

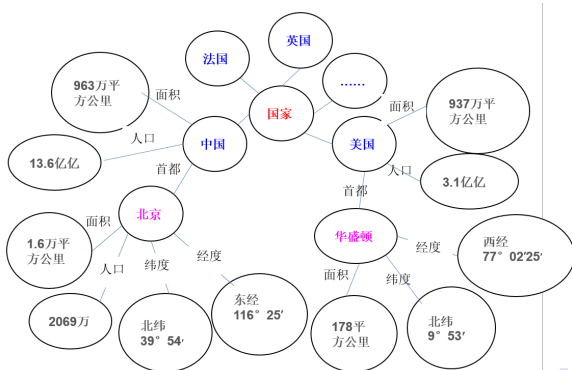
# 知识图谱 (Knowledge Graphs)

- 知识图谱简介
- 知识图谱中的推理

\*Slides based on those of Jiahai Wang and Fei Wu

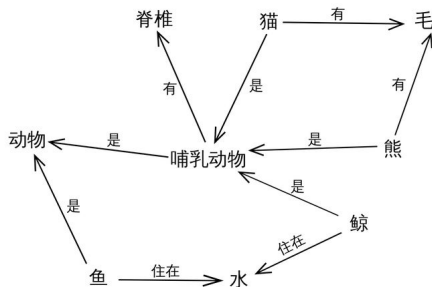
# 什么是知识图谱

- 知识图谱是一张有向图，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
- 知识图谱可以看作是三元组(triple)的集合
  - (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



# 语义网络 (Semantic Network)

- 知识图谱本质上是语义网络
- 语义网络是一种由节点和边组成的“有向”或“无向”图。其中，节点表示概念，边表示概念间的语义关系。
- 语义网络是由Quillian于1968年作为人类联想记忆的一个心理学模型提出的
- 随后在AI中用于自然语言理解。



# 知识图谱的提出

- 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。
- 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。
- Google、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。

# 知识图谱应用：新一代搜索引擎

传统搜索：基于关键词匹配

知识图谱可用来更好地查询复杂的关联信息



刘德华的妻子的出生地



百度一下

Q 网页 知道 资讯 贴吧 文库 图片 地图 采购 视频 更多

百度为您找到相关结果约712,000个

搜索工具



朱丽倩出生地:

来自百度百科

## 马来西亚槟城

朱丽倩(Carol), 1966年4月6日出生于马来西亚槟城。曾当选为马来西亚选美小姐, 后做了平面模特。1984年, 参加马来西亚槟城的'新潮小姐'选美获得季军, 之后赴香港学美容; 1985年至1987年... 详情 >

## 刘德华的老婆是谁 朱丽倩的身世背景\_有图有真相



刘德华的老婆是朱丽倩(Carol), 1966年4月6日出生于马来西亚槟城。曾当选为马来西亚选美小姐, 后做了平面模特。1984年, 她和姐妹参加马来西亚槟城的新潮小姐选美获得季军, 之后赴香港学美容; 1985年至1...

有图有真相 百度快照

## 刘德华的老婆叫什么名 刘德华妻子简历 - 达人家族

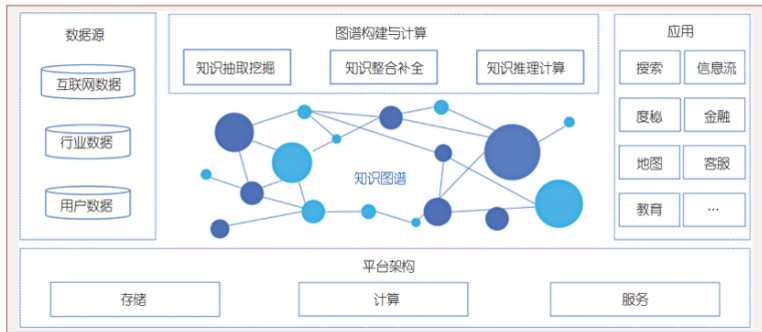


2022年1月14日 1、刘德华的妻子是朱丽倩(Carol), 朱丽倩1966年4月6日出生于马来西亚槟城。曾当选为马来西亚选美小姐, 后做了平面模特。2、1984年, 她和姐妹参加马来西亚槟城的新潮小姐选美获得季军...

