

Web信息处理与应用



第十一节 知识图谱导论

徐童

2023.11.13

- 回顾信息检索部分课程

信息检索部分

第3节 网页文字处理



第4节 网页索引



第5节 查询与评估

第6、7节 排序（上、下）

文本表征、排序学习算法、PageRank及衍生模型

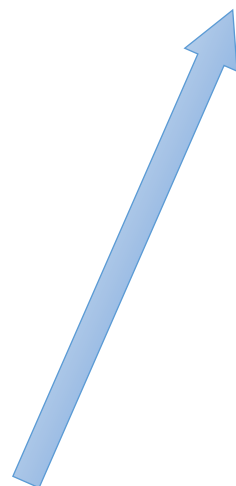


第8、9节 个性化检索（上、下）

基础个性化算法、情境感知技术、数据预处理



第10节 多模态搜索



- 然而，传统信息检索返回的是“文档的集合”，而非信息

Baidu 百度

欧文

百度一下

网页 图片 贴吧 文库 资讯 知道 地图 视频 采购 更多

百度为您找到相关结果约100,000,000个

搜索工具

[欧文\(美国现役篮球运动员\) - 百度百科](#)

 生涯：篮球（布鲁克林篮网队）
生日：1992年3月23日
重要事件：2016年奥运会男篮冠军
简介：[欧文](#)一般指凯里·[欧文](#)。凯里·[欧文](#)（Kyrie Irving...
[早年经历](#) [职业生涯](#) [国家队经历](#) [生涯数据](#) [获奖记录](#) [更多 >](#)

百度百科

[欧文的最新相关信息](#)

 [舆论开始反弹,欧文有救了?球员工会不满篮网,名记为...](#)
欧文的事情持续发酵，球迷们都想知道，此事何去何从？没有人知道下一步在哪里。但如今给人的感觉是，篮网队步步为营。最终的结...
篮郭先生 9小时前

 [重拳出击,篮网要借机裁掉欧文,一举两得 他的NBA生涯...](#)
时间来到11月8日NBA常规赛继续进行，然而篮网队球星欧文依旧是深陷麻烦，欧文的不当言论事件还在继续发酵中。篮网队连续重拳...
篮球大部落 10小时前

[\[流言板\]记者:篮网预计欧文不能完成6件事情,有可能直接回归球队](#)
[流言板]记者:篮网预计欧文不能完成6件事情,有可能直接回归球队由虎扑篮球资讯 发表在篮球资讯
<https://bbs.hupu.com/502> 虎扑11月08日讯 近日,欧文事件持续发酵,NBA知名记...
虎扑体育 8小时前

[查看更多 >](#)

- 如今，用户对“信息”的需求更为迫切



美国总统



百度一下

Q 网页 资讯 图片 视频 贴吧 知道 文库 采购 地图 更多

百度为您找到相关结果约100,000,000个

搜索工具



美国 - 总统:

来自百度百科

约瑟夫·拜登

约瑟夫·拜登，即小约瑟夫·罗宾内特·拜登（Joseph Robinette Biden, Jr），一般称作乔·拜登（Joe Biden），1942年11月20日出生于美国宾夕法尼亚州的一个爱尔兰裔中产阶级家庭，1970年... [详情 >](#)

人们已不再满足于单纯呈现原始的文档，
而需要更加**精炼**的知识表达与更加**直观**的需求解决

- 不仅提供“信息”，还提供信息的“拓展”

Baidu 鸡腿的热量 百度一下 百度首页 消息 设置

鸡腿热量:
214大卡/100g
每100克鸡腿含碳水化合物0克, 脂肪11.06克, 蛋白质26.8克, 纤维素0克。 [详情>>](#)

[鸡腿的热量, 鸡腿减肥 - 薄荷食物库](#)

别名: 鸡腿肉 **热量:** 181 大卡(100克可食部分) 分类: 蛋类、肉类及制品
红绿灯: 评价: 鸡的脂肪多囤积在皮下, 鸡腿带皮, 所以脂肪含量不低, 减肥期间建议去皮后食用。...

[去皮鸡腿的热量, 去皮鸡腿减肥 - 薄荷食物库](#)

全面分析去皮鸡腿的热量, 营养价值, 去皮鸡腿减肥功效, 食用效果以及食用注意事项等等, 薄荷网食物库提供。

[鸡腿的热量是多少 鸡腿热量高吗 百度知道](#)
2个回答 · 回答时间: 2019年5月10日
最佳答案: 热量高, 每100克鸡腿所含热量214大卡。 鸡肉肉质细嫩, 滋味鲜美, 由于其味较淡, 因此可适用于各种料理中。蛋白质的含量颇多, 在肉之中, 可以说是蛋白质最高...

[更多关于鸡腿的热量的问题>>](#)

相关食物 展开

- 黄瓜鸡蛋减肥法: 加速脂肪的消耗和排毒
- 鸡胸肉: 可改善记忆, 优质肉类
- 苦瓜减肥食谱: 降低胆固醇和甘油三酯
- 水果减肥法: 全日只吃水果不吃主食
- 酸奶减肥法: 消除便秘的救世主
- 西红柿炒鸡蛋: 饭馆里经常找不到蛋
- 卤牛肉: 强筋健骨的传统名菜
- 鸡腿炖土豆: 以鸡腿土豆制成

相关减肥方法 展开

- 冬瓜减肥法
- 水煮蛋减肥
- 香蕉减肥法
- 21天减肥法

从查询条件的扩展，到查询内容的拓展

用户希望实现 一次查询，多重服务 的信息关联

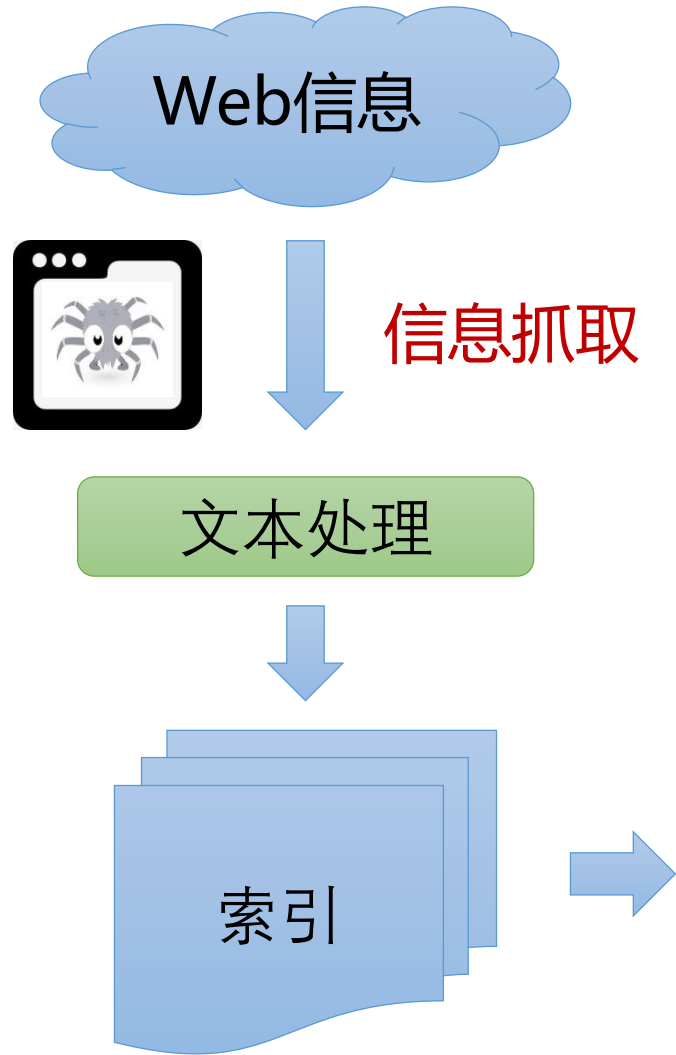
- 实现信息拓展依赖于信息之间的“关联”



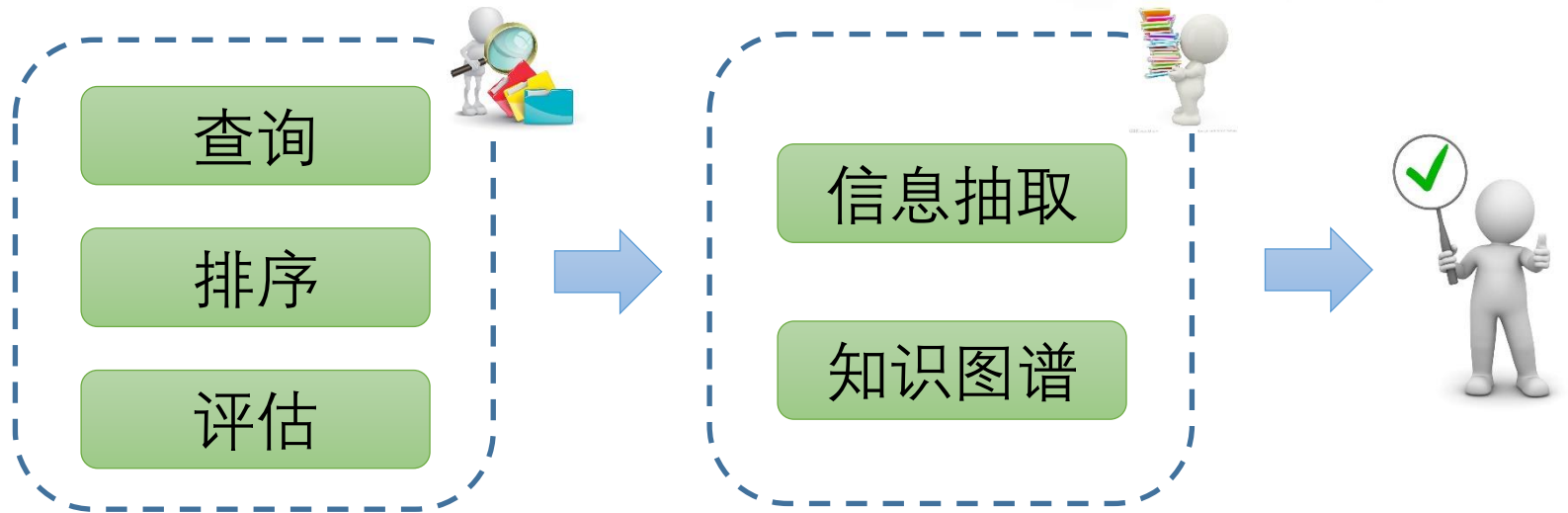
实现信息拓展的前提，是实现信息之间的**关联**

如何提炼规范化的信息，并挖掘信息之间的**关联关系**？

- 本课程所要解决的问题



第十个问题：
如何从文档中总结并关联知识？

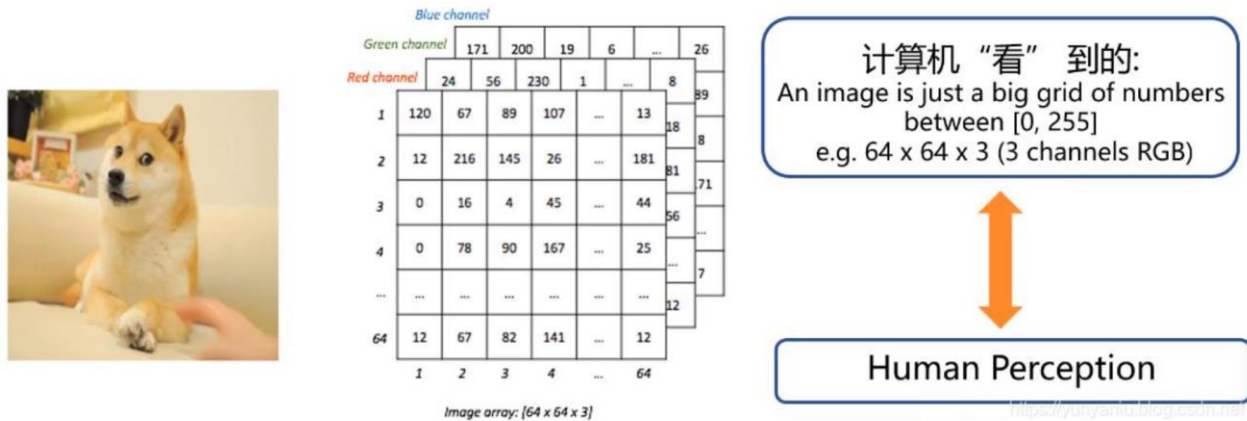


- **如何让机器理解人类的知识?**

- 人类心智的一个重要特征是抽象表达和传承复杂知识

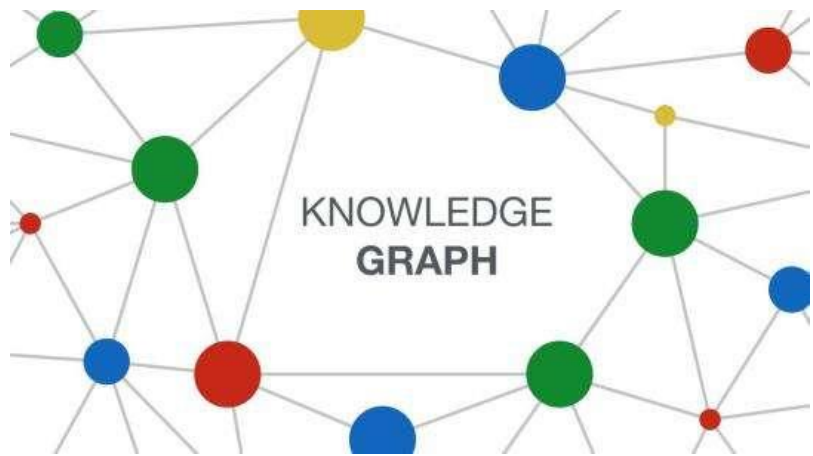
- 人类的自然语言, 绘画、音乐, 以及模型、公式等, 都是知识表达的方式

- 然而, 机器如何理解这些方式所表达的知识?



