

# 司法知识图谱

KnowledgeGraph

曾祥辉  
2017-09-20

---

## 目录

---

1

知识图谱概述

2

图谱构建及应用技术

3

总结与参考资料

# 知识图谱概述

- 2012年Google 提出知识图谱，优化搜索
- 知识图谱是指以**有向图**的形式表达**实体**和**概念**的**属性**及其之间**关系**的知识库，**主要关注实体**
- 知识图谱实质是一种语义网络，一种知识工程的知识表示，适合集成大规模知识

奥巴马

百度一下

网页 新闻 贴吧 知道 音乐 图片 视频 地图 文库 更多»

百度为您找到相关结果约15,400,000个

搜索工具

[贝拉克·侯赛因·奥巴马](#) [百度百科](#)



职业：律师、政府官员  
生日：1961年8月4日  
代表作品：《我相信变革》《我父亲的梦想》《无畏的希望》  
简介：贝拉克·侯赛因·奥巴马，美国民主党籍政治家，第44任美国总统，为美国历史上第一位非裔美国人（美国黑人）总统。...

[人物经历](#) [主要事件](#) [主要作品](#) [家庭生活](#) [人物评价](#)

[baike.baidu.com/](#)

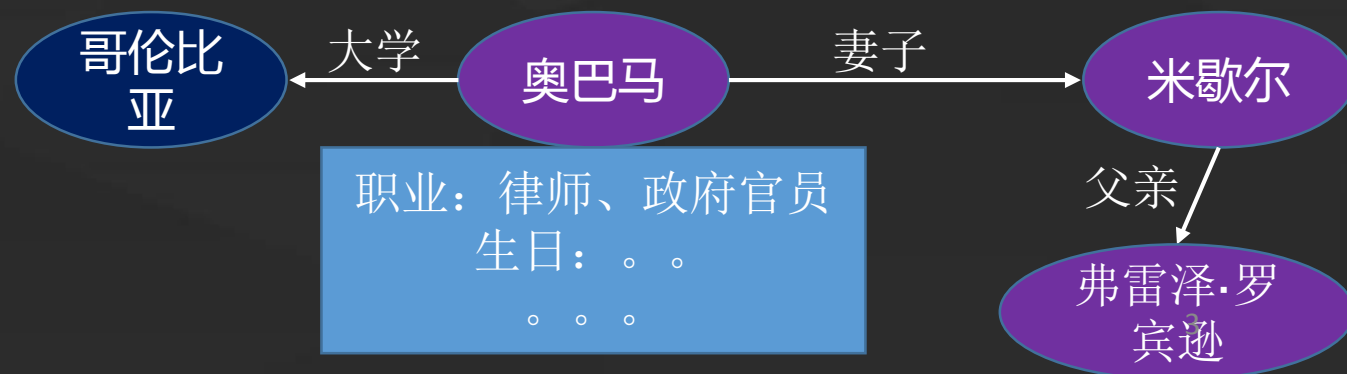
[奥巴马](#) [百度百科](#)

圣枪游侠·卢锡安是《英雄联盟》中的一名英雄。卢锡安将灌输着远古力量的遗物作为武器，以坚定守护神的身份对抗着亡灵生物。他冷酷的信念从未动摇，即使面对着令...