达人家族 百度快照

- 是指在建立知识图谱中使用的技术，是融合认知计算、知识表示与推理、信息检索与抽取、自然语言处理与语义Web、数据挖掘与机器学习等的交叉研究。
- 知识图谱构建解决如何设计计算机算法，从客观世界或者互联网的各种数据资源中获取客观世界的知识。
- 知识图谱应用主要研究如何利用知识图谱，建立基于知识的系统，并提供知识服务，更好地解决实际问题。

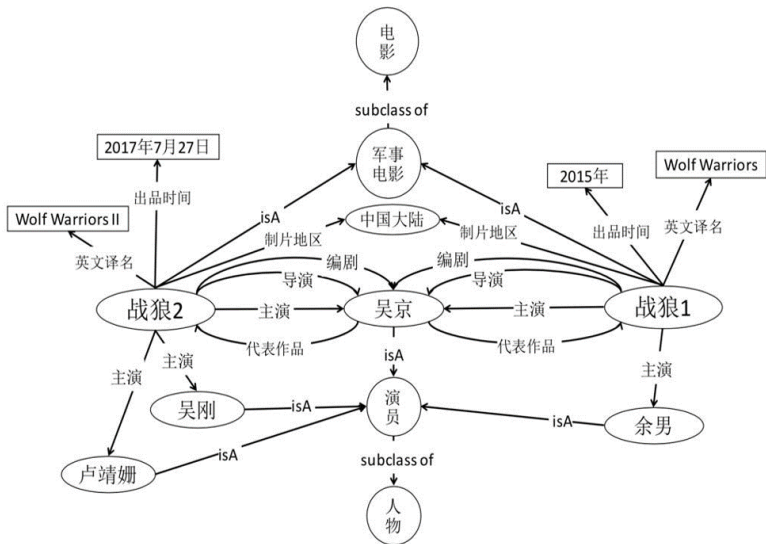
# 百度知识图谱技术视图



- 1 面向海量数据的知识图谱构建技术
- 2 大规模知识图谱补全技术：基于已有知识图谱开展知识挖掘，对缺失的图谱关系进行补全
- 3 智能知识图谱认知技术：主要研究基于给定知识图谱的深度语义解析技术
- 4 超大规模高性能分布式图索引及存储计算技术：研究面向海量知识数据的图存储、图索引、图计算和应用框架技术
- 5 知识图谱应用技术：实现知识图谱在搜索、问答、对话等产品中的规模化应用





# 在智慧推荐中的应用





根据知识图谱中的路径为战狼1的观众推荐战狼2


# 在智能问答中的应用


 NATURAL LANGUAGE

 MATH INPUT

 EXTENDED KEYBOARD

 EXAMPLES

 UPLOAD

 RANDOM

Input interpretation

Edward Heath cause of death

Result

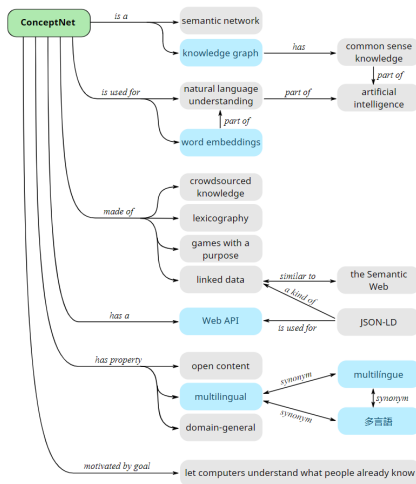
pneumonia

Basic information

full name	Edward Richard George Heath
date of birth	Sunday, July 9, 1916 (105 years ago)
place of birth	Broadstairs, Kent, United Kingdom
date of death	Sunday, July 17, 2005 (age: 89 years) (16 years ago)
place of death	Salisbury, Wiltshire, United Kingdom