- 本节课开始，我们进入“知识图谱”部分的介绍

知识图谱部分详情



第11节 知识图谱概述



第12节 实体抽取与对齐



第13节 关系与事件抽取



第14节 知识图谱与图计算



第15节 知识图谱应用

- 知识组织形式的发展
- 信息抽取任务概述
- 知识图谱的基本概念
- 更多类型的知识图谱
 - 事理图谱
 - 多模态知识图谱

- 关于知识组织的早期设想
- 记忆机器MEMEX, 1945年

"Wholly new forms of **encyclopedias** will appear, ready made with a **mesh of associative trails** running through them, ready to be dropped into the memex and there amplified"

As We May Think, The Atlantic, 1945

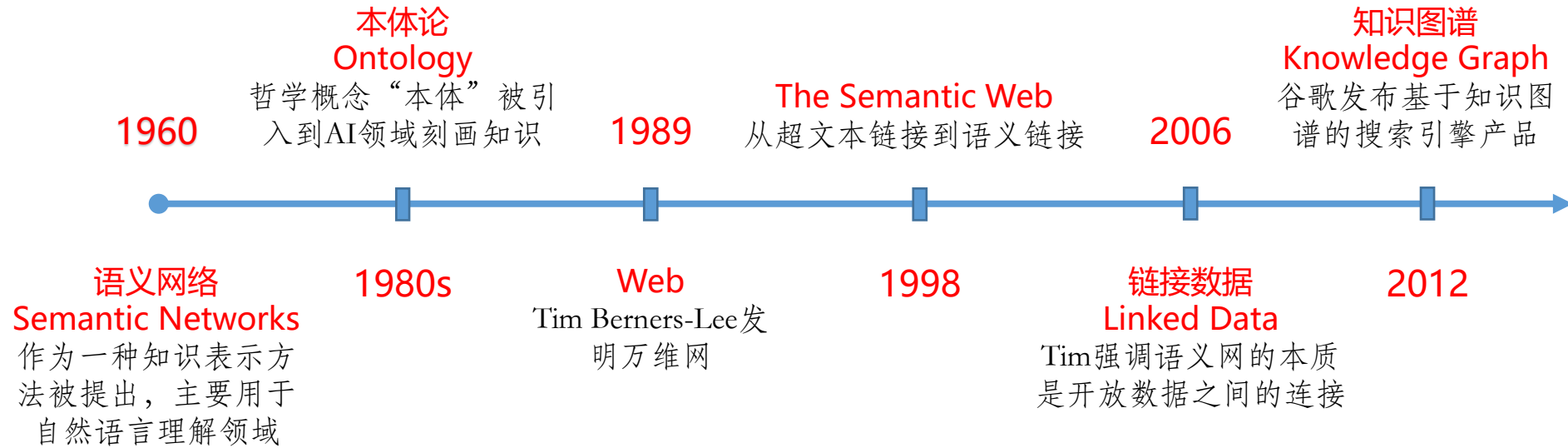
↑ 这一构想已初步具备了今天超文本和超连接的概念



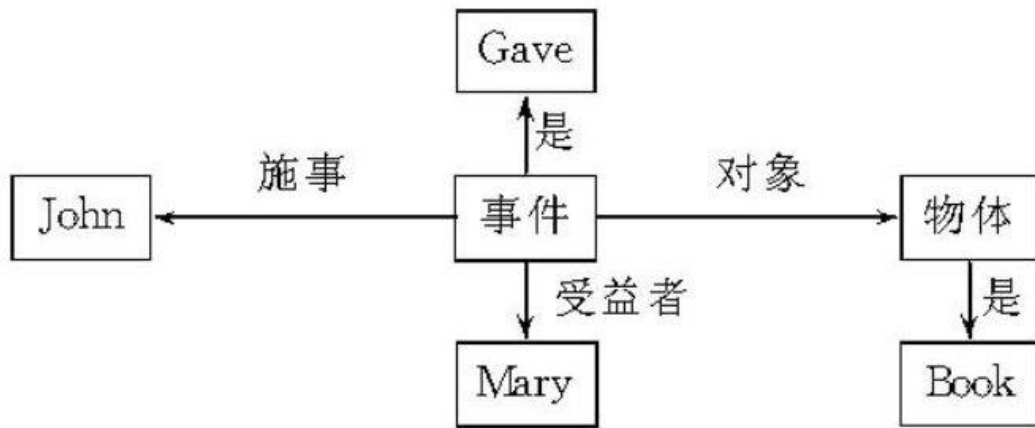
Vannevar Bush 教授
(顶级大佬，头衔太多不一一列举了)

- 如何将构想变成现实

- 知识图谱的发展历程：从语义网络与本体论衍生而来



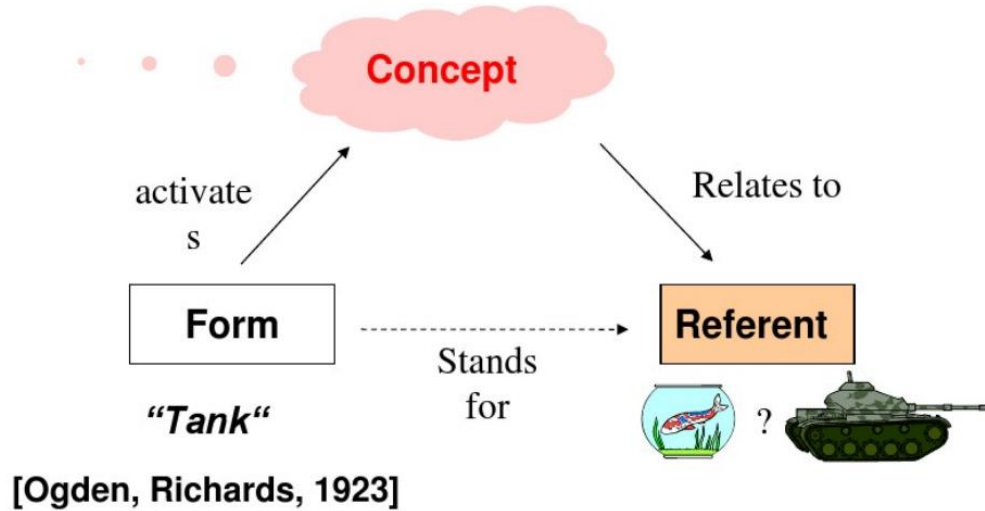
- **雏形理念：语义网络 (Semantic Networks)**
- 一种以有向图结构表达人类知识构造的形式
 - 部分内容在第三节课（“网页文字处理”）有所提及
 - 例如：同义词/相关词，上下位关系，WordNet等



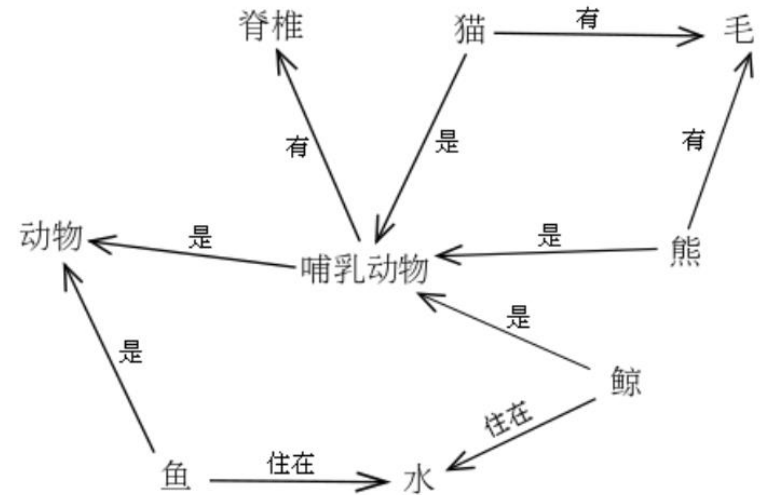
```
dog, domestic dog, Canis familiaris
=> canine, canid
=> carnivore
=> placental, placental mammal, eutherian, eutherian mammal
=> mammal
=> vertebrate, craniate
=> chordate
=> animal, animate being, beast, brute, creature, fauna
=> ...
```

• 雏形理念：本体论 (Ontology)

- 本体是指一种形式化的，对于共享概念体系的明确而又详细的说明
 - 指特定领域之中存在的对象类型或概念，及其属性和相互关系



符号三角形：符号、概念、事物三者关系



“哺乳动物”概念与相关事物形成的语义网络

- **雏形理念：本体论 (Ontology)**

- 本体论中的五元组表示法： $O = \{C, R, F, A, I\}$
 - C – Concept, 概念集合, 通常以术语形式组织, 如 “中科大学学生”
 - R – Relations, 描述概念或实例间语义关系, 如 “同班同学”
 - F – Functions, 一组特殊关系, 其中第N个元素由其他n-1个元素决定
 - 如, 作业的得分由题目难度、完成情况、格式规范性等因素决定
 - A – Axioms, 公理, 例如A是B的父亲, B是C的父亲, 得到A与C的关系
 - I – Instances, 具体的个体, 如 “助教” 是 “中科大学学生” 的实例

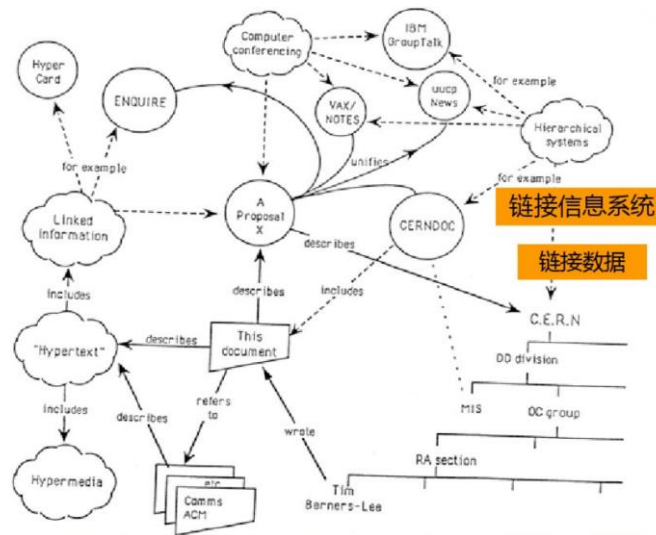
- **雏形理念：语义网 (Semantic Web)**

- 1998年，由Tim Berners-Lee（又是他！）提出

- 核心思想：通过给万维网上的文档（如HTML、XML文档）添加能够被计算机所理解的语义“元数据”（Meta data），从而使整个互联网成为一个通用的信息交换媒介。

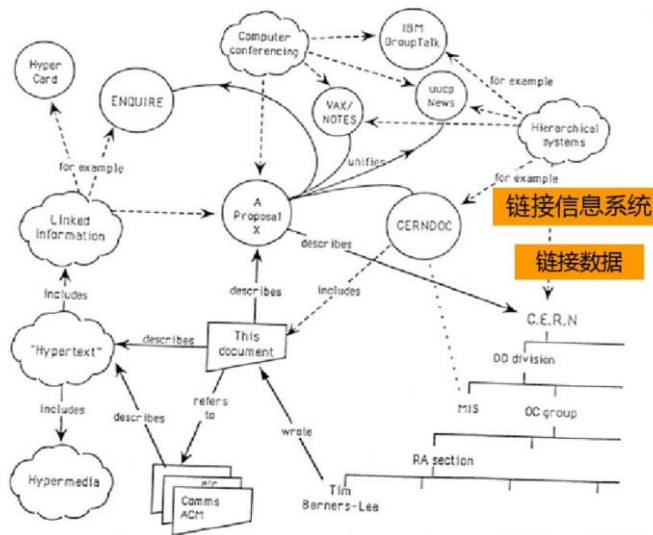
- 重要理念：信息应以 **图** 的形式组织，图中节点表达任何事物，链接表达事物之间的关联

- 以 **图** 为中心的信息组织方式可以有效提升检索效率，并在开放网络环境中快速增长



- 雏形理念：语义网 (Semantic Web)

- 1998年，由Tim Berners-Lee (又是他!) 提出
- 基本的万维网 (WWW) 和语义网络的区别？核心区别：服务对象不同
 - 万维网在组织信息资源时主要面向“人”的信息发布和获取，侧重于可视效果，即信息的显示格式和样式 (如HTML)。
 - 而语义网必须侧重于信息的语义内容，并考虑计算机对文本内容的“理解”以及它们之间的相互交流和沟通。



• 形形色色的语义网



- 知识组织形式的发展
- **信息抽取任务概述**
- 知识图谱的基本概念
- 更多类型的知识图谱
 - 事理图谱
 - 多模态知识图谱

- **信息抽取的含义**
- 从语料中抽取指定的事件、事实等信息，形成结构化的数据
 - 被抽取的信息以预先定义的、结构化的形式描述。
 - 为后续的情报分析、自动文摘、问答系统等等一系列应用提供服务。
 - 从整体文档中抽取细粒度、结构化信息！

