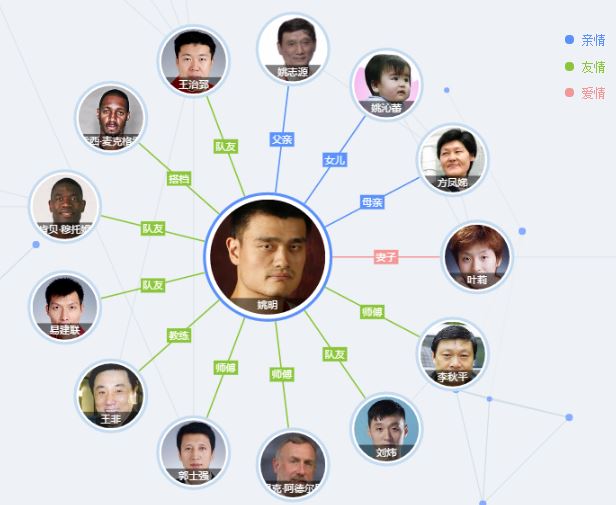
1.什么是知识图谱？

随着互联网中用户生成内容(User Generated Content, UGC)和开放链接数据(Linked Open Data, LOD)等大量RDF(Resource Description Framework)数据被发布。互联网又逐步从仅包含网页与网页之间超链接的文档万维网(Web of Document)转变为包含大量描述各种实体和实体之间丰富关系的数据万维网(Web of Data)。

在此背景下，知识图谱（Knowledge Graph）正式被Google于2012年5月提出，其目标在于改善搜索结果，描述真实世界中存在的各种实体和概念，以及这些实体、概念之间的关联关系。紧随其后，国内外的其它互联网搜索引擎公司也纷纷构建了自己的知识图谱，如微软的Probase、搜狗的知立方、百度的知心。知识图谱在语义搜索、智能问答、数据挖掘、数字图书馆、推荐系统等领域有着广泛的应用。

下图是搜狗知立方“姚明”的关系图：



谷歌S. Amit的论文《Introducing the Knowledge Graph: Things, Not Strings》

阿米特·辛格尔博士通过“The world is not made of strings, but is made of things”这句话来介绍他们的知识图谱的，此处的“thing”是和传统的互联网上的网页相比较：知识图谱的目标在于描述真实世界中存在的各种实体和概念，以及这些实体、概念之间的关联关系。

其实，知识图谱就是下一代搜索引擎或知识计算引擎。

**2.生活中存在知识图谱的应用吗？**

知识图谱的出现进一步敲开了语义搜索的大门，搜索引擎提供的已经不是通向答案的链接，还有答案本身。下图展示Google搜索结果的快照，当用户搜索“刘德华的年龄”时，其结果包括：

(1).列出了相关的网页文档检索结果；

(2).在网页文档的上方给出了搜索的直接精确答案“54岁”；

(3).并且列出了相关的人物“梁朝伟”、“周润发”以及他们各自的年龄；

(4).同时在右侧以知识卡片（Knowledge Card）的形式列出了“刘德华”的相关信息，包括：出生年月、出生地点、身高、相关的电影、专辑等。

知识卡片为用户所输入的查询条件中所包含的实体或搜索返回的答案提供详细的结构化信息，是特定于查询（Query Specific）的知识图谱。



这些检索结果看似简单，但这些场景背后蕴含着极其丰富的信息：

首先，搜索引擎需要知道用户输入中的“刘德华”代表的是一个人；

其次，需要同时明白“年龄”一词所代表什么含义；

最后，还需要在后台有丰富的知识图谱数据的支撑，才能回答用户问题。

同时，知识图谱还在其他方面为搜索引擎的智能化提供了可能，辛格尔博士指出：搜索引擎需要在答案、对话和预测三个主要功能上进行改进。另外，知识图谱在智能问题、知识工程、数据挖掘和数字图书馆等领域也具有广泛的意义。

那么，传统的搜索引擎发展到知识图谱之间又有哪些变化呢？下图是雅虎的搜索引擎，它是基于传统的关键字查找并返回与关键字相关的文档序列的方法，其中包括“姚明”和“身高”关键字，越关联的分数越高，排序越靠前。



主要的变化包括以下几点：

(1).信息抽取目标发生了变化，传统的文本指定抽取(ACE)=>海量数据的发现(KBP);

(2).从文本分析为核心转变成了知识发现为核心；

(3).让计算机真正理解用户的查询需求，给出准确答案而不是给出相关的链接序列。

**3.百度知心和搜狗知立方**  
        百度知心访问链接：<http://tupu.baidu.com/xiaoyuan/>  
        其中百度知识图谱“章子怡”人物关系图谱如下所示。



知识挖掘包括：Named entity mining 命名实体挖掘、AVP mining 属性-值对挖掘、Hyponymy learning 上下位学习、Related entity mining 相关实体挖掘。



搜狗知立方整体框架图如下所示，其中下部分的实体对齐、属性对齐是我现在研究的部分。主要包括以下部分：

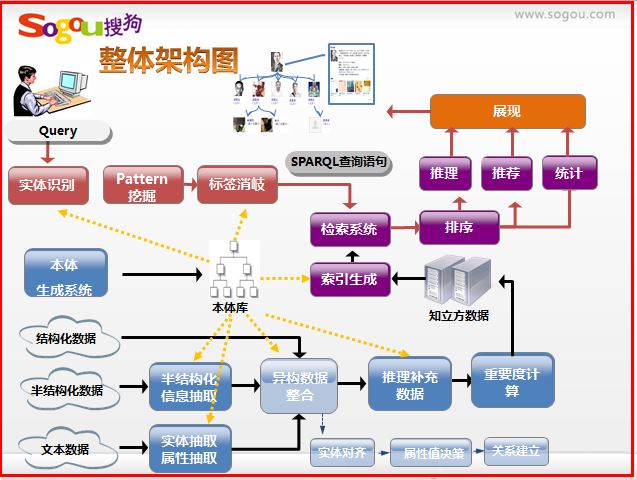
本体构建（各类型实体挖掘、属性名称挖掘、编辑系统）

实例构建（纯文本属性、实体抽取、半结构化数据抽取）

异构数据整合（实体对齐、属性值决策、关系建立）

实体重要度计算

推理完善数据



知立方信息展现及应用：

(1).提供知识库信息的展示载体，将知识库的信息转化为用户可以理解的内容；

(2).提供更加丰富的富文本信息（不局限于文字，增添图片、动画、表格等）；

(3).提供更友好的用户交互体验：增加更多的用户交互元素，如图片浏览、点击试听。并能够引导用户在更短的时间获取更多的信息。

如下图所示：“刘德华”分别点击上方基本信息，点击歌曲，点击属性标签，点击具体的电影



再如重名、系列实体展现如下： “李娜 ”点击其他的同名人物、“十大元帅”点击某个具体的人物、“速度与激情”点击更多，展示更多的系列实体，同时包括问答系统的应用。