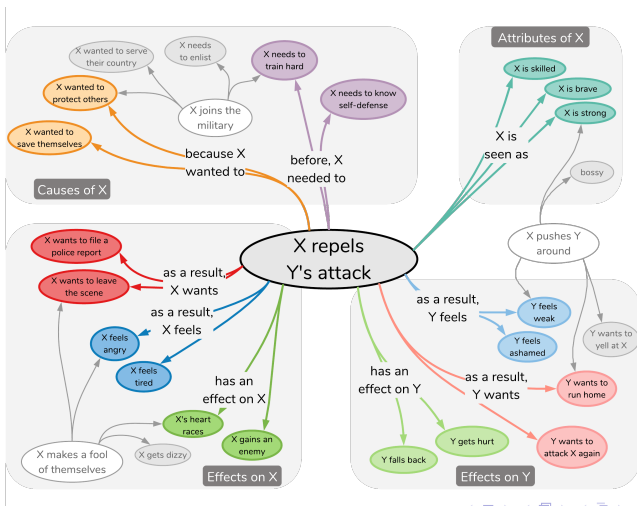
# ConceptNet

起源于众包项目 Open Mind Common Sense, 由MIT媒体实验室于1999年发起

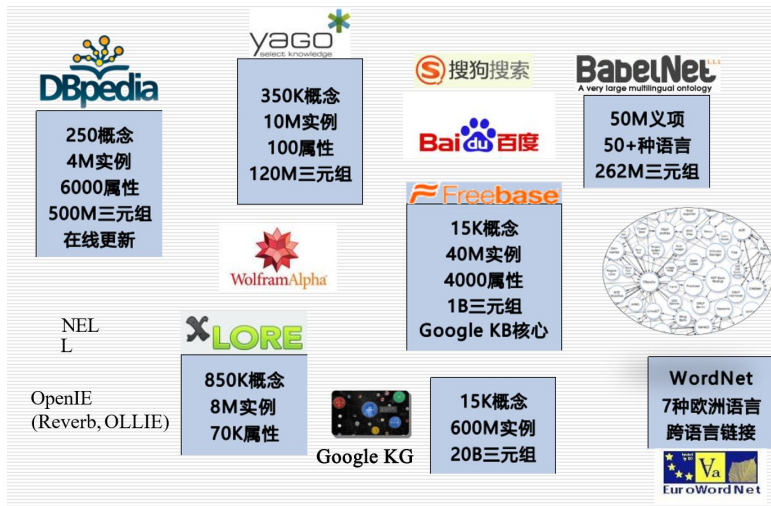


# ATOMIC常识知识图谱

由美国Allen研究所提出，关于日常事件及其因果，基于众包事件：X对Y进行反击



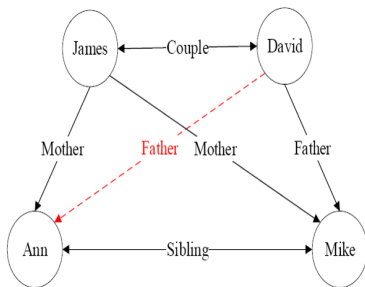
# 现有知识图谱资源



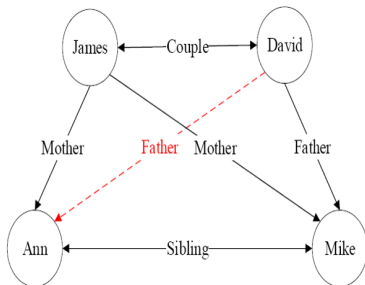
- 知识图谱中的三元组可以表示为一阶逻辑中的原子公式，从而为基于知识图谱的推理创造了条件
- 例如，从<奥巴马，出生地，夏威夷>和<夏威夷，属于，美国>两个三元组，可推理得到<奥巴马，国籍，美国>

# 知识图谱补全

- 给定一个简单的家庭关系知识图谱
- 图中形如 $\langle \text{James}, \text{Couple}, \text{David} \rangle$ 的关系可用一阶逻辑公式来描述, 即 $\text{Couple}(\text{James}, \text{David})$
- 从图中已有关系可推知David和Ann具有父女关系, 但这一关系在图中初始图(无红线)中并不存在



- 问题：如何从知识图谱中推理得到  $Father(David, Ann)$
- 解：希望学习到规则  
$$\forall x \forall y \forall z. Mother(z, y) \wedge Couple(x, z) \rightarrow Father(x, y)$$