凯里·欧文

篮球运动员



凯里·安德鲁·欧文为美国职业篮球运动员，场上主打控球后卫，现效力于NBA布鲁克林篮网。在2011年NBA选秀中以第一顺位被克利夫兰骑士选中，2016年为球队夺得NBA总冠军。在2016年夏季奥林匹克运动会代表美国参与男子篮球比赛夺得金牌。 [维基百科](#)

出生信息：1992年3月23日（30岁），澳大利亚墨尔本

身高：1.88米

臂展：1.93米

子女：阿兹乌里·伊丽莎白·欧文

父母：德雷德里克·欧文，伊丽莎白·厄文

兄弟姐妹：伦敦·欧文，艾莎·欧文

电影



德鲁大叔
2018年

Uncle Drew:
Chapter 4

空中大灌篮
2: 新传奇
2021年

Uncle Drew:
Chapter 2

信息抽取是整合与分析的基础



↑ 企业关联

↓ 竞争分析

天眼查 TianYanCha.com 国家中小企业发展基金旗下 官方备案企业征信机构

查公司 查老板 查关系

北京百度网讯科技有限公司

天眼一下 VIP会员

公司背景 368 司法风险 999+ 经营风险 67 公司发展 999+ 经营状况 999+ 知识产权 999+ 历史信息 999+

竞品信息 2234

序号	产品名称	当前融资轮次	估值	成立日期	产品标签	所属地	简介
1	魅闪科技	A轮	-	2019-04-29	人工智能	广东	网红打造系统研发商
2	元戎启行	Pre-A轮	-	2019-02-18	人工智能	广东	自动驾驶运营服务提供商
3	觉非科技	A轮	-	2019-01-30	人工智能	北京	自动驾驶技术研发商
4	蝌蚪有读	-	-	2019-01-29	企业服务	上海	IT服务商
5	深蓝爱	-	-	2019-01-04	人工智能	北京	智能办公系统研发商

- **信息抽取与信息检索**

- 两者密切相关，却又存在鲜明差异

- 功能不同

- 检索：从文档集合中找文档子集
- 抽取：从文本中获取用户感兴趣的事实信息

- 处理技术不同

- 检索：可以采用统计与关键词等技术
- 抽取：需要借助自然语言处理技术

- 使用领域不同

- 检索：通常领域无关
- 抽取：通常领域相关（借助领域知识辅助抽取）



- **信息抽取的内容**

- 核心的8字方针：“抽取实体，确定关系”

- 实体：即命名实体，指文本中的基本构成块，如人、机构等
- 属性：实体的特征，如人的年龄、机构的类型等
- 关系：实体之间存在的联系，也称事实，如公司和地址之间的位置关系、公司与人之间的雇佣关系
- 事件：实体的行为或实体参与的活动

- **信息抽取的基本任务**
- MUC (Message Understanding Conference) 会议
 - 美国国防部研究计划署 (DARPA) 资助
 - 是否还记得这家神奇的机构? 1969年, 互联网雏形ARPANet由此诞生
 - 该会议主要测评信息抽取系统, 自87年共举行7次 (MUC-1...MUC-7)
- 在MUC-7 (1998年) 上, 定义了5类基本的信息抽取任务
 - 命名实体NE、模板元素TE、共指关系CR、模板关系TR、背景模板ST

- **信息抽取的基本任务**
- **命名实体NE** (实体抽取)
- 命名实体抽取是信息抽取最重要的任务
- 命名实体是文本中基本的信息元素，是正确理解文本的基础
 - 狭义：指现实世界中具体或抽象的实体，如人、组织、地点等
 - 如：水果摊/Org, 老板 郝哥/Person
 - 广义：还可以包含日期和时间、数量表达式等

- **信息抽取的基本任务**
- **模板元素TE** (属性抽取)
- 模板元素又称为实体的属性，目的在于更加清楚、完整地描述命名实体
- 通过槽 (Slots) 描述了命名实体的基本信息
 - 槽：名称、类别、种类等
 - 例如：郝哥表示，这都是**大棚的瓜**，你嫌贵我还嫌贵呢。
 - ◆ TE：**瓜是大棚里产的** (属性)

- 信息抽取的基本任务

- 共指关系CR

- 如果不同的命名实体表达了相同的含义，即为共指关系，也称为等价概念
- 共指关系的抽取任务在于抽取关于共指表达的信息
 - 包括那些已在命名实体和模板元素任务中作了标记的，对于某个命名实体的所有表述
 - 例如：我开水果摊儿的，能卖给你生瓜蛋子啊.....要是不熟，我自己吃了它，满意了吧？
 - ◆ CR：生瓜蛋子和它均代指 郝哥卖的瓜

- **信息抽取的基本任务**
- **模板关系TR** (关系抽取)
- 实体之间的各种关系, 又称为事实
- 通过关系抽取, 将实体关联起来, 并为推理奠定基础
 - 例如, 职务 (Post_of)、雇佣关系 (Employee_of)、生产关系 (Product_of) 等
 - 如:
 - Post_of (老板, 郝哥)
 - Employee_of (水果摊, 郝哥)

- **信息抽取的基本任务**
- 场景模板ST (事件抽取)
- 又称事件, 是指实体发生的事件
 - 例如: 会议 (Time<...>, Spot<...>, Convener<...>, Topic<...>)
 - 常见的新闻事件描述模板 5W1H
 - Who、When、Where、What、Why、How
 - 例如: 那天 (When), 有一个人 (Who) 来水果摊 (Where) 买瓜, 因为秤有问题 (Why), 用西瓜刀 (How) "萨日朗" (What) 了郝哥

- **信息抽取的基本任务**

- 一个正儿八经的信息抽取实例：人民日报1998-01-07

19980107-06-016-001意大利总理普罗迪 4 日说，欧洲国家将采取行动，共同对付库尔德难民涌入问题。普罗迪 4 日晚召开了由意外长、内政和国防部长参加的紧急会议，商讨应付库尔德难民问题的对策。会前，普罗迪说，“在经过最初的混乱后，欧洲国家的行动已经大大加强”，今后几天内将在此问题上进行系统合作。

- 信息抽取的基本任务

- NE实体抽取结果示例

<NamedEntities>

<PersonList>

库尔德 (occurrence: 1/1/15; 1/2/19;) (类似于倒排表的形式, 但表示在文中的位置)

普罗迪 (occurrence: 1/1/3; 1/2/0; 1/3/2;)

</PersonList>

<OrgList></OrgList>

</NamedEntities>

- TR关系抽取结果示例

<EntityRelations>

post_of(意大利总理, 普罗迪)

</EntityRelations>

- 信息抽取的基本任务

- ST事件抽取结果示例

<EventTemplateInstatnces>

<ConferenceInfo>

<Time> 4 日晚 (1998-01)</Time>

<Spot> 意大利</Spot>

<Converner> 普罗迪</Converner>

<Title>由意外长、内政和国防部长
参加的紧急会议

</Title>

</ConferenceInfo>

</EventTemplateInstatnces>

会议时间 Time	4 日晚 (1998-01)	
会议地点 Spot	意大利	
召集人 Convener	姓名/团体名称 Name	普罗迪
	机构、职位 Org/Post	意大利总理
会议名 / 标题 Conf-Title	由意外长、内政和国防部长参加的紧急会议	

- 知识组织形式的发展
- 信息抽取任务概述
- **知识图谱的基本概念**
- 更多类型的知识图谱
 - 事理图谱
 - 多模态知识图谱

- 知识图谱的诞生

- 2012年5月16日，Google知识图谱（Knowledge Graph, KG）正式发布

- 除了显示网页文档的连接列表外，还提供结构化的、详细的有关主题的信息。

- 其目标是，用户可以使用此功能直接解决查询的问题，而不必导航到其他网站并自行汇总信息。

The screenshot shows a Google search for "obama birthday". The search results include a Knowledge Graph card for Barack Obama, displaying his birth date as August 4, 1961 (58 years old), and a list of related search terms like "特朗普" (Trump), "米歇尔·奥巴马" (Michelle Obama), and "乔治·W·布什" (George W. Bush). To the right, a detailed information box for Barack Obama is visible, including his full name, title as former president, and biographical details.

Google obama birthday

全部 图片 新闻 地图 视频 更多 设置 工具

找到约 80,800,000 条结果 (用时 0.54 秒)

贝拉克·奥巴马 / 出生日期

1961 年 8 月 4 日 (58 岁)

用户还搜索了

- 特朗普 1946 年 6 月 14 日
- 米歇尔·奥巴马 1964 年 1 月 17 日
- 乔治·W·布什 1946 年 7 月 6 日

反馈

贝拉克·奥巴马 (Barack Obama) 前美国总统

贝拉克·奥巴马，全名贝拉克·侯赛因·奥巴马二世，是美国政治人物，从2009年至2017年任第44任美国总统。他是首位拥有非裔血统的美国总统。2005年至2008年代表伊利诺伊州担任美国联邦参议员，从1997年至2004年担任伊利诺伊州参议员时居于芝加哥。卸任总统后，定居首都华盛顿特区。 [维基百科](#)

生于：1961 年 8 月 4 日 (58 岁)，夏威夷州檀香山 Kapi'olani Medical Center for Women & Children

全名：Barack Hussein Obama II

身高：6' 1"

总统任期：2009 年 1 月 20 日 - 2017 年 1 月 20 日

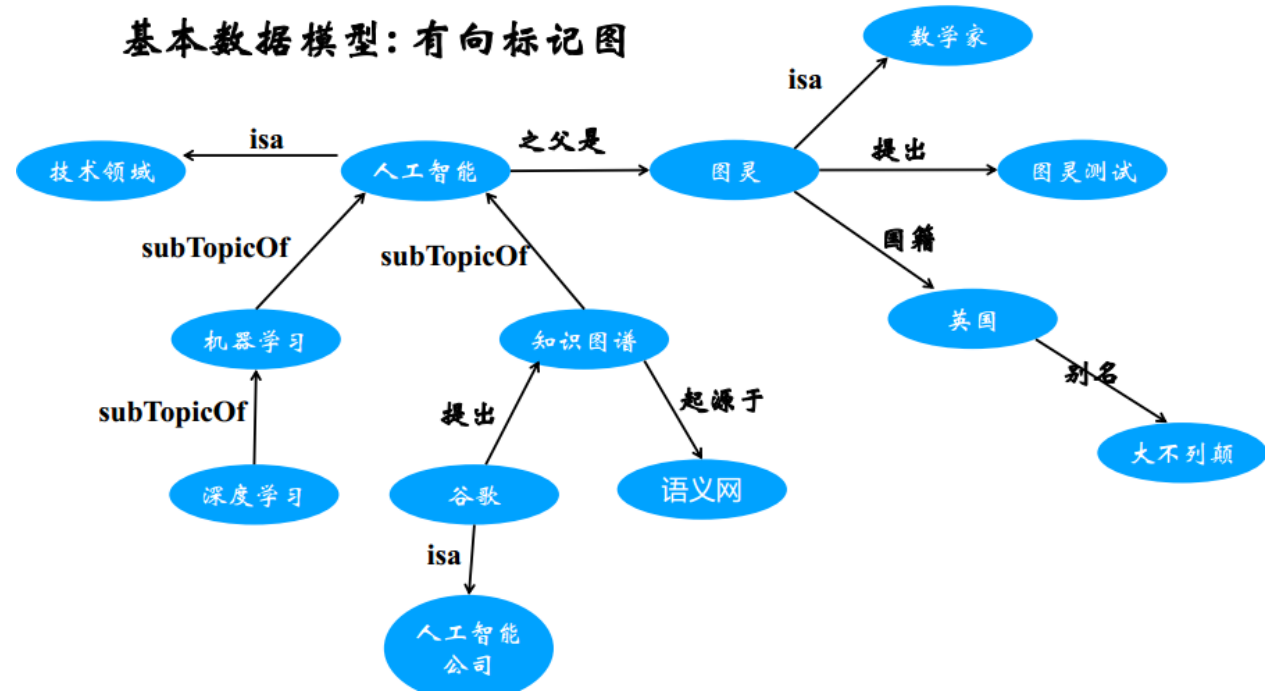
- 知识图谱的符号表示

- 由前述各种网络衍生而来，知识图谱呈现出类似的形式

- 由结点和结点之间的边组成，
结点表示概念（或实体），
边表示关系（或属性）。

- 在数学上，知识图谱表现为一个有向图

- 方向表示主被动关系，表达相互关系请用双向图

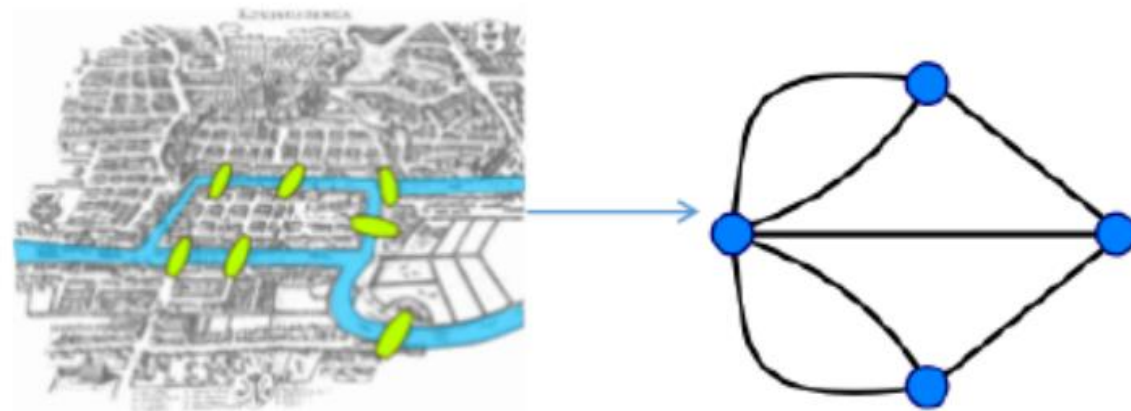


- **知识图谱的基本元素：节点**
- 一般而言，知识图谱中的节点用来表示概念 (Concept) 和实体。
- 实体 (Entity / Object / Instance)
 - 能够独立存在的，作为一切属性的基础和万物本原的东西 —— 黑格尔



