

# 第一次作业：对 knowledge graph 的初步了解与概念消化

58119125 蒋卓洋

## 1. What is a knowledge graph?

在课后，我计划分为四步来检索综述“什么是知识图谱”这一问题，阅读相关科普文章和理论文献，得出一个大致的答案。

### (1) 尝试寻找定义

首先，我先理解了“知识”的基本定义。如果说信息是指外部的客观事实，那么知识就是对外部客观规律的归纳和总结。而知识图谱就是由一条条知识组成的，或者说知识图谱将信息构建一条条知识，描述了他们的自身属性和相互关系，再在总体上形成一个巨大的知识框架。

通俗地讲，知识图谱就是把所有不同种类的信息（Heterogeneous Information）连接在一起而得到的一个关系网络，本质上是一种揭示实体之间关系的语义网络，可以对现实世界的事物及其相互关系进行形式化地描述，其结点代表实体（entity）或者概念（concept），边代表实体/概念之间的各种语义关系。

因此，可以从实际应用的角度出发换个角度定义知识图谱。知识图谱是一种有向多关系图（Directed Multi-relational Graph），多关系图一般包含多种类型的节点和多种类型的边，我们用“实体（Entity）”来表达图里的节点、用“关系（Relation）”来表达图里的“边”，具体来说，实体指的是现实世界中的事物，关系则用来表达不同实体之间的某种联系。

综上，我们已经可以初步概括知识图谱的定义，之后，我们还可以进一步用数学语言来描述它。我们用  $G=\langle V, E \rangle$  来表示作为有向图的知识图谱，其中  $V$  代表有向图中的节点的集合，从而表示知识图谱中的实体（Entity）集合， $E$  代表有向图中的边的集合，从而表示知识图谱中的关系（Relation）和属性（Property）集合。

### (2) 理解熟悉概念

#### ① 应用角度理解：

我们可以先从不同的应用视角去审视知识图谱的概念：

- 在 Web 视角下，知识图谱如同简单文本之间的超链接一样，通过建立数据之间的语义链接，支持语义搜索。
- 在自然语言处理视角下，知识图谱就是从文本中抽取语义和结构化的数据。
- 在知识表示视角下，知识图谱是采用计算机符号表示和处理知识的方法。
- 在人工智能视角下，知识图谱是利用知识库来辅助理解人类语言的工具。
- 在数据库视角下，知识图谱是利用图的方式去存储知识的方法。

而从应用动机上看，知识图谱的主要目标是用来描述真实世界中存在的各种实体和概念，以及他们之间的强关系，我们用关系去描述两个实体之间的关联，例如姚明和火箭队之间的关系，他们的属性，我们就用“属性--值对”来刻画它的内在特性，比如说我们的人物，他有年龄、身高、体重属性。知识图谱可以通过人为构建与定义，去描述各种概念之间的弱关系，例如：“忘了订单号”和“找回订单号”之间的关系。

#### ② 支撑理论探索：

首先对一些专有名词的定义进行进一步探索。

- 实体：指的是具有可区别性且独立存在的某种事物。如某一个人、某一个城市、某一种植物等、某一种商品等等。世界万物有具体事物组成，此指实体。实体是知识图谱中的最基本元素，不同的实体间存在不同的关系。
- 语义类（概念）：具有同种特性的实体构成的集合，如国家、民族、书籍、电脑等。概念主要指集合、类别、对象类型、事物的种类，例如人物、地理等。
- 内容：通常作为实体和语义类的名字、描述、解释等，可以由文本、图像、影视

频等来表达。

- d. 属性(值): 从一个实体指向它的属性值。不同的属性类型对应于不同类型属性的边。属性值主要指对象指定属性的值。如图 1 所示的“面积”、“人口”、“首都”是几种不同的属性。属性值主要指对象指定属性的值,例如 960 万平方公里等。
- e. 关系: 形式化为一个函数,它把  $kk$  个点映射到一个布尔值。在知识图谱上,关系则是一个把  $kk$  个图节点(实体、语义类、属性值)映射到布尔值的函数。