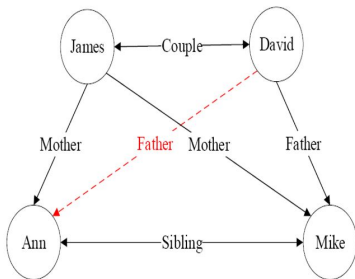




- 归纳逻辑程序设计 (Inductive Logic Programming, ILP) 是机器学习和逻辑程序设计交叉领域的研究内容。
- ILP使用一阶谓词逻辑进行知识表示，通过修改和扩充逻辑表达式对现有知识归纳，完成推理任务。
- 作为ILP的代表性方法，FOIL (First Order Inductive Learner) 通过逐步覆盖实现规则推理。

- $\forall x \forall y \forall z. \text{Mother}(z, y) \wedge \text{Couple}(x, z) \rightarrow \text{Father}(x, y)$
- $\text{Father}(x, y)$  是已知的目标谓词
- $\text{Mother}(z, y) \wedge \text{Couple}(x, z)$  是前提约束谓词，学习得到
- 推理手段: positive examples + negative examples + background knowledge examples  $\Rightarrow$  hypothesis

# 正例和反例

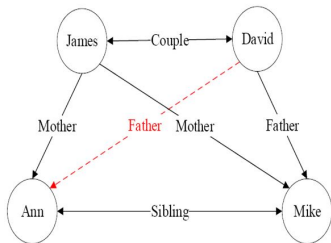


一个简单的家庭关系知识图谱

- **目标谓词:**  $Father(x, y)$
- 目标谓词只有一个正例  $Father(David, Mike)$ 。

反例在知识图谱中一般不会显式给出，但可从知识图谱中构造出来。如从知识图谱中已经知道  $Couple(David, James)$  成立，则  $Father(David, James)$  可作为目标谓词  $P$  的一个反例，记为  $\neg Father(David, James)$ 。

- 只能在已知两个实体的关系且确定其关系与目标谓词相悖时，才能将这两个实体用于构建目标谓词的反例，而不能在不知两个实体是否满足目标谓词前提下将它们来构造目标谓词的反例。



- **目标谓词:**  $Father(x, y)$
- **背景知识:** 知识图谱中目标谓词以外的其他谓词实例化结果, 如  $Sibling(Ann, Mike)$

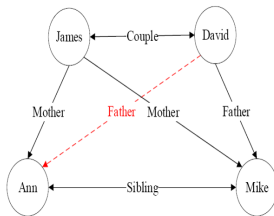
- $\forall x \forall y \forall z. Mother(z, y) \wedge Couple(x, z) \rightarrow Father(x, y)$
- 从一般到特殊，逐步给目标谓词添加前提约束谓词，直到所构成的推理规则不覆盖任何反例
- 哪些谓词可以作为目标谓词的前提约束谓词？
- 直观思想：覆盖的正例尽可能多

FOIL信息增益值计算方法如下：

$$FOIL\_Gain = \widehat{m}_+ \cdot \left( \log_2 \frac{\widehat{m}_+}{\widehat{m}_+ + \widehat{m}_-} - \log_2 \frac{m_+}{m_+ + m_-} \right)$$

其中， $\widehat{m}_+$ 和 $\widehat{m}_-$ 是增加前提约束谓词后所得新推理规则覆盖的正例和反例的数量， $m_+$ 和 $m_-$ 是原推理规则所覆盖的正例和反例数量。

# 举例说明



背景知识 样例集合	Sibling(Ann, Mike)	目标谓词 训练样例 集合	Father(David, Mike)
	Couple(David, James)		$\neg$ Father(David, James)
	Mother(James, Ann)		$\neg$ Father(James, Ann)
	Mother(James, Mike)		$\neg$ Father(James, Mike)
			$\neg$ Father(Ann, Mike)

- *Mother*(·,·)
- *Sibling*(·,·)
- *Couple*(·,·)



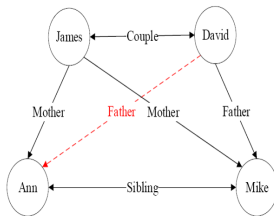
依次将谓词加入到推理规则中作为前提约束谓词，  
并计算所得新推理规则的FOIL增益值。  
基于计算所得FOIL增益值来选择最佳前提约束谓词。

