

什么是知识图谱

如果一个人掌握的知识多一些，总会表现的更聪明，所以一个高中生穿越回唐朝也能一统天下。同样，知识对于人工智能也很重要，像 IBM 的 Watson 就是因为知识丰富，才在危险边缘 (Jeopardy) 游戏中击败了人类，展现出强大的智能。

那么，什么是知识呢？似乎是很简单的问题，但细究起来并不容易回答。牛顿第二定律肯定是知识，“如果下雨，路面会湿”这种因果关系也是知识，“天是蓝的”这种废话其实也是知识 -- 原因很简单，如果你不说，计算机就不知道。

所以说，我们身边的很多知识对计算机都是很重要的，关键是用什么方式把这些知识表示出来，让计算机能读懂。人们提出了很多知识表示方法，典型的包括语义网络、产生式规则、框架系统等。2012 年 5 月 16 日谷歌在语义网络 (Semantic Web) 的基础上提出知识图谱 (Knowledge Graph) 的概念，受到广泛关注[1]。

什么是知识图谱呢？简单地说，知识图谱是用图来对知识进行表示的方法。如图 1 所示，将“奥巴马”，“美国”，“迈克尔”等名字 (称为实体) 作为图中的节点，再将实体之间的关系表示成节点之间的有向连接，就形成了一个实体-关系图。这个图可以表示诸如“奥巴马是美国总统”，“奥巴马和迈克尔是夫妻”等知识。再复杂一些，可以为图中的实体标记上属性，来表示诸如“奥巴马是男性”，“奥巴马的生日是 1961 年 8 月 4 号”等知识。

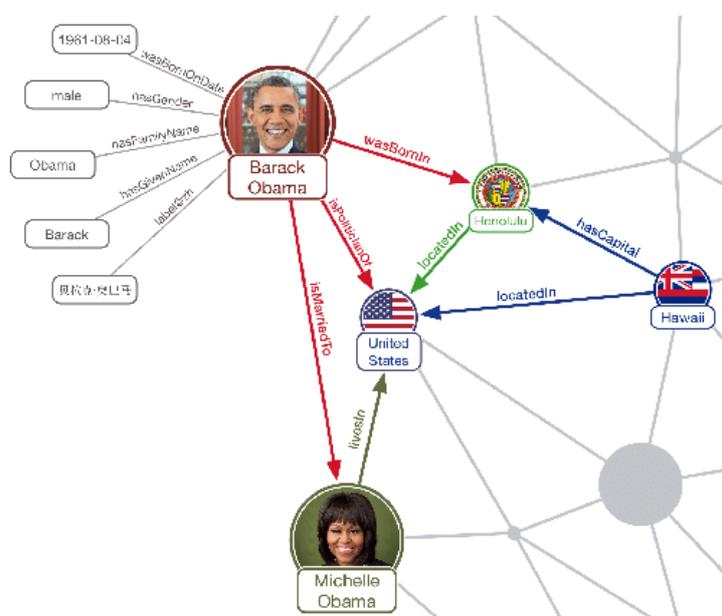


图 1: 知识图谱的例子

形式上，图谱中的知识可用每条边所代表的三元组来表示，如“奥巴马→总统→美国”。然而，因为这些孤立的三元组连接成了网络，使得我们可以推导出更多衍生知识。例如，如果我们知道了美国总统是奥巴马，而美国首都是华盛顿，那么就可以很容易推导出“奥巴马的工作地点在华盛顿”。因此，知识图谱所代表知识远比直观看到的实体间的连接要丰富。

构造一个知识图谱并不困难。如果已经有确定的实体和关系集合，只要把这些关系整理成三元组的形式就可以了。如果仅有一些文本资料，则需要对这些资料进行处理，从中抽取了实体和关系。这一方法通常用在开放域任务中，例如要造一个关于足球比赛的知识图谱，一般需要从新闻报导中把比赛这个事件抓取出来，并提取相关信息。

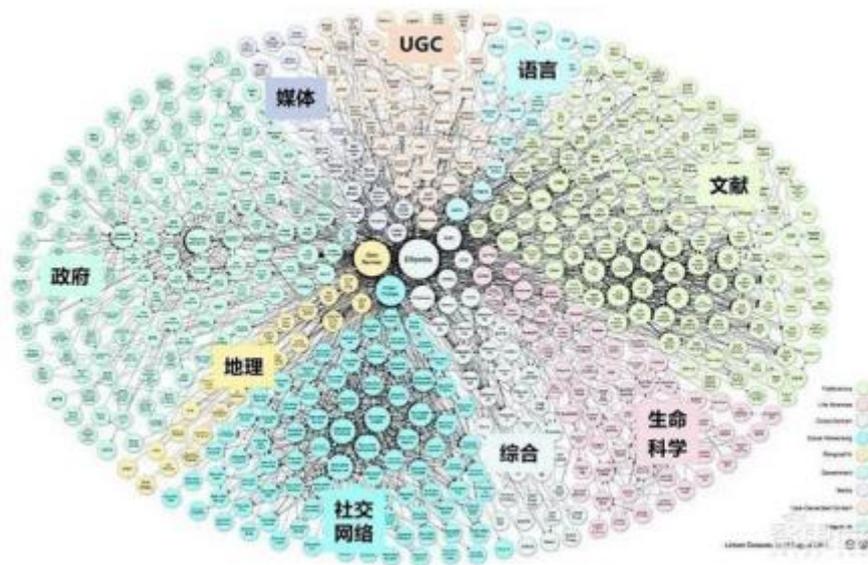


图 2：特定领域知识图谱 [2]

有了知识图谱，我们可以做很多事情。例如在搜索引擎中输入“奥巴马的工作地点”，如果没有知识图谱，引擎只会通过“奥巴马”“工作地点”等关键词去查找网页，如果没有这样的网页，搜索引擎是不能返回恰当结果的。如果有了知识图谱，搜索引擎就可以理解用户的输入目的（而不仅是字串），从而返回“华盛顿”这一正确结果。同样的技术在问答系统（如 Watson）和推荐系统中也有广泛应用。

[1]刘知远，韩旭，孙茂松, 知识图谱与深度学习, 清华大学出版社，2020 年 5 月第 1 版，245 页，ISBN: 9787302538523

[2]智东西，把知识变成图谱一共需要花几步？

<https://zhuanlan.zhihu.com/p/56903119>