- **知识图谱的基本元素：边**

- 一般而言，知识图谱中的边用来表示关系（Relation）和属性（Attribute）。
 - 关系：侧重实体（Entity）之间的关联，例如“**高王**”：姚明**高王**小四
 - 属性：用于描述实体的特征，例如尺寸，颜色、组成等等
- 点和边组成知识图谱的基本单位：**三元组**（头实体-关系-尾实体）



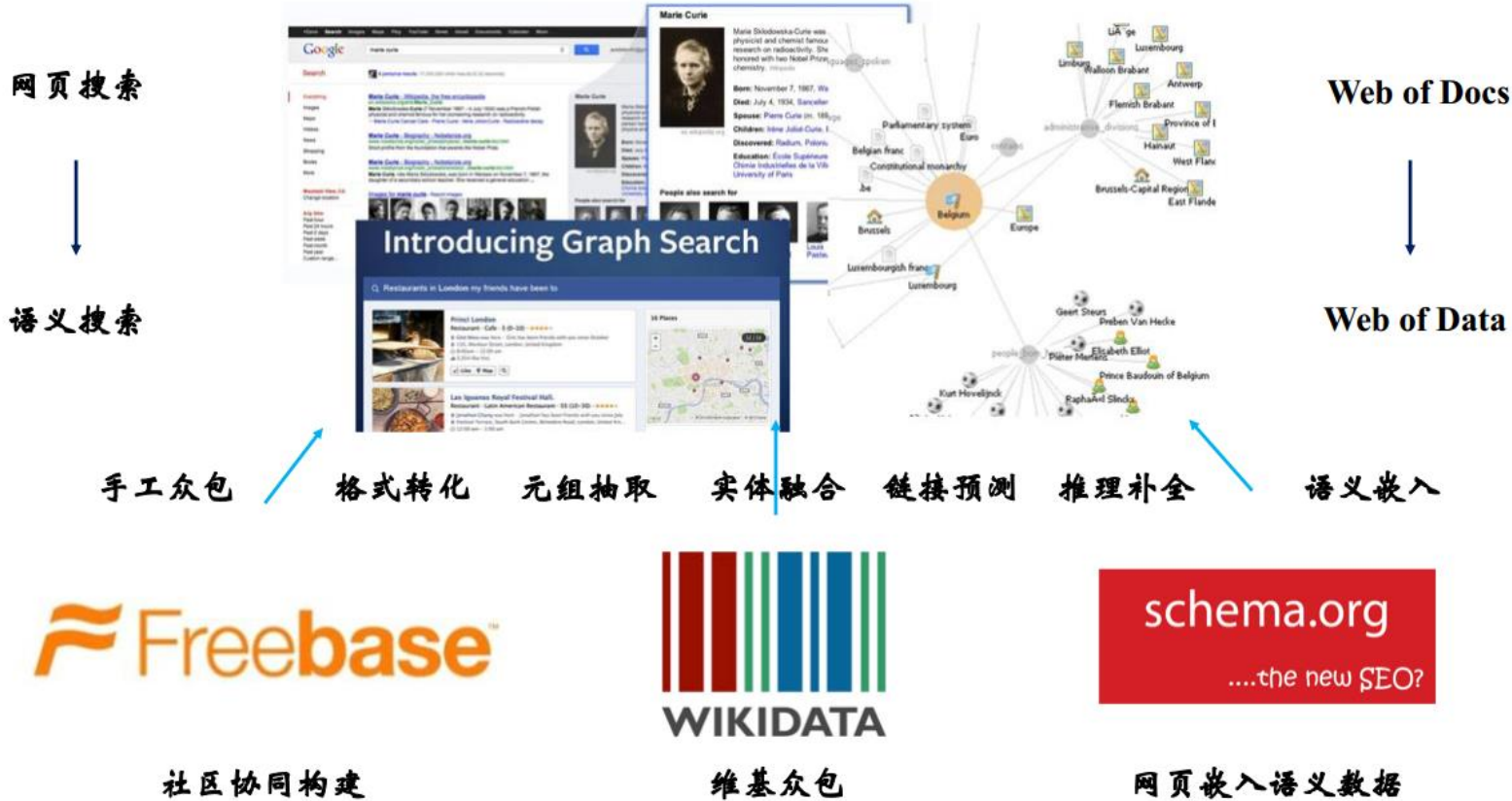
- **知识图谱的优点**

- 知识图谱至少可从以下三个层面提升搜索的效果：
 - 找到最想要的信息：不再需要用户自行浏览、阅读和总结，而将信息直接呈现
 - 提供最全面的摘要：对搜索对象进行总结，使得用户获得更完整的信息和关联
 - 让搜索更有深度和广度：构建完整知识体系，使用户获得意想不到的新发现



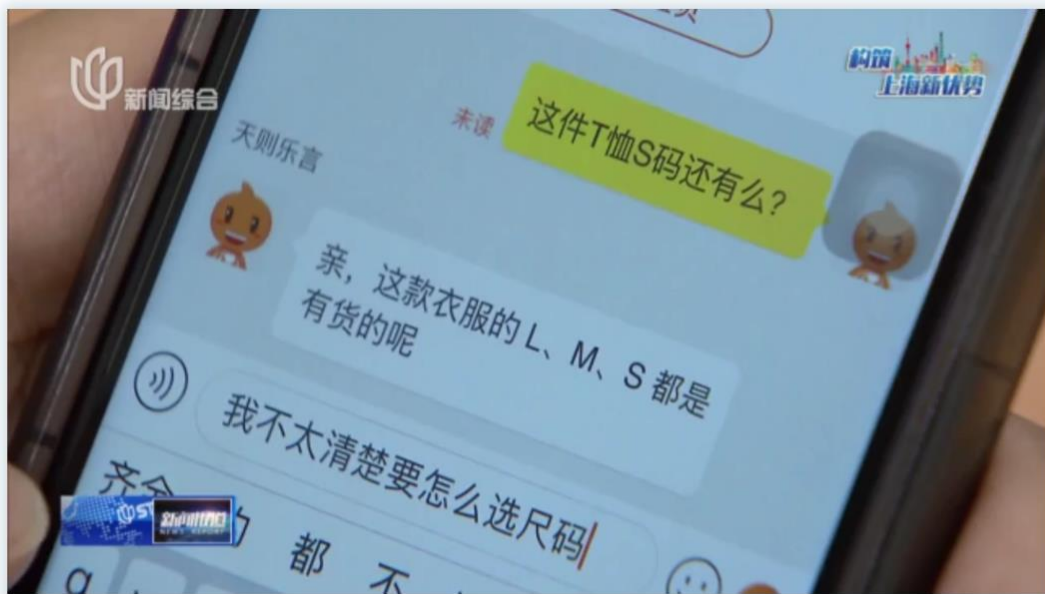
知识图谱相关应用 (1) : 语义搜索

通过建立事物之间的联系, 实现更准确、更直接、更完整的搜索



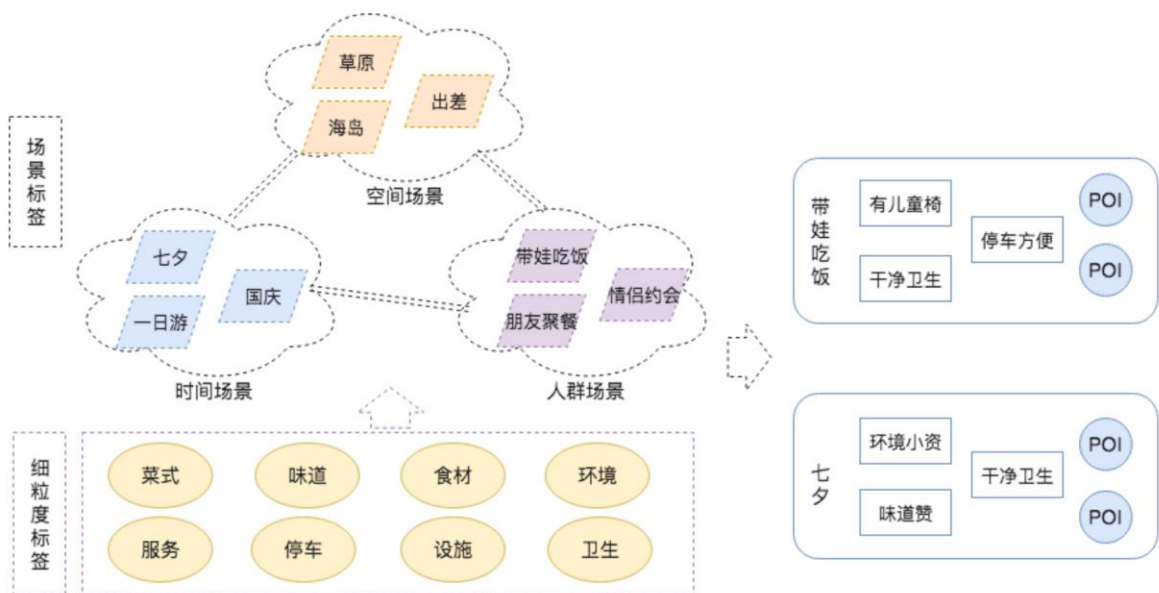
- **知识图谱相关应用（2）：问答系统**

- 在理解用户意图的基础上，将知识图谱作为大脑，基于推理能力提升交互体验



• 知识图谱相关应用 (3) : 推荐系统

• 利用知识图谱提高推荐系统的推荐多样性和可解释性, 提升推荐性能



用户查询理解



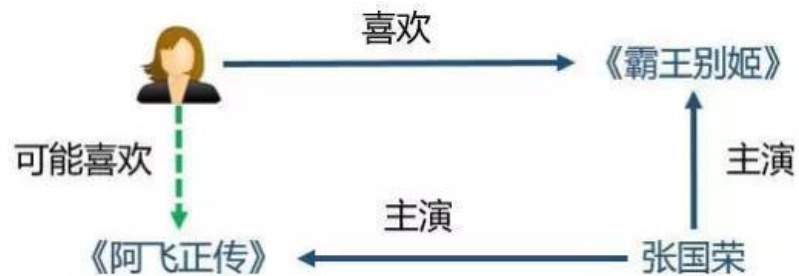
场景推荐



知识图谱相关应用 (3) : 推荐系统

引入与物品相关的知识，提升精确性和多样性

精确性：知识图谱为物品引入了更多的语义关系，可以深层次地发现用户兴趣。



多样性：通过知识图谱中不同的关系链接种类，有利于推荐结果的发散。



知识图谱工程技术路线

- 高关联、规范化的知识图谱极大提升了数据质量，在各个垂直领域都发挥着重要作用，被各大互联网大厂视为人工智能“新基建”。



- **公开的知识图谱数据**

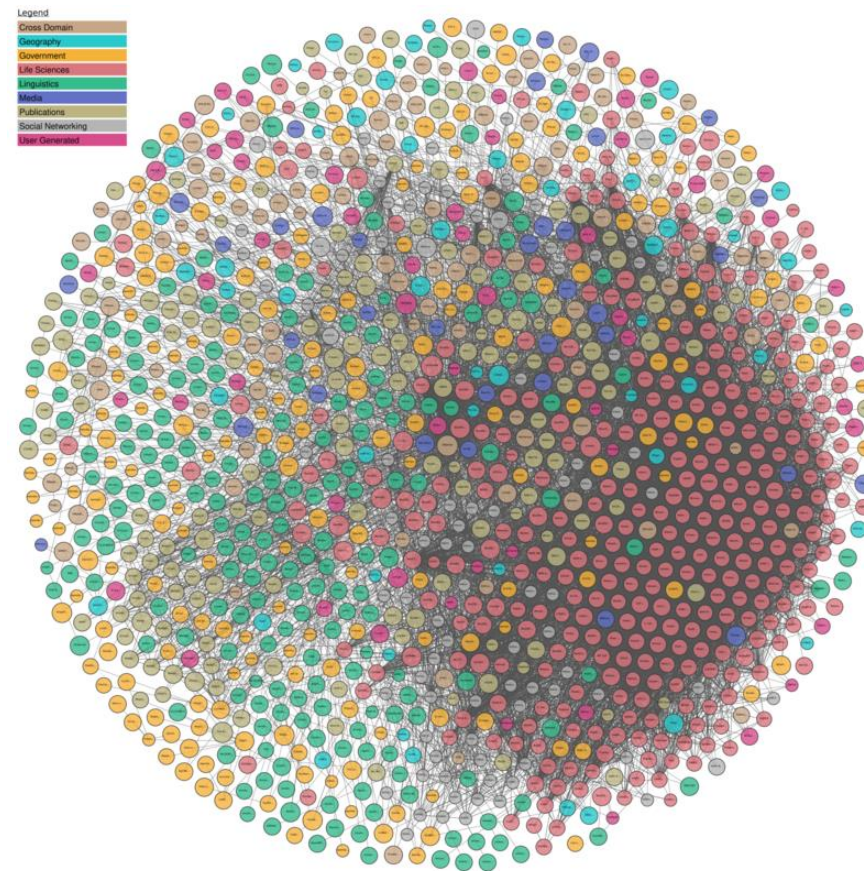
- 目前，已有多个通用或专业知识图谱开放，典型的中文KG知识图谱有：
 - CN-Dbpedia：由复旦大学知识工场实验室研发并维护的大规模通用领域结构化百科，包含1700万个实体，2亿个三元组
 - Zhishi.me：通过从开放的百科数据中抽取结构化数据，包含1000万个实体，1.2亿个三元组
 - XLORE：由清华大学开放，从异构的跨语言在线百科中抽取结构化信息，包含1600万个实体，44万个属性
- 通过OpenKG (<http://www.openkg.cn>)，可以找到丰富的开放图谱及工具。

The logo for CN-DBpedia, featuring the text "CN-DBpedia" in white on a dark grey rectangular background.The logo for Zhishi.me, featuring the text "ZHISHI" in blue with white dots above the letters, and "me" in white on a blue background.The logo for XLORE, featuring the text "XLORE" in multi-colored letters (X in blue, L in green, O in red, R in yellow, E in green) and a blue square with a white 'X' on the right. Below the text is the Chinese text "中英文跨语言百科知识图谱".