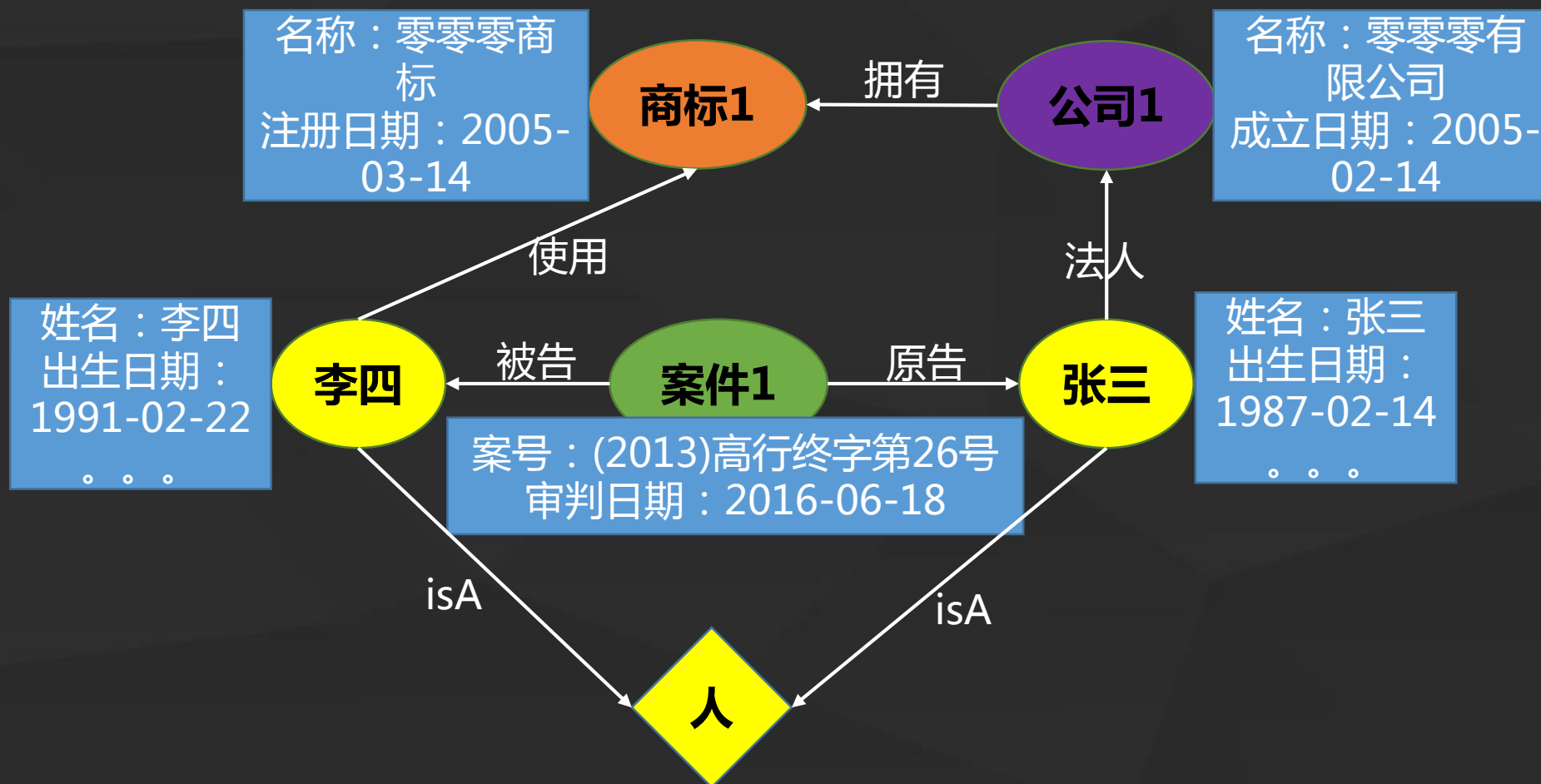
其他含义：  
[鱼类](#)  
[美国第44任总统](#)  
[查看“奥巴马”全部7个含义>>](#)

**历年时代周刊年度风云人物**

 <b>普京</b> 被认为俄罗斯铁腕总统	 <b>克林顿</b> 曾任美国总统	 <b>小布什</b> 美国第43任总统	 <b>比尔·盖茨</b> 微软公司联合创始人
 <b>肯尼迪</b> 美国第35任总统	 <b>罗斯福</b> 美国蝉联4届的总统	 <b>斯大林</b> 苏联共产党中央总书记	 <b>里根</b> 演员出身的美国总统



# 知识图谱概述



# 知识图谱概述

## 知识表示方式

	优点	缺点
文本	资源丰富，最常见的知识存储方式	信息获取慢，机器难以获取所需信息
二维表	机器获取信息效率高，技术链成熟	复杂信息或关系表达困难，多度查询速度慢，模式难扩展
图	自由，适合复杂关系信息表达和查询，易扩展，多度查询效率高	技术组件尚未成熟