有了上述定义,我们可以基于三元组表示知识图谱,三元组是知识图谱的一种通用表示方式。一个三元组就是一个关系。三元组的基本形式主要包括(实体 1-关系-实体 2)和(实体-属性-属性值)等。每个实体(概念的外延)可用一个全局唯一确定的 ID 来标识,每个属性-属性值对(attribute-value pair, AVP)可用来刻画实体的内在特性,而关系可用来连接两个实体,刻画它们之间的关联。

### (3) 了解工程过程

知识图谱有自顶向下和自底向上两种构建方式。自顶向下构建是指借助百科类数据源,提取本体和模式信息,并加入知识库中。自底向上构建是指借助一定的技术手段,从公开的数据中提取资源,选择其中置信度较高的信息,经人工审核后,加入知识库中。

#### ① 知识表示与建模:

将实体的语义信息表示为稠密低维实值向量,进而在低维空间中高效计算实体、关系及其之间的复杂语义关联,对知识库的构建、推理、融合以及应用均具有重要的意义。

#### ② 知识获取:

就是从各种类型的数据源中提取实体、属性以及实体间的相互关系,在此基础上形成本体的知识表述。知识图谱的构建过程中存在大量的非结构化或者是半结构化数据,这些数据在知识图谱的构建过程中需要通过自然语言处理的方法进行信息抽取。从这些数据中,我们可以提取出实体、关系和属性。

#### ③ 知识融合:

主要工作是把结构化的数据以及信息抽取提炼到的实体信息,甚至第三方知识库进行实体对齐和实体消歧。这一阶段的输出应该是从各个数据源融合的各种本体信息。

#### ④ 知识加工:

其中有重要的工作比如知识图谱的补全。常用的知识图谱的补全方法包括:基于本体推理的补全方法、相关的推理机制实现以及基于图结构和关系路径特征的补全方法。

#### ⑤ 知识储存

#### ⑥ 知识应用(知识查询与推理)

### (4) 溯源相关理论

#### ① 知识表示(Knowledge Representation):

知识的表示就是对知识的一种描述,或者说是知识的一组约定,一种计算机可以接受的用于描述知识的数据结构。某种意义上讲,表示可视为数据结构及其处理机制的综合:表示= 数据结构+处理机制。常见的有产生式规则、语义网、框架法等。

#### ② 专家系统(Expert System):

专家系统被定义为一种交互式可靠的基于计算机的决策系统,它使用事实和启发式方法来解决复杂的决策问题。专家系统通常由人机交互界面、知识库、推理机、解释器、综合数据库、知识获取等 6 个部分构成。

为了使计算机能运用专家的领域知识,必须要采用一定的方式表示知识。目前常用的知识表示方式有产生式规则、语义网络、框架、状态空间、逻辑模式、脚本、过程、面向对象等。

### ③ 语义网络 (Semantic Network)

语义网络是一种用图来表示知识的结构化方式。在一个语义网络中,信息被表达为一组结点,结点通过一组带标记的有向直线彼此相连,用于表示结点间的关系。

## 2. What is its role in machine intelligence?

首先,知识图谱与知识表示与推理 (Knowledge Representation and Reasoning) 有关,知识图谱构建也需要用到知识表示技术与理论,而知识表示与推理是人工智能领域里的传统分支,偏重于符号逻辑工作,目的是将人类的知识用符号逻辑表示,然后能够使用符号逻辑推理机从这些知识里推理出隐含的知识。这也是我们人类智能里非常重要的一部分。

其次,知识图谱需要用到人工智能领域中许多基础方法,机器学习、深度学习和自然语言处理等人工智能传统领域,都有知识图谱所需的理论基础与工具。例如,做各种领域的知识图谱时,语料缺失是高效高质做出知识图谱的最大障碍,结合深度学习和迁移学习的方法或许就能对这个难题的解决有所帮助。

再次,知识图谱有丰富的语义关系,概念、属性、关系等语义关系可以很好的应用到自然语言处理相关任务上,例如分词,短语理解,文本理解等任务上。通过知识图谱可以让机器能更好的去理解自然语言,进一步的更好的理解用户的意图,文本的含义。也就是说,知识图谱也能反过来帮助人工智能其他领域的方法构建与总体发展。