- **公开的知识图谱数据**

- 一些更大规模的数据：

- Linked Life Data是面向医疗领域的大规模知识库，包括基因、蛋白质、分子反应、药物等方面的103亿个RDF三元组。
- 开放数据云图（Linked Open Data cloud diagram）显示了已经被发表为关联数据格式的数据集。已经包含1239个数据集，含有16147个链接（2019年3月）



<http://linkedlifedata.com/>

<http://lod-cloud.net/>

- **相关会议与测评**

- 全国知识图谱与语义计算大会（CCKS）
 - 已成为国内知识图谱、语义技术、链接数据等领域的核心会议

2022全国知识图谱与语义计算大会

China Conference on Knowledge Graph and Semantic Computing – 秦皇岛 – 2022年

- CCKS每年组织有关知识图谱与语义计算相关的测评，2022年的测评包括：
 - 千言通用信息抽取竞赛评测
 - 基于知识图谱的优质文章识别
 - 带条件的分层级多答案问答
 -

- 知识组织形式的发展
- 信息抽取任务概述
- 知识图谱的基本概念
- **更多类型的知识图谱**
 - 事理图谱
 - 多模态知识图谱

- **事理图谱 (Event Logic Knowledge Graph)**
- 哲学问题：这个世界是怎样构成的？我们人类是怎样认识这个世界的？
 - 知识图谱：世界由物质（实体）及其之间的关系构成
 - 另一种思路：世界由**事实**构成，而事实包含物质和事件，人类通过事件来认识世界
 - 相比于前者，后者有着更好的 **动态性**

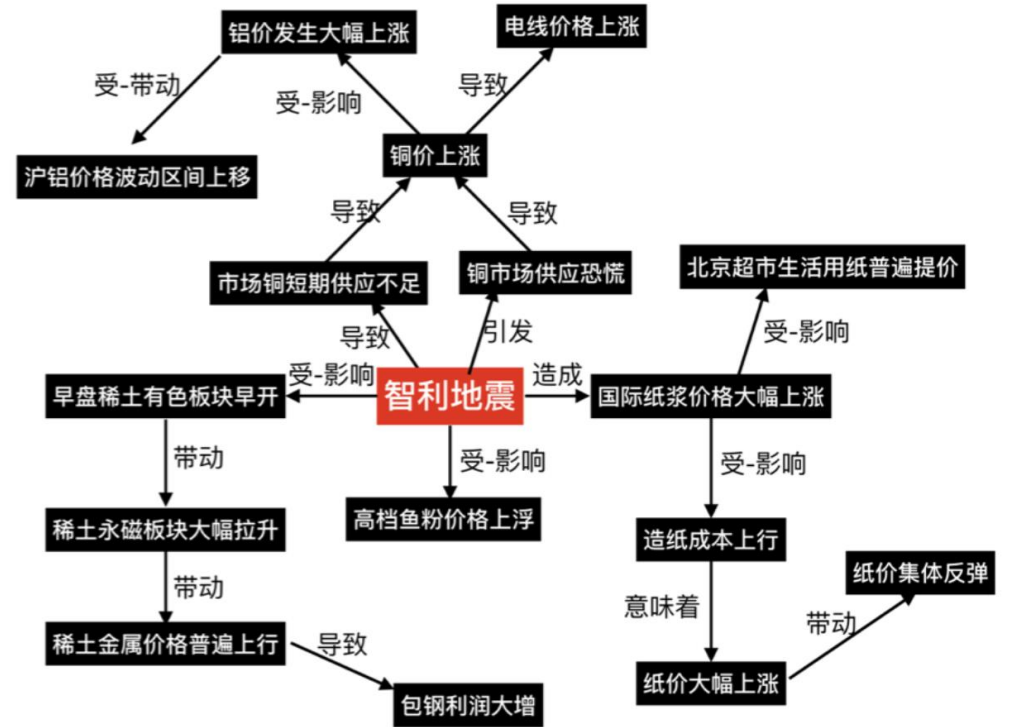


“世界是事实的总和，而非物质的总和”。

——维特根斯坦《逻辑哲学论》

• 事理图谱的价值

- 事理图谱所要描绘的是一个逻辑社会，研究对象是谓词性事件及其内外联系。
- 事理图谱与知识图谱的组织形式相仿，实体通过头尾相连，形成有向图的组织性质
- 借助图谱中的事理逻辑链接，可以形成对于事件的推理



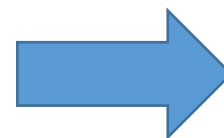
智利地震事件因果图谱（部分）

- **事理图谱的定义**

- 事件, 比如: “结婚”, “买房”
- 事理逻辑 (事件之间的演化规律与模式)

比如: “结婚” - (导致) > “买房”

- 事理图谱 = 事件+逻辑
 - 其本质是一个事理逻辑知识库
 - 不仅呈现了事件本身的发生, 还揭示了它们之间潜在的发展关系和影响链条

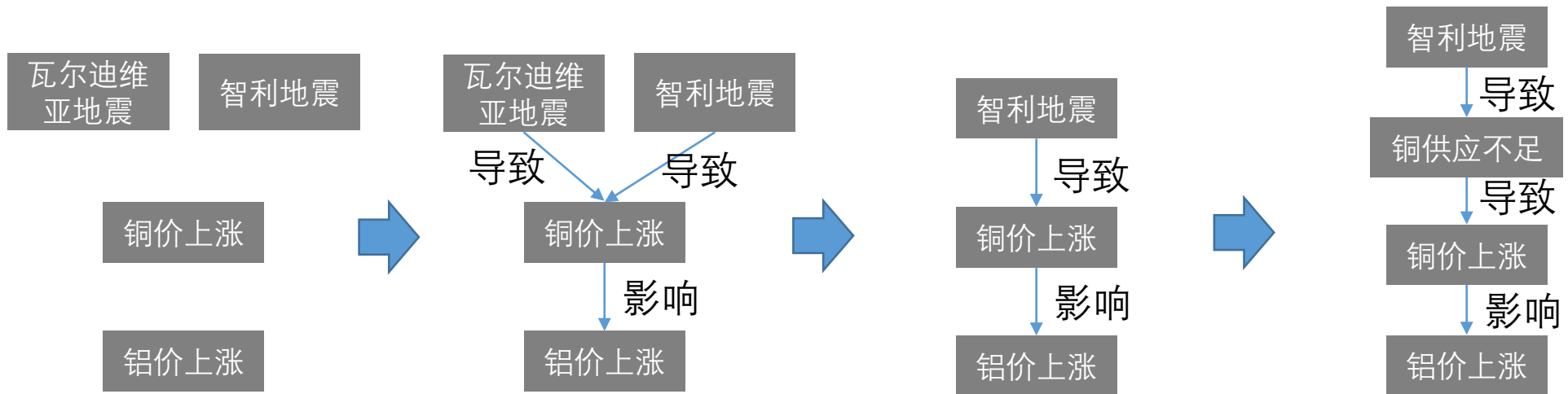
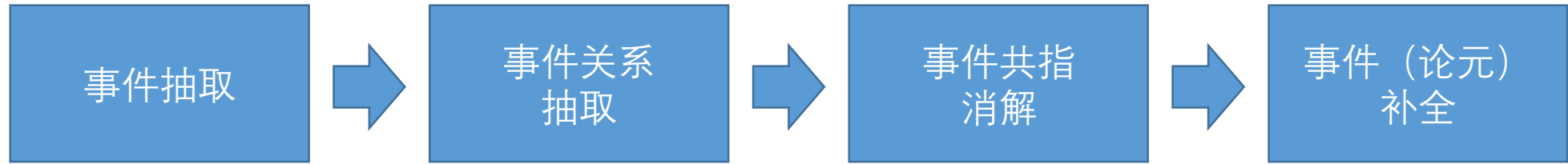


• 事理图谱VS知识图谱

- 差异主要体现在描述内容、结构和表示、应用领域和时间维度上

	知识图谱	事理图谱
描述知识	万物本体	逻辑社会
研究对象	名词性实体及其属性、关系	谓词性事件及其内外（空间、时间域）联系
构建目标	万物互联	全逻辑库，逻辑演化模型
回答问题	When、Who、What、Where	Why、How
组织形式	有向图	有向图
知识形式	<实体，属性，属性值>,<实体，关系，实体> 三元组	<事件，论元集合，逻辑关系> 多元组
知识确定	事实是确定的	逻辑不确定，有转移概率
知识状态	相对静态，变化缓慢	动态的
知识敏感	精确性要求极高，实时性要求极高	可一定容错，参考逻辑
构建难点	知识本体的搭建、知识抽取与融合	事件的表示、事件的抽取；与知识图谱的融合

• 事理图谱的构建



• 事件之间的关系

事件	含义	形式化	举例
因果事理	某一事件导致某一事件发生	A导致B	<地震,房屋倒塌>
条件事理	某事件条件下另一事件发生	如果A那么B	<限制放宽,立即增产>
反转事理	某事件与另一事件形成对立	虽然A但是B	<起步晚,发展快>
顺承事理	某事件紧接着另一事件发生	A接着B	<去旅游,买火车票>
上下位事理	某事件是另一事件的上位或下位事件	A是B的一类	<地震,地质灾害>
组成事理	某事件是另一事件的组成部分	A组成B	<灭火,火灾救援>
并发事理	某事件与另一事件同时发生	A同时B	<睡觉,闭眼>

- 事件之间的关系

- 顺承关系

- 两个事件在时间上的相继发生的偏序关系
- 通过转移概率来表示从一个事件按照时序关系顺承到下一个事件的置信度

“吃过午饭之后，小明到前台买单，然后离开了餐馆”



- **事件之间的关系**

- 因果关系

- 前一个事件（原因）导致后一个事件（结果）
- 满足发生时间上的偏序约束→属于顺承关系
- 通过因果强度值（0到1之间）来表示因果关系的置信度

“核泄漏引起了严重的海洋污染”