# 知识图谱概述

## 应用方向

- 信息搜索和可视化优化
- 为自然语言理解提供背景知识库
- 问答系统、医疗诊断、金融反欺诈、电商搜索推荐



## 目录



1

知识图谱概述

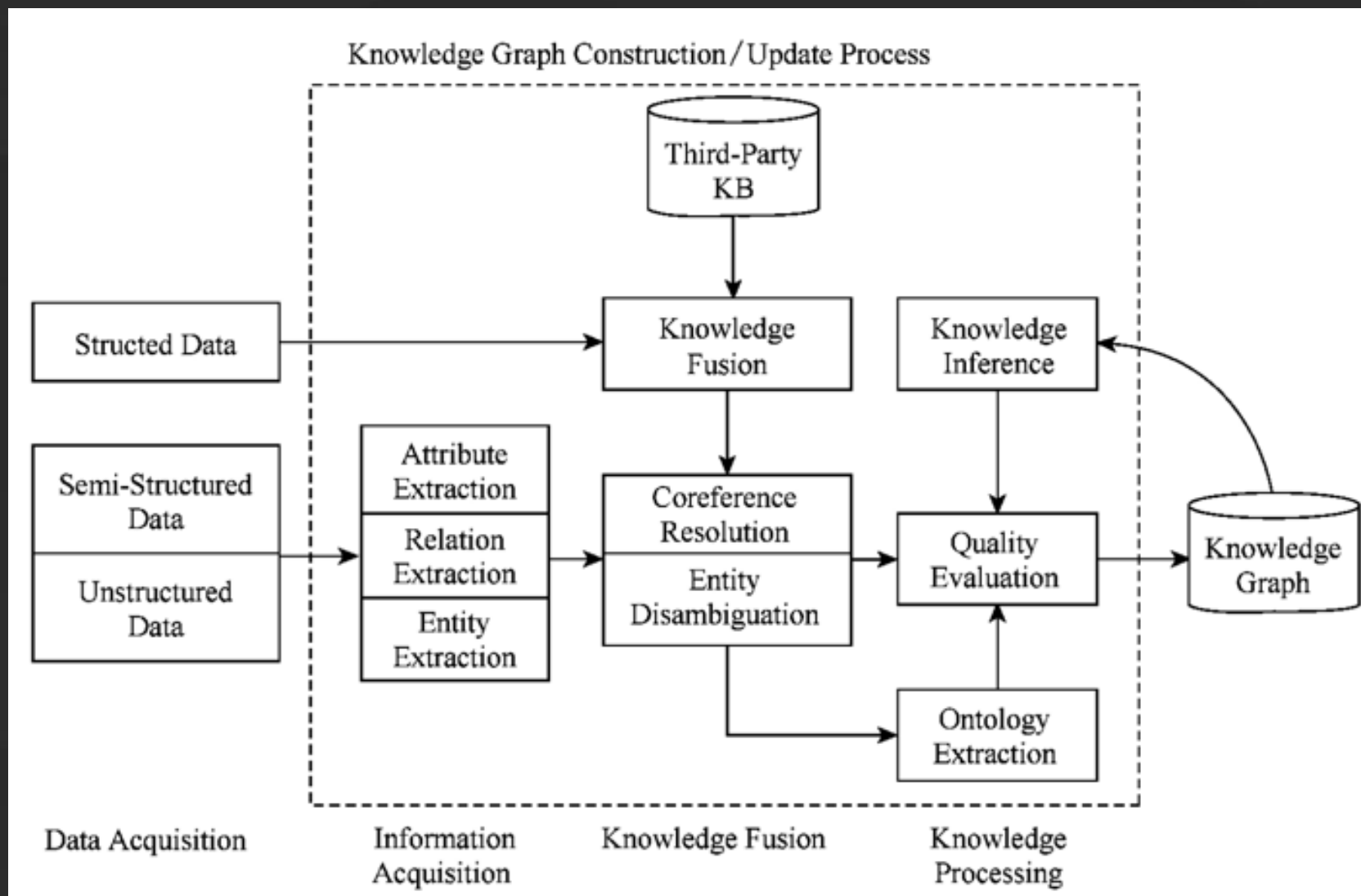
2

图谱构建及应用技术

3

总结及参考资料

# 图谱构建及应用技术





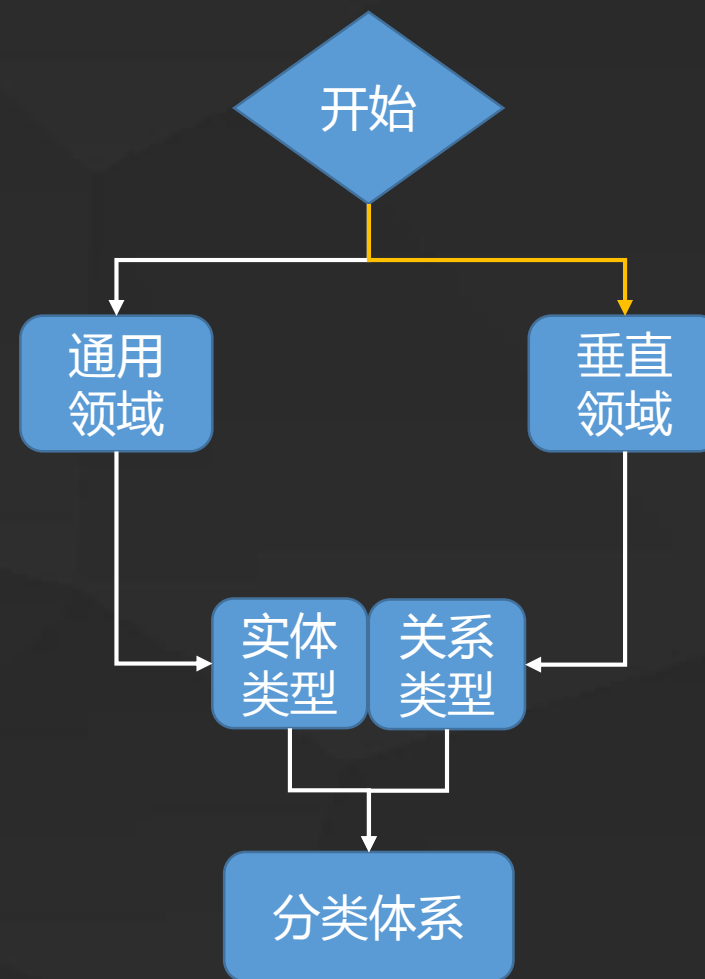
# 图谱构建及应用技术

## ➤ 明确需求

- 通用或垂直领域
- 业务需求关注的实体类型、关系类型
- 分类体系