- 前提约束谓词为空集时，推理规则所覆盖的正例和反例的样本数分别是训练样本中正例和反例的数量，  
即 $m_+ = 1, m_- = 4$ 。
- 将 $Mother(x, y)$ 作为前提约束谓词加入，可得到推理规则  
 $Mother(x, y) \rightarrow Father(x, y)$
- 覆盖正例和反例数量分别为0和2，即 $\hat{m}_+ = 0, \hat{m}_- = 2$ 。
- 由于 $\hat{m}_+ = 0$ ，代入FOIL Gain公式时会出现负无穷的情况，  
此时FOIL Gain记为NA (Not Available)
- 考虑 $Couple(x, z) \rightarrow Father(x, y)$
- 覆盖正例和反例数量分别为1和1，信息增益值为1.32

推理规则		推理规则涵盖的正例和反例数		FOIL信息增益值
目标谓词	前提约束谓词	正例	反例	信息增益值
$Father(x, y)$ ←	空集	$m_+ = 1$	$m_- = 4$	$FOIL\_Gain$
$Father(x, y)$ ←	$Mother(x, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 2$	NA
	$Mother(x, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 2$	NA
	$Mother(y, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	$Mother(y, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	$Mother(z, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	$Mother(z, y)$	$\widehat{m}_+ = 1$	$\widehat{m}_- = 3$	0.32
	$Sibling(x, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	$Sibling(x, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	$Sibling(y, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$Sibling(y, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$Sibling(z, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$Sibling(z, y)$	$\widehat{m}_+ = 1$	$\widehat{m}_- = 2$	0.74
	$Couple(x, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	<b><math>Couple(x, z)</math></b>	<b><math>\widehat{m}_+ = 1</math></b>	<b><math>\widehat{m}_- = 1</math></b>	1.32
	$Couple(y, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$Couple(y, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$Couple(z, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 2$	NA
	$Couple(z, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA



# 确定谓词



<b>Back-ground knowledge</b>	<p>Sibling(Ann, Mike)</p> <p>Couple(David, James)</p> <p>Mother(James, Ann)</p> <p>Mother(James, Mike)</p>
<b>Positive and negative samples</b>	<p>Father(David, Mike)</p> <p><math>\neg</math>Father(David, James)</p> <p><math>\neg</math>Father(James, Ann)</p> <p><math>\neg</math>Father(James, Mike)</p> <p><math>\neg</math>Father(Ann, Mike)</p>

- $Couple(x, z)$  加入后信息增益最大
- 将  $Couple(x, z)$  加入推理规则，得到  $Couple(x, z) \rightarrow Father(x, y)$  新推理规则
- 将训练样例中与该推理规则不符的样例去掉。这里不符指当  $Couple(x, z)$  中  $x$  取值为 David 时，与  $Father(David, )$  或  $\neg Father(David, )$  无法匹配的实例。
- 训练样本集中只有正例  $Father(David, Mike)$  和负例  $\neg Father(David, James)$  两个实例

推理规则		推理规则涵盖的 正例和反例数		FOIL信息增益值
现有规则	拟加入前提 约束谓词	正例	反例	信息增益值
$Father(x, y) \leftarrow Couple(x, z)$		$m_+ = 1$	$m_- = 1$	1.32
$Father(x, y) \leftarrow Couple(x, z)$	$\wedge Mother(x, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Mother(x, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Mother(y, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Mother(y, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Mother(z, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge \mathbf{Mother}(z, y)$	$\mathbf{\widehat{m}_+ = 1}$	$\mathbf{\widehat{m}_- = 0}$	<b>1</b>
	$\wedge Sibling(x, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Sibling(x, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Sibling(y, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Sibling(y, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Sibling(z, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Sibling(z, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Couple(x, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 1$	NA
	$\wedge Couple(x, z)$	$\widehat{m}_+ = 1$	$\widehat{m}_- = 1$	0
	$\wedge Couple(y, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Couple(y, z)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Couple(z, x)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA
	$\wedge Couple(z, y)$	$\widehat{m}_+ = 0$	$\widehat{m}_- = 0$	NA

# 学习结束并用于推理

- $Mother(z, y)$  加入信息增益最大
- 得到  $Mother(z, y) \wedge Couple(x, z) \rightarrow Father(x, y)$
- 该推理规则覆盖训练样本集合中所有正例且不覆盖任意反例，因此算法学习结束。
- 已知  $Mother(James, Ann)$ ,  $Couple(David, James)$ , 于是  $Father(David, Ann)$