核泄漏

导致

海洋污染

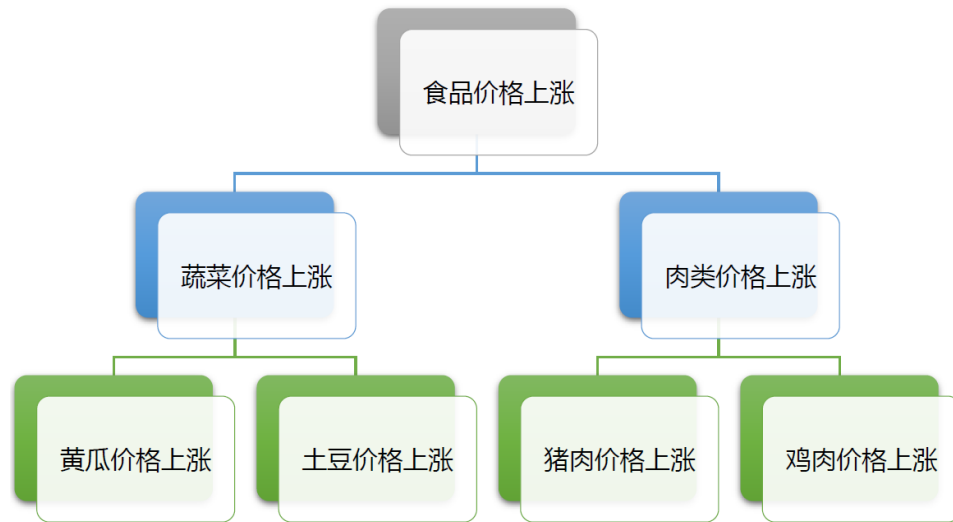
- 事件之间的关系

- 上下位关系

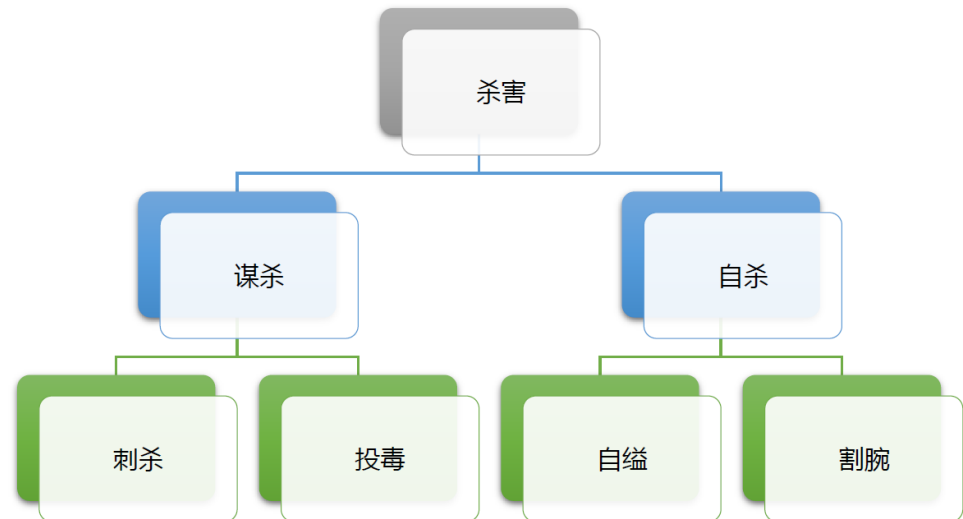
- 表示两个事件间的包含，从属等关系

- 分为名词性上下位和动词性上下位

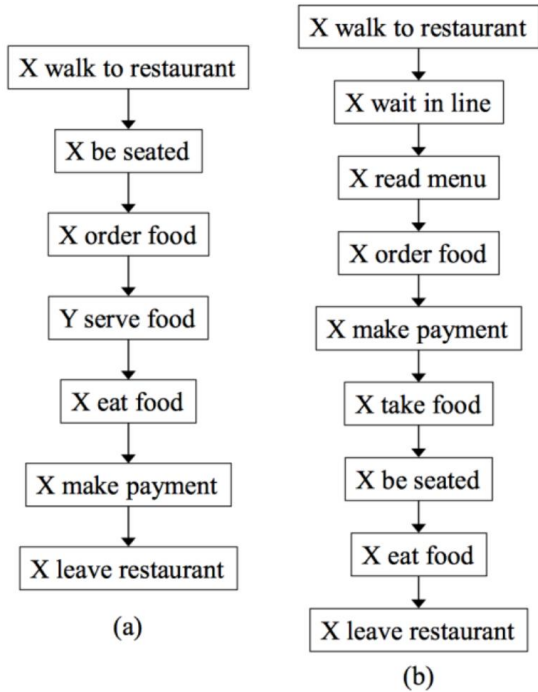
名词性上下位关系



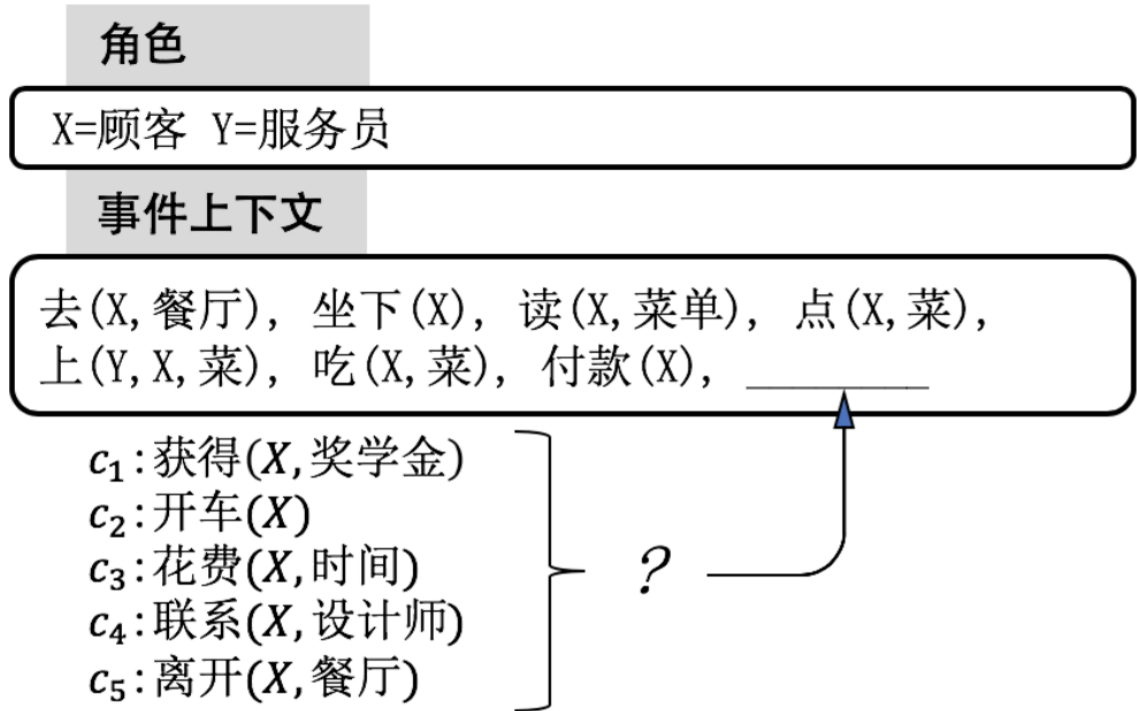
动词性上下位关系



- 更丰富的知识图谱类型
- 事理图谱的应用
 - 脚本事件 (script event) 预测



餐厅场景的脚本事件



餐厅场景的脚本事件预测

- **更丰富的知识图谱类型**
- 事理图谱的应用
 - 情报分析：
 - 有助于整合和分析大量的情报数据，提供全面的情报图景，支持决策制定和战略规划

巴勒斯坦以色列再爆发军事冲突

更新至2023年11月12日 07:16

全部 198

各方发声 59

伤亡状况 25

冲突进展 57

当地华人 5

时间倒序

- 今天 07:16 **新**
以色列总理称 Hamas 已失去对加沙地带北部的控制
- 昨天 14:21
内塔尼亚胡称不寻求占领、征服或统治加沙
- 昨天 12:42
加沙最大医院遭以军袭击
- 昨天 06:59
燃料耗尽 加沙地带奥达医院医生用手机照明做手术
- 11月10日 23:55
本轮巴以冲突已致双方至少12661人死亡

- **事理图谱上存在的不足**
- 关系的边界比较难以界定
- 对于事件关系的研究大量集中于因果关系等，对于其他关系的研究较少
- 事件的定义不够明确，从而引起抽取事件比较困难

事件	含义	形式化	举例
因果事理	某一事件导致某一事件发生	A导致B	<地震,房屋倒塌>
条件事理	某事件条件下另一事件发生	如果A那么B	<限制放宽,立即增产>
反转事理	某事件与另一事件形成对立	虽然A但是B	<起步晚,发展快>
顺承事理	某事件紧接着另一事件发生	A接着B	<去旅游,买火车票>
上下位事理	某事件是另一事件的上位或下位事件	A是B的一类	<地震,地质灾害>
组成事理	某事件是另一事件的组成部分	A组成B	<灭火,火灾救援>
并发事理	某事件与另一事件同时发生	A同时B	<睡觉,闭眼>



• 事理图谱无力解决的问题

我不敢苟同他的说法，首先我认为老坛酸菜面应该配三鹿奶粉，这样更能净化恒河水，但后果就是导致全球的金融危机，甚至富士山大爆炸，再说根据乘法口诀表推算出核弹中加满牛奶能威力更大，我俗称它大男孩，有人会觉得这观点偏激了，当然从达尔文的进化论可以得知牛顿被苹果砸中后发现了奶酪要用嘴吃，也可以得出路易十五与奥巴马视频聊天通话中谈到是先有鸡，还是先有蛋的话题，总体来说，这都不影响上帝与夏娃在一起创造银河 😊

2022-08-06 05:53 312 回复

我并不赞同你的观点，根据爱因斯坦提出的狭义相对论，我认为关谷健次郎在某种力的作用下，也可以达到田螺的口感，换句话说，只要有足够的张力，椅子是完全可以打电话给野比大雄，那么巴西的袋鼠问题就能得到完美的解决，这样也能间接的使用考拉的力量，使冰川融化，起到龙卷风摧毁停车场的的作用，某种意义上来说，只要高尔夫球的转速足够快，蜘蛛侠南天门大战蒋门神的状况也可以得到改善

2022-08-06 09:20 27 回复

仅仅依靠事理图谱并不能有效解析 / 辨识所有事物间的关系

如何有效结合 **常识** 判断逻辑的合理性？

- **事理图谱数据集**

- Event-QA: <https://github.com/tarcisiosouza/Event-QA>

包含 1000 个语义查询和相应的英语、德语和葡萄牙语的语言表

包含一个以事件为中心的知识图谱，包含超过 97 万个事件

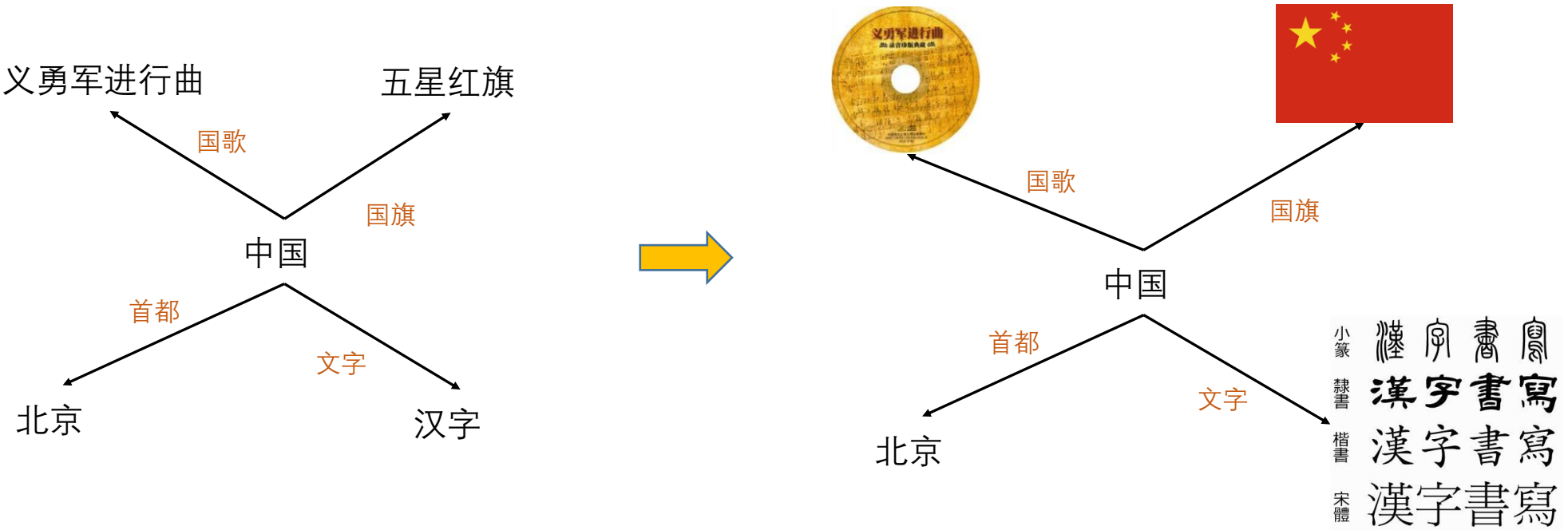
- ETKG: https://github.com/xcwujie123/Hainan_KG