## ➤ 确定数据来源

- 通用图谱：百科网站、互联网文本等
- 领域图谱：垂直网站（法律咨询网站、文书网）、期刊、书籍等



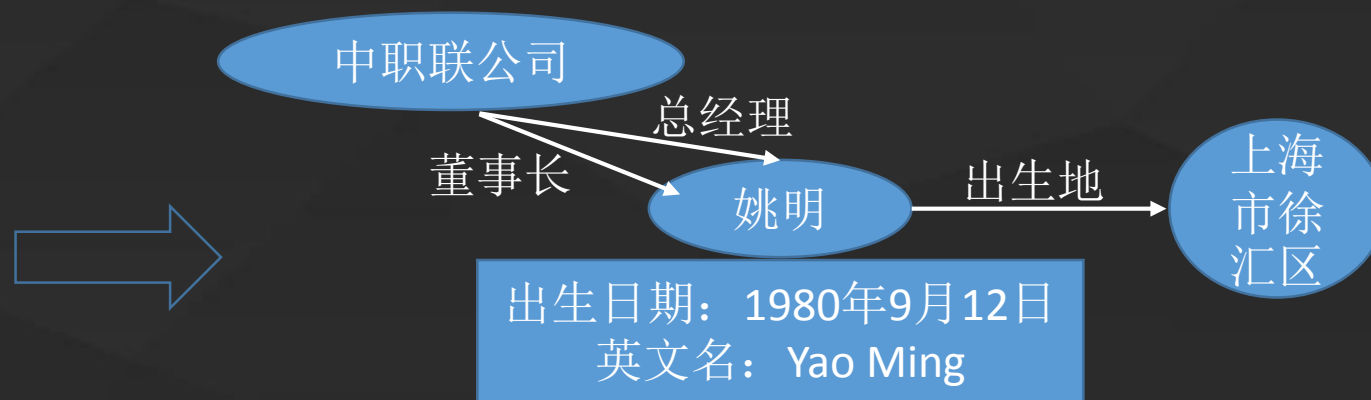
# 图谱构建及应用技术

## ➤ 知识抽取

### ● 实体抽取（NER）及关系（属性）抽取

- 基于规则和词典的方法：在词法、句法分析基础上，见效快；规则难以穷尽、瓶颈
- 基于统计的方法：带标语料难以获得，尤其是垂直领域
- 二者结合的方法：半监督学习，bootstrapping，种子学习+pattern，效果有待优化
- 神经网络：LSTM+CRF

