。
- *Couple*(*x*, *y*)是一阶谓词，*Couple*是图中实体之间具有的关系，*x*和*y*是谓词变量。
- 问题：如何从知识图谱中推理得到 *Father*(David, Ann)
- 能否从具体例子中归纳学习到下述规则：

$$\frac{(\forall x)(\forall y)(\forall z)(Mother(z, y) \wedge Couple(x, z) \rightarrow Father(x, y))}{\text{前提约束谓词 (学习得到)}} \quad \frac{}{\text{目标谓词 (已知)}}$$

前提约束谓词  
(学习得到)

目标谓词  
(已知)

# 现有知识图谱资源

- 依靠人工构建的知识资源

- 早期知识资源建立是通过人工添加和群体智能合作编辑得到，如英文WordNet和Cyc项目，以及中文的HowNet。Cyc是一个通用的世界知识库，始建于1984年，其目的是将上百万条知识编码为机器可处理形式，并在此基础上实现知识推理等人工智能相关任务。Cyc包含了50万实体，接近3万个关系以及5百万事实。

- 基于群体智能的知识图谱

- 维基百科是至今利用群体智能建立的互联网上最大的知识资源，因此出现了很多使用维基百科构建知识库的项目，如DBpedia、YAGO和Freebase等。DBpedia以构建本体的形式对知识条目进行组织。YAGO融合WordNet良好的概念层次结构和维基百科中的大量实体数据。Freebase是基于维基百科，使用群体智能方法建立的包含6800万实体的结构化数据的知识图谱。清华大学和上海交通大学通过利用互动百科、百度百科，建立大规模知识图谱XLORE和Zhishi.me。XLORE以英文维基百科为桥梁，通过跨语言链接技术，建立融合了四大中英文百科数据的跨语言知识库。

# 现有知识图谱资源

- **基于互联网上链接数据构建的知识资源**

- 国际万维网组织W3C于2007年发起的开放互联数据项目（Linked Open Data, LOD），为实现网络环境下的知识发布、互联、共享和服务提供了创新技术，为智能搜索、知识问答和语义集成提供了创新源动力。Sean Bechhofer 等人在科学领域自建了一个近似于Linked Data的语义数据资源，该资源包含更准确的学术用语，并能很好地反映研究者的影响力。

- **基于机器学习和信息抽取构建的知识图谱**

- 从互联网数据自动获取知识是建立可持续发展知识图谱的发展趋势。这类知识图谱构建的特点是面向互联网的大规模、开放、异构环境，利用机器学习和信息抽取技术自动获取Web上的信息构建知识库。如华盛顿大学图灵中心的KnowItAll和TextRunner项目、卡内基梅隆大学的“永不停歇的语言学习者”（Never-Ending Language Learner, NELL）项目都是这种类型的知识库。

# 现有知识图谱资源

- **维基百科 (Wikipedia)** 由维基媒体基金会负责运营的一个自由内容、自由编辑的多语言知识库。
- **DBpedia** 由2007年德国柏林自由大学及莱比锡大学的研究者从维基百科里萃取结构化知识的项目开始建立。
- **YAGO** 由德国马克斯-普朗克研究所 (MPI) 构建的大型多语言的语义知识库，是从10个维基百科以不同语言提取事实和事实的组合。
- **XLORE** 是清华大学构建的基于中、英文维基和百度百科的开放知识平台，是第一个中英文知识规模较为平衡的大规模中英文知识图谱。