以游客的活动为中心建立的出行事理图谱，从6166个匿名游客的旅游信息中提取

- 知识组织形式的发展
- 信息抽取任务概述
- 知识图谱的基本概念
- **更多类型的知识图谱**
 - 事理图谱
 - **多模态知识图谱**

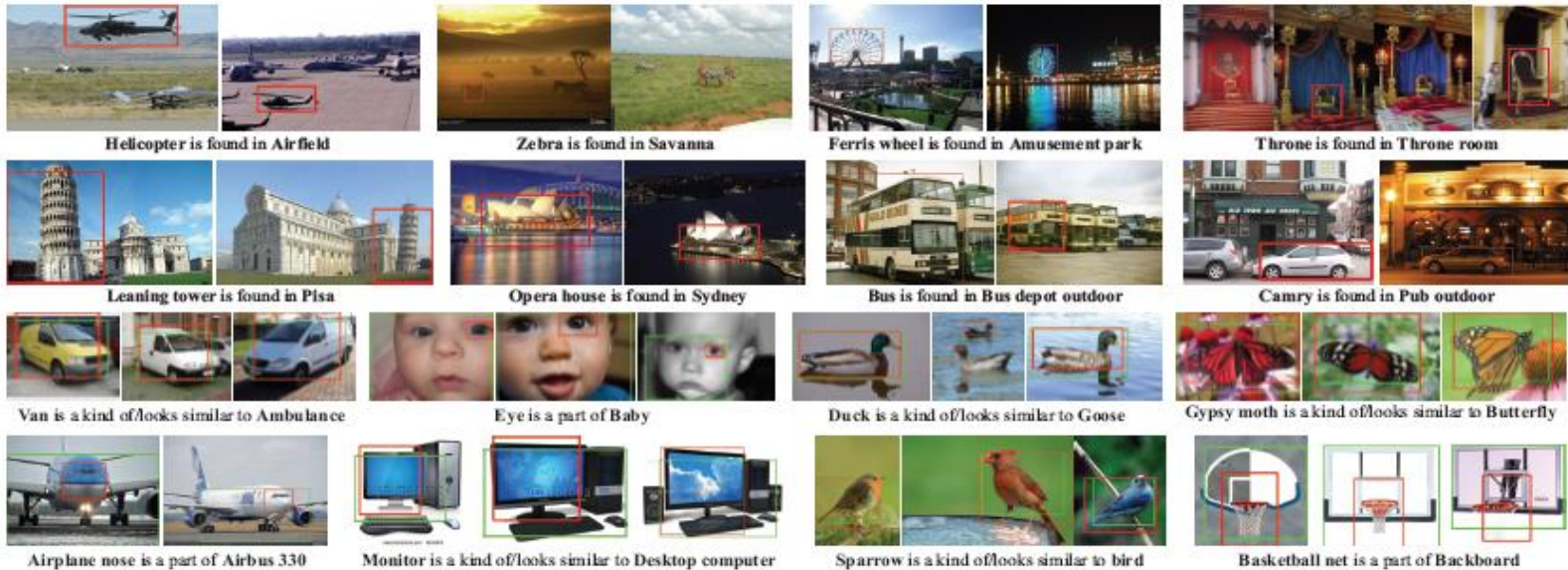
• 多模态知识图谱

- 模态：某种类型的信息或者存储信息的表示形式
- 传统图谱以文本模态为主，难以描述复杂的现实世界信息



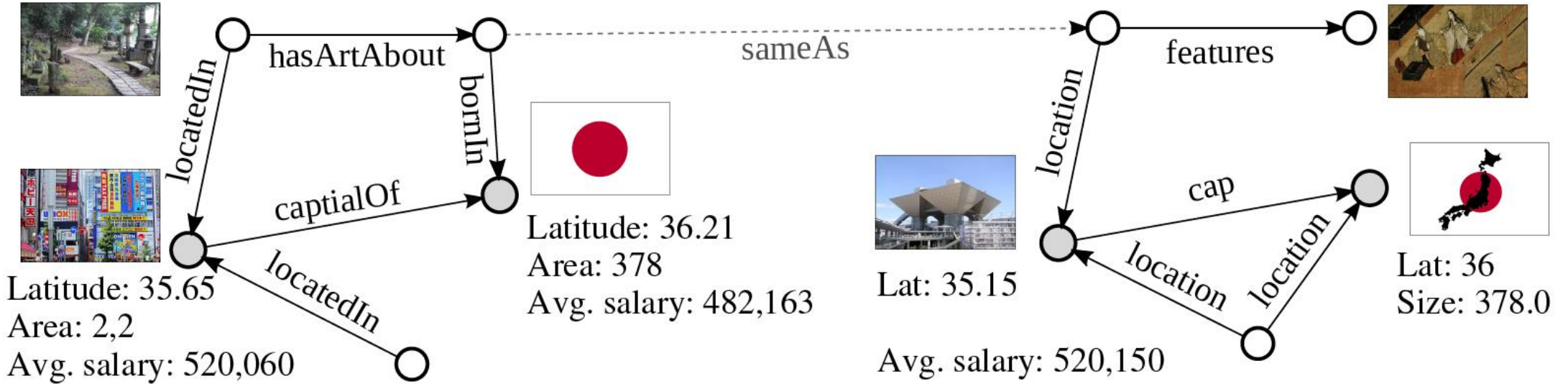
• 从视觉信息中挖掘事物的关系

- 2013年，CMU提出了名为NEIL的知识抽取程序，持续标注图片数据，并从中自动抽取实体间关联
 - NEIL = Never Ending Image Learner