姚明 (Yao Ming), 1980年9月12日出生于上海市徐汇区，祖籍江苏省苏州市吴江区震泽镇，前中国职业篮球运动员，司职中锋，现任中职联公司董事长兼总经理。

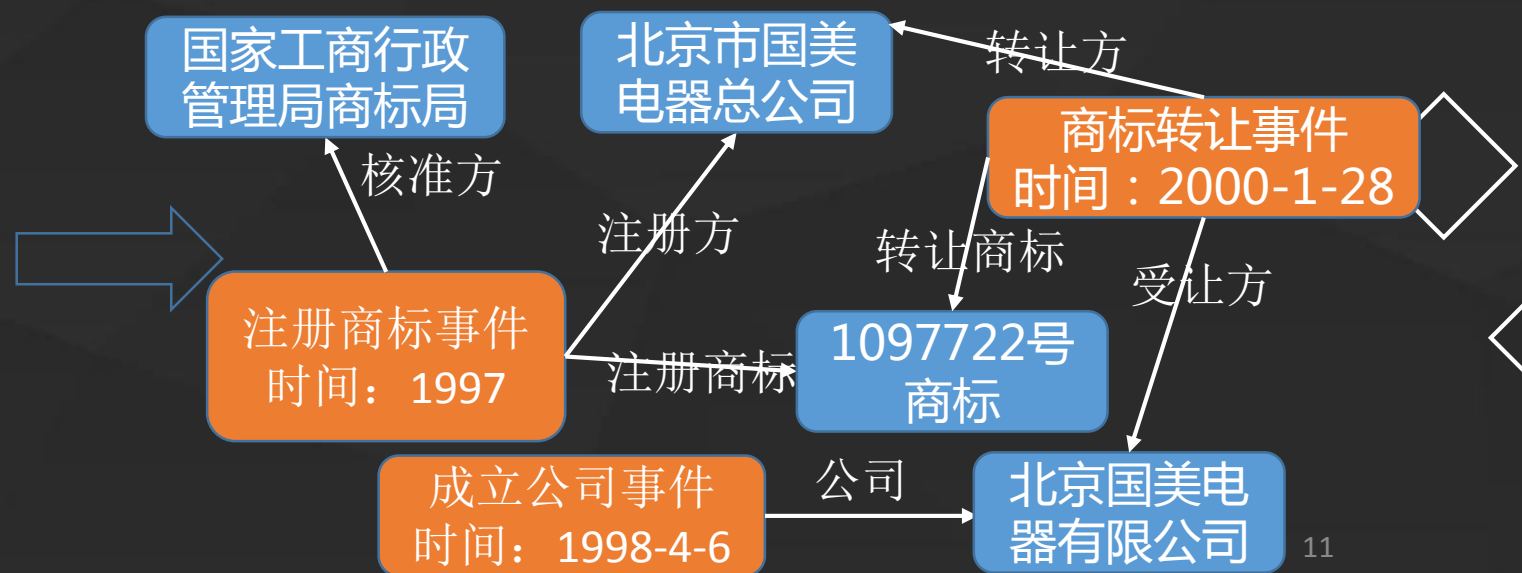


# 图谱构建及应用技术

## ➤ 知识抽取

- 事件抽取：静态转向动态，将事件作为实体的一种，拓宽实体关系，丰富图谱  
事件识别与分类：触发词、机器学习分类  
事件要素的抽取，包括实体和属性：句法、语义分析
- 概念抽取：将概念识别引入，主要丰富图谱中IsA的关系，建立层级关系

北京国美电器有限公司成立于1998年4月6日。1997年，北京市国美电器总公司向国家工商行政管理局商标局申请注册了“国美电器”商标（其中“电器”放弃专用权，下同），商标注册证号为第1097722号，注册有效期限自1997年9月7日至2007年9月6日止，核定服务项目为第35类。2000年1月28日，经国家工商行政管理局商标局核准第1097722号商标转让注册，受让人名义为北京国美电器有限公司即本案原告。



# 图谱构建及应用技术

## ➤ 知识融合

- 实体对齐：不同数据源同一实体、关系或属性的对齐，基于规则或者统计
- 关系（属性）对齐：如出生日期与出生时间
- 知识验证：来源可靠性，概率评估

百度百科

中文名	刘德华
外文名	Andy Lau, Lau Tak Wah
别 名	华仔, 华Dee, 华哥等
国 籍	中国
民 族	汉族
星 座	天秤座
血 型	AB型
身 高	174cm
体 重	63kg
出生地	香港新界大埔镇泰亨村
出生日期	1961年9月27日
职 业	演员, 歌手, 填词人, 制片人
毕业院校	可立中学, 第十期无线艺员训练班