综上,正是因为可以基于这些在机器智能基础领域的既有方法,并辅以人工智能发展过程中产生的分支技术与理论,我们才可以更好地发展知识图谱;也正是因为知识图谱可以对其他智能领域的发展有新的帮助,知识图谱才会成为今天机器智能发展的一个新的突破点。

而换一个角度来说,如果把机器智能与人类智能做一个对照,那么人类视觉对对应计算机视觉、人类阅读与听力对应语言识别和自然语言处理、人类肢体运动对应机器人智能工程。以上这些从总体上看都是感知和学习,如果再进一步对人类的特性进行观察,那么“思考与创造”就是一个重要的人类特性,而“知识”对人类来说就是这种特性的重要前置条件和过程变量,所以知识图谱可以说是机器由感知到认知、由智能到博学的桥梁。

## 3. 3 techniques of Knowledge Graph.

### (1) 用基于大数据的企业图谱建筑构建的问答系统

基于知识图谱的企业知识问答系统,应用知识图谱及自然语言处理的方法,实现了智能的语义问答、检索和答案的可视化显示等功能。这将极大地改善人机交互方式,使得在查询数据库时无需数据库相关的专业技能及培训,显著降低非专业技术人员查询数据库和获取信息的难度。可视化的查询结果可以降低人们在从查询结果中获取信息、分析信息的难度,并且图的方式也极大地便利了人们通过现有信息进行推理、拓展出新的信息。

### (2) 基于知识图谱技术的招标代理

招标代理机构应用知识图谱技术提升采购文件编制质效,强化知识资产在招标采购领域的深度应用的方法,对不断提升招标代理机构业务专业化水平,推动采购文件编制活动向更智能,更高效的方向发展。通过研究基于知识图谱技术的知识搜索,知识推荐功能在采购文件编制全过程的有效应用,对提升整个招投标活动的效率效益,以及发挥招标采购相关知识资产的应用价值具有实践意义。

### (3) 基于知识图谱技术的网络不良信息检测和过滤

随着信息产业的发展,网络在给大家提供信息服务的同时也带来了一定的安全隐患。而现在的网络信息检测和过滤形式单一,无法满足信息产业发展的需要。知识图谱在信息检索中的高速发展为不良信息的检索和过滤带来了新的契机。基于现阶段知识图谱构建的发展现状结合网络不良信息的数据源,可以总结出面向网络不良信息知识图谱构建方法和步骤。

## 参考文献

- [1] 漆桂林、高桓、吴天星, 《知识图谱研究进展》[J]. 情报工程, 2017, 3(1)
- [2] 徐增林等, 《知识图谱技术综述》[J]. 电子科技大学学报, 2016, 7(4)
- [3] 机器学习研究会, 《最全知识图谱综述: 概念以及构建技术》[DB/OL] . 2021-1-7
- [4] 王昊奋、漆桂林、陈华钧, 《知识图谱: 方法、实践与应用》, 电子工业出版社, 2019-8
- [5] 袁安云. 基于大数据的企业图谱的研究与应用[D]. 华南理工大学, 2017.
- [6] 薛朋强. 面向网络不良信息知识图谱构建方法研究[D]. 新疆大学, 2017.
- [7] 程志强. 基于新浪微博主题的用户影响力研究[D]. 东北大学, 2013.
- [8] 知乎文章: 知识图谱之语义网络篇, 作者: 漆桂林,  
<https://zhuanlan.zhihu.com/p/28276520>
- [9] 知乎文章: 为什么需要知识图谱? 什么是知识图谱? ——KG 的前世今生, 作者: SimmerChan, <https://zhuanlan.zhihu.com/p/31726910>
- [10] 知乎文章: 把知识变成图谱一共需要花几步? 作者: 智东西,  
<https://zhuanlan.zhihu.com/p/56903119>
- [11] 知乎文章: 知识图谱-从入门到跑路(1), 作者: cavities  
<https://zhuanlan.zhihu.com/p/62824358>
- [12] 知乎答案: 知识图谱的构建流程?, 作者: Hooke,  
<https://www.zhihu.com/question/299907037/answer/519394870>
- [13] 知乎答案: 知识图谱的构建流程?, 作者: 陈运文,  
<https://www.zhihu.com/question/299907037/answer/537482952>