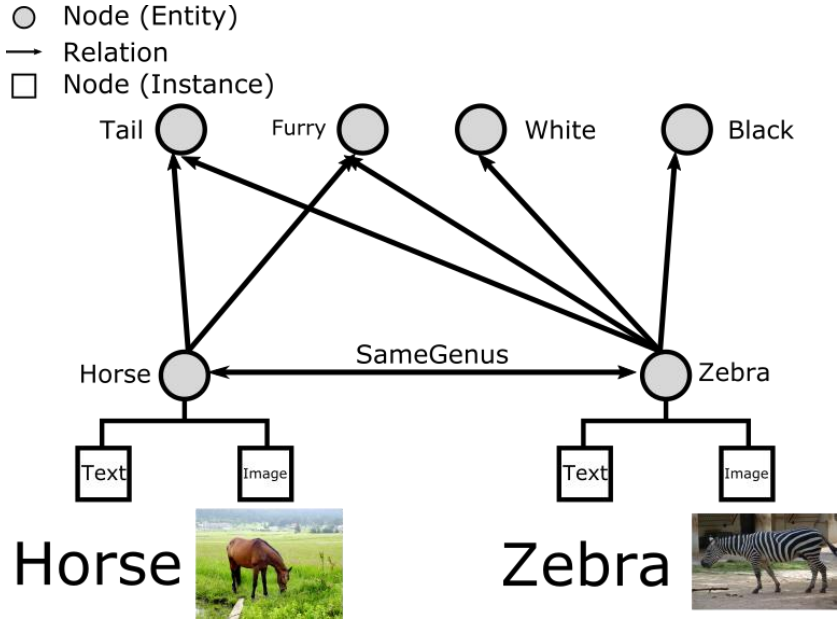
• 多模态知识图谱

• 现实世界语义模态日益丰富，有效表示与整合多模态知识成为趋势

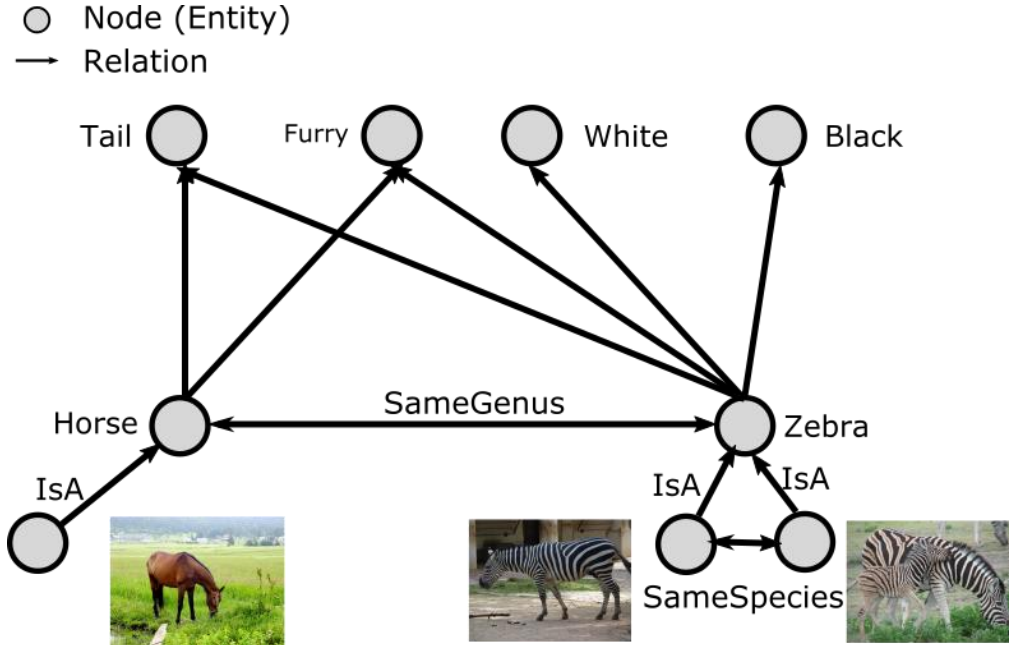


• 多模态知识图谱

• 多模态知识图谱可笼统分为**属性多模态**与**实体多模态**两大类

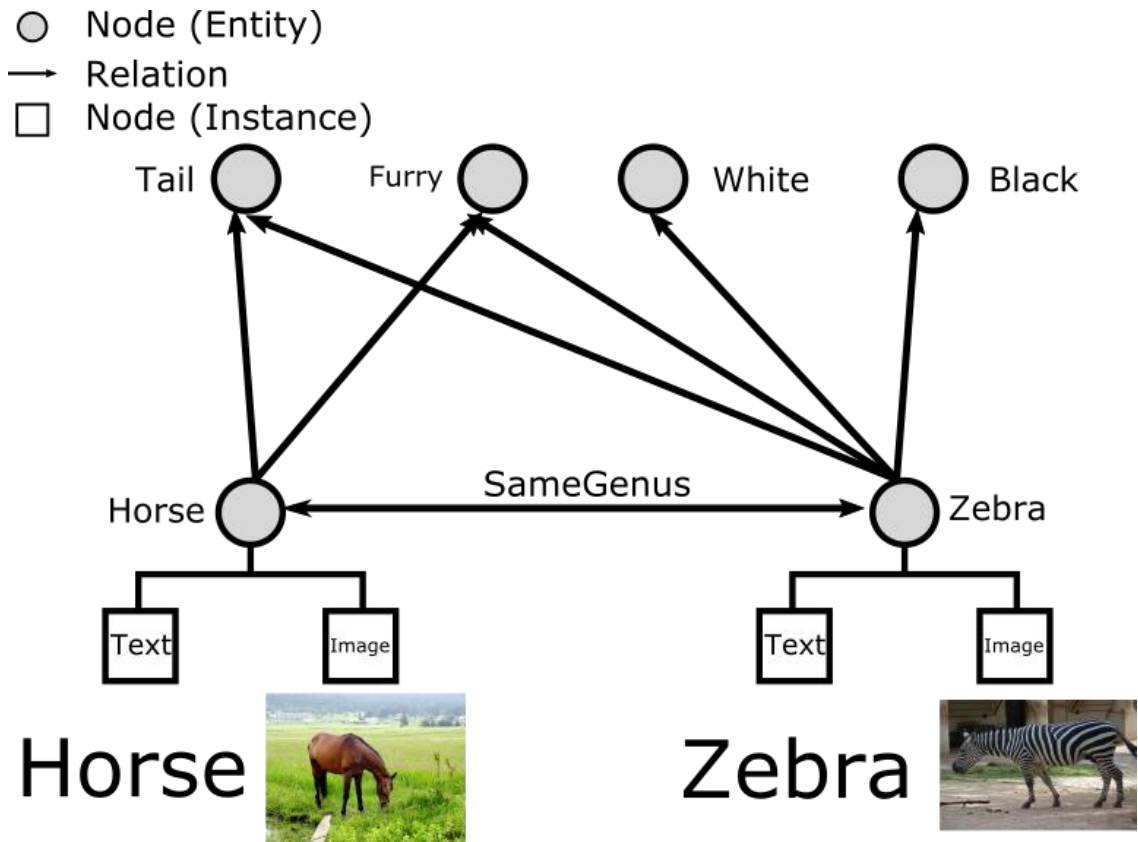


属性多模态知识图谱



实体多模态知识图谱

• 多模态知识图谱：属性多模态



属性多模态知识图谱

实体

- 抽象概念
- 具有多模态的表示

关系

- 实体之间的概念联系

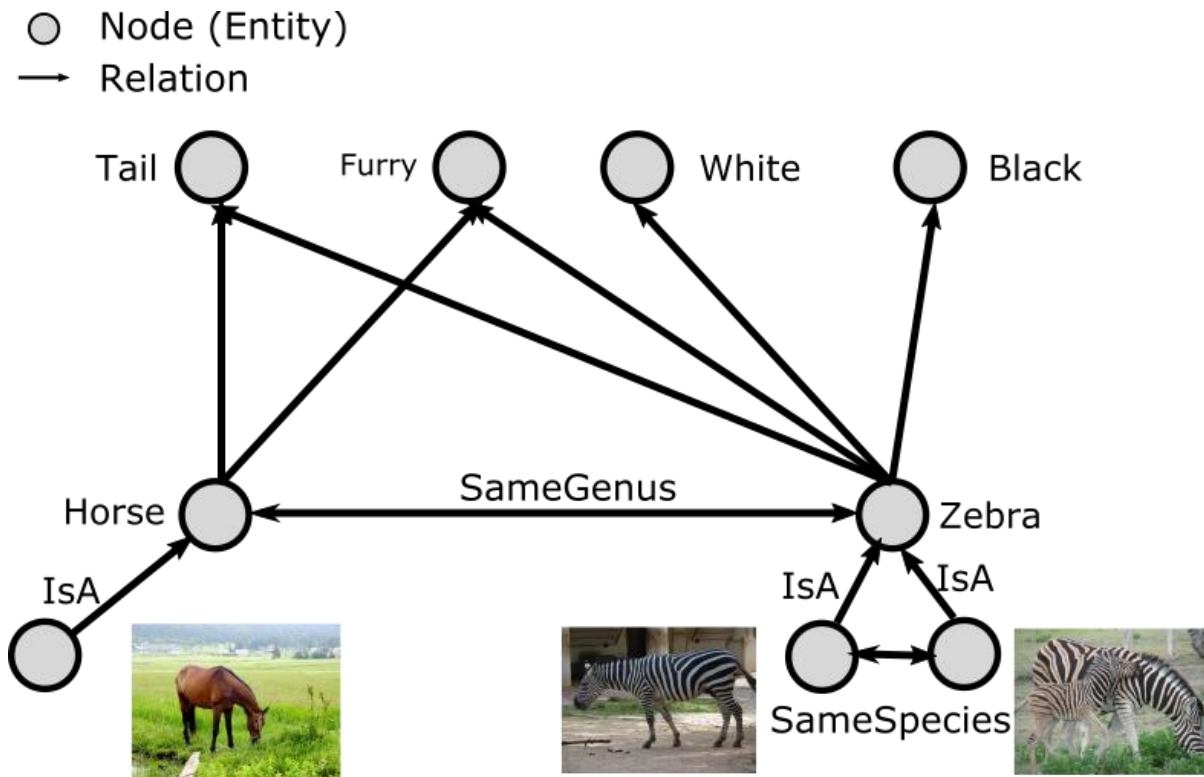
优点

- 易于从现有图谱中进行扩展，即通用图谱可以轻松扩展，概念和关系不用改变
- 可以推理视觉知识

缺点

- 实体仍仅限于文本概念描述
- 1个概念可以对应N张图片，但1张图片无法对应N个概念

• 多模态知识图谱：实体多模态



实体多模态知识图谱

实体

- 多模态的实体

关系

- 逻辑关系——相同模态
- 物理关系——跨模态
- 语义关系

优点

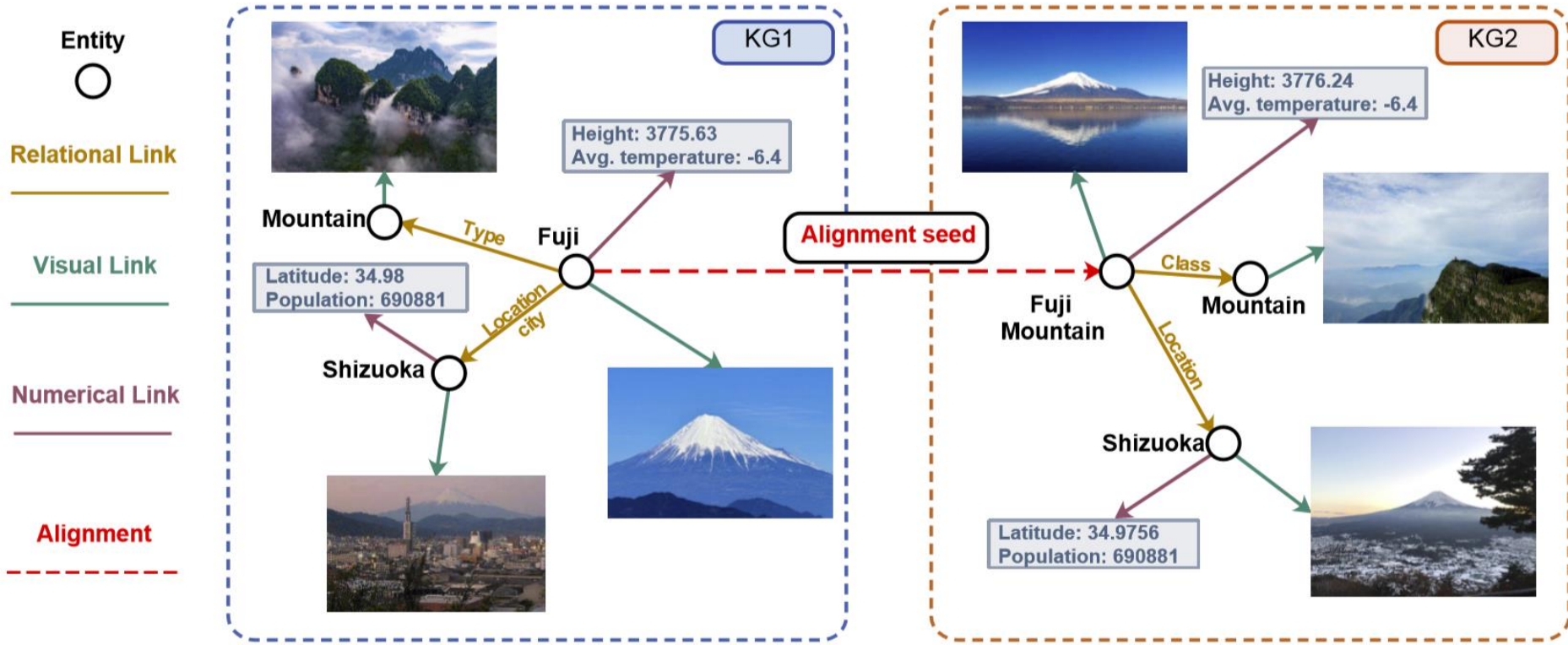
- 视觉语义信息丰富，场景多源化
- 关系丰富

缺点

- 图谱庞大
- 符号复杂

• 多模态知识图谱的作用

- 模态知识互补：不同模态协同服务于实体的理解



• 多模态知识图谱的作用

• 模态信息搜索：以不同模态作为信息入口

Q 全部 图片 新闻 地图 视频 更多 工具

找到约 262,000 条结果 (用时 0.28 秒)

新浪体育
欧文和萧华今日完成会面富有成效且获得谅解
据此前的报道，欧文因为反犹事件被篮网禁赛，必须要完成六大事项才能回归。（修楚）. 声明：新浪网独家稿件，未经授权禁止转载！ 关键词...
6 小时前

搜狐
肖华和欧文完成会面双方进行富有成效的谈话且收到谅解_篮网_...
北京时间11月9日消息，篮网老板肖华和篮网队球星欧文完成会面，双方进行富有成效的谈话，欧文收到谅解...
6 小时前

手机网易网
欧文与萧华完成会面为他与篮网继续合作铺平道路_手机网易网
10月底，欧文在社交平台分享了一部纪录片，改编自一本充斥着反犹比喻的书，篮网老板蔡崇信当时发推表示：“我对欧文似乎支持了一部基于一本充满反犹太...
6 小时前

国际新闻
“反犹”争议发酵，欧文归队要过6关
【环球时报特约记者徐海睿】欧文的大麻烦来了！“竞技场”网站记者查拉尼亚6日称，针对欧文此前的反犹行为，NBA篮网队已提出6项要求，欧文只有全部完成...
2 天前

新闻

简介

凯里·安德鲁·欧文为美国职业篮球运动员，场上主打控球后卫，现效力于NBA布鲁克林篮网。在2011年NBA选秀中以第一顺位被克利夫兰骑士选中，2016年随骑士队夺得NBA总冠军。在2016年夏季奥林匹克运动会代表美国参与男子篮球比赛夺得金牌。

出生信息

身高: 1.88 米
臂展: 1.93 米
子女: 阿兹乌里·伊丽莎白·欧文
父母: 德雷德里克·欧文, 伊丽莎白·厄文
兄弟姐妹: 伦敦·欧文, 艾莎·欧文

结构化数据

声明对此知识面板的所有权 提供反馈

凯里·欧文
篮球运动员

概览 统计信息 新闻 电影



www.youtube.com > watch
欧文训练师单挑2米08中国小伙，在身高优势面前 - YouTube
Nike以及亞馬遜將全面封殺歐文！道歉僅僅只是一個開始！想住職業生涯已晚！【Kyrie Irving】#nba#人物#篮网#詹姆斯#...
YouTube · 篮球微传记 · 2022年7月26日



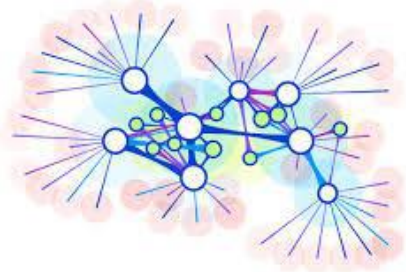
视频

www.youtube.com > watch
NBA首位巨星力挺欧文！杰伦布朗撑他 - YouTube
NBA首位巨星力挺欧文！杰伦布朗撑他：不认为欧文是反犹太主义者... Jaylen brown和kyrie Irving的宗教信仰相同。
YouTube · 篮球微传记 · 1 天前

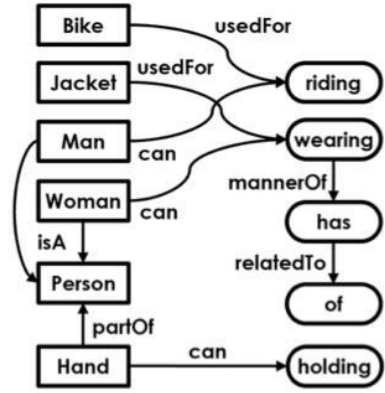
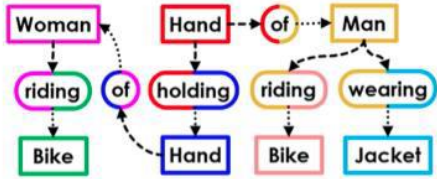
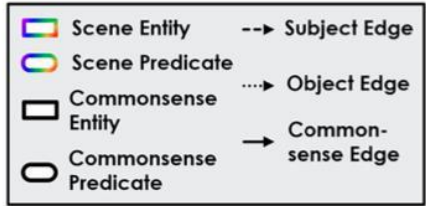
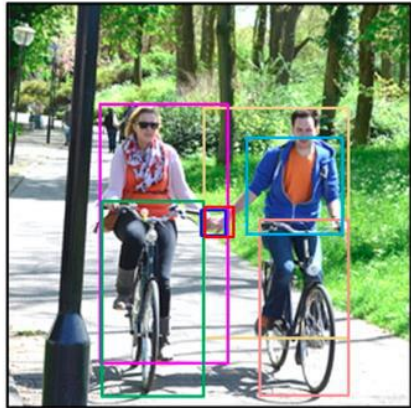


- 多模态知识图谱的作用

- 模态语义增强：利用知识图谱增强多模态任务



知识注入



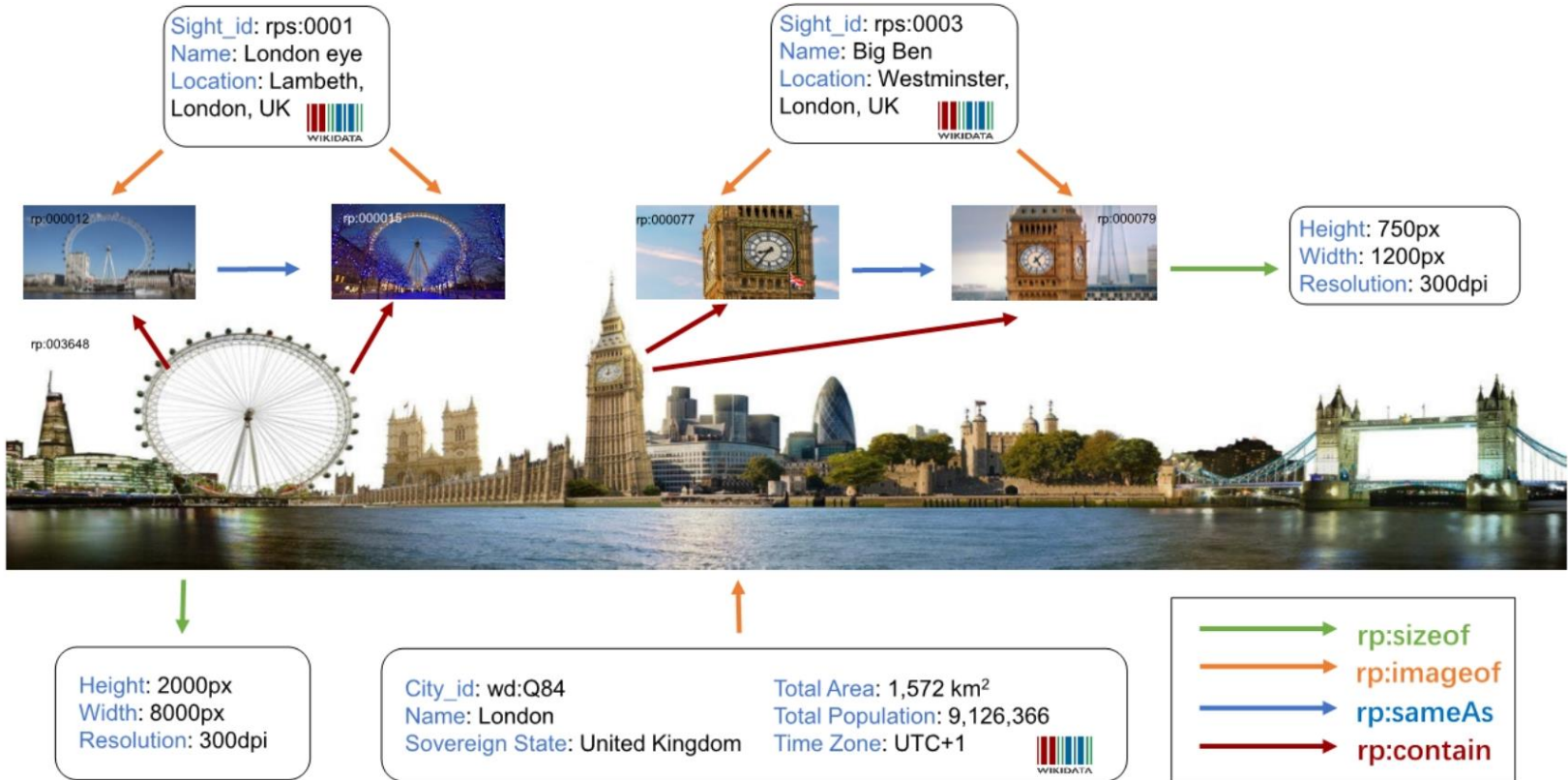
GB-NET

- 视觉语义理解
- 视觉场景图构建
- 视觉可解释性
-

• 多模态知识图谱例子

• 开放多模态知识图谱的典型代表：Richpedia

- 由东南大学认知智能研究所漆桂林老师团队提出，摆脱传统知识图谱中实体局限于文本的束缚



本章小结

实体识别

- 信息抽取概述：定义、五种基本任务
- 知识图谱概述
 - 发展历史、雏形思想
 - 基本要素：点（实体）、边（关系）、三元组
- 命名实体识别
 - 基于词典或规则的基础方法
 - 基于统计的方法：分类、序列标注