刘德华

刘德华

刘德华

知识图谱



摄于2005年2月24日，第29届香港国际电影节

**男艺人**

罗马拼音 Lau Tak Wah

英文名 Andy Lau

昵称 华仔  
刘天王  
华Dee  
刘华  
华神（台湾尊称）  
老大（台湾粉丝另一尊称）  
刀仔

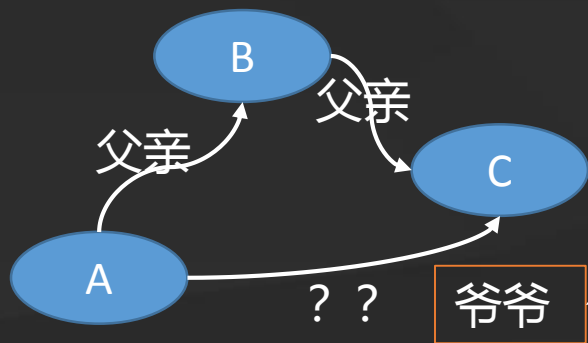
别名 紫慧（作词笔名）  
慧果（佛教法名）

维基百科

# 图谱构建及应用技术

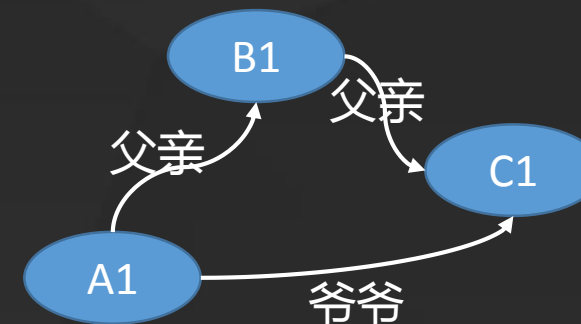
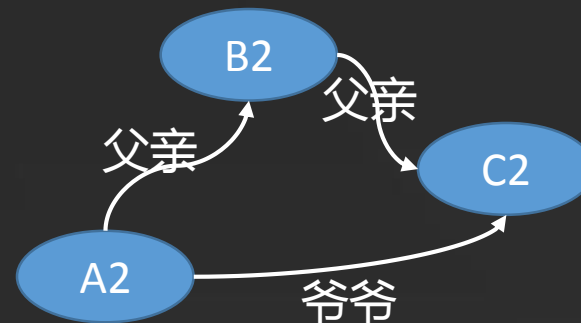
## ➤ 知识推理

- 基于规则的推理：如父亲的父亲是爷爷
- 基于统计的推理：如图中三角关系的推断



父亲的父亲是爷爷

爷爷



# 图谱构建及应用技术

## ➤ 知识分布式表示

- 向量化： $h+r=e$ ，word2vec，graph2vec?
- TransE、TransR、TransH。。
- <https://github.com/wuxiyu/transE>源码
- 李涓子老师和刘知远老师的工作



# 图谱构建及应用技术

## ➤ 搜索优化

- ◆ 实体链接
- ◆ 关系识别
- ◆ 路径推理

## ➤ 推荐系统

- ◆ 实体链接
- ◆ 分布式表示
- ◆ 相似性计算

## ➤ 问答系统

- ◆ 意图识别
- ◆ 实体链接及关系识别
- ◆ 路径推理

## 总结

- 面向业务需求，决定用什么技术
- 工程性质及快速迭代，粒度由大到小
- 有效果的技术就是最好的技术
- 重视人工协作的力量



## 参考资料

- 从语义网到知识图谱——语义技术工程化的回顾与反思

<http://www.wtoutiao.com/p/181x8bc.html>

- 降低知识图谱的构造成本

<http://weibo.com/p/1001603966996583691220>

- 知识图谱研究进展综述

[http://www.360doc.com/content/17/0325/18/99071\\_640071618.shtml](http://www.360doc.com/content/17/0325/18/99071_640071618.shtml)

- 知识图谱与深度学习

<http://blog.csdn.net/starzhou/article/details/71169636>

请指正，谢谢！