# 知识图谱推理: FOIL (First Order Inductive Learner)

$$(\forall x)(\forall y)(\forall z)(Mother(z, y) \wedge Couple(x, z) \rightarrow Father(x, y))$$



前提约束谓词  
(学习得到)



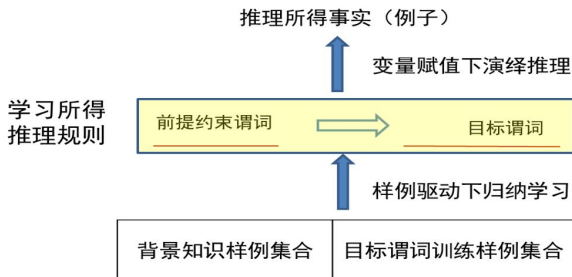
目标谓词  
(已知)

推理手段: *positive examples* + *negative examples* + *background knowledge examples*  $\Rightarrow$  **hypothesis**

背景知识 样例 集合	Sibling	(Ann, Mike)	目标谓词 训练样例 集合	Father(David, Mike)
	Couple	(David, James)		$\neg$ Father(David, James)
	Mother	(James, Ann)		$\neg$ Father(James, Ann)
	Mother	(James, Mike)		$\neg$ Father(James, Mike)
集合	Mother	(James, Mike)		$\neg$ Father(Ann, Mike)

给定目标谓词, FOIL算法从实例(正例、反例、背景样例)出发, 不断测试所得到推理规则是否还包含反例, 一旦不包含负例, 则学习结束, 展示了“**归纳学习**”能力。

# 知识图谱推理：FOIL (First Order Inductive Learner)



给定目标谓词，FOIL算法从实例（正例、反例、背景知识样例）出发，不断测试所得推理规则是否还包含反例，一旦不包含，则学习结束，由此充分展示了“归纳学习”的能力。在学得推理规则后，再给推理规则中的变量赋予具体例子，经过“演绎”得到新的知识

# 知识图谱推理：路径排序推理

基本思想是将实体之间的关联路径作为特征，来学习目标关系的分类器。  
工作流程主要分为三步：

- (1) 特征抽取：生成并选择路径特征集合。生成路径的方式有随机游走（random walk）、广度优先搜索、深度优先搜索等。
- (2) 特征计算：计算每个训练样例的特征值。该特征值可以表示为布尔值，表示实体 $s$ 到实体 $t$ 之间是否存在指定路径；
- (3) 分类器训练：根据训练样例的特征值，为目标关系训练分类器。当训练好分类器后，即可将该分类器用于推理两个实体之间是否存在目标关系。

(1) 目标关系: *Father*

(2) 对于目标关系*Father*, 生成四组训练样例, 一个为正例、三个为负例:

正例: (David, Mike)

负例: (David, James), (James, Ann), (James, Mike)

(3) 从知识图谱采样得到路径, 每一路径链接上述每个训练样例中两个实体:

(David, Mike)对应路径: *Couple*  $\rightarrow$  *Mother*

(David, James)对应路径: *Father*  $\rightarrow$  *Mother*<sup>-1</sup> (*Mother*<sup>-1</sup>与*Mother*为相反关系)

(James, Ann)对应路径: *Mother*  $\rightarrow$  *Sibling*

(James, Mike)对应路径: *Couple*  $\rightarrow$  *Father*

# 举例说明

(4) 对于每一个正例/负例，判断上述四条路径可否链接其包含的两个实体，将可链接（记为1）和不可链接（记为0）作为特征，于是每一个正例/负例得到一个四维特征向量：

(David, Mike):  $\{[1, 0, 0, 0], 1\}$

(David, James):  $\{[0, 1, 0, 0], -1\}$

(James, Ann):  $\{[0, 0, 1, 0], -1\}$

(James, Mike):  $\{[0, 0, 1, 1], -1\}$

(5) 依据(4)中的训练样本，训练分类器 $M$ 。

(6) 预测。对于图2.3.1中形如(David, Ann)的样例，得到其特征值为 $[1, 0, 0, 0]$ ，将特征向量输入到分类器 $M$ 中，如果分类器 $M$ 给出分类结果为1，则 $Father(\text{David}, \text{Ann})$ 成立。