

知识图谱相关学习

知识图谱

知识图谱是结构化的语义知识库，用于描述物理世界中的概念及其相互关系，由结构化、半结构化和非结构化数据（如关系数据库、XML、JSON、百科、图片、音频、视频、文本）经过加工、处理、整合，形成实体及其关系的三元组构成。

常用的存储方法包括使用 **RDF（Resource Description Framework）** 规范或图数据库（如 **Neo4j**），从而实现知识的快速响应和推理。

知识图谱的构建过程主要包含三个阶段：**知识抽取、知识融合和知识加工**。

- 知识抽取从各种类型的数据源中提取出实体、属性以及实体间的相互关系，在此基础上形成本体化的知识表达；
- 知识融合对获得的新知识进行整合，以消除矛盾和歧义，比如某些实体可能有多种表达，某个特定称谓也许对应于多个不同的实体等；
- 知识加工对经过融合的新知识进行质量评估之后（部分需要人工参与甄别），将合格的部分加入到知识库中，以确保知识库的质量。

知识图谱问答

我的理解：现在的项目大概率也会用到这部分

知识图谱问答(Knowledge Graph Question Answering, KGQA) 是指在问答场景中使用知识图谱作为主要数据源来获取答案的技术，其基于知识图谱进行推理，从而为给定的问题提供准确的答案。（为知识增强提供更准确的参照实体）

知识图谱问答根据答案获得的方式可分为**语义解析**和**信息检索**两种方法。

- 语义解析方法是**将非结构化的自然语言问题映射为一系列结构化逻辑形式**，例如语义图和高级查询语言(如 SPARQL, Cypher 等)。
- 信息检索方法则旨在**将自然语言问题和知识库中的实体和关系映射为同一低维空间中的特征向量**，将任务转化为问题向量与知识图谱中对应实体关系向量之间的相似度匹配任务以获得候选答案。



图 1 知识图谱问答方法

语义解析

语义解析方法旨在将 NLP 领域的语义解析应用于 知识图谱问题翻译，主要基于词汇表映射完成对知识图谱信息的利用，可通过系统生成的查询语句或逻辑形式解释系统的答案来源，可解释性较高。

传统语义解析共有直接映射和复述两种方法。（直接生成和检索生成）

直接生成：

语义解析技术的核心在于将输入问题翻译为能在图上执行的查询语句。

直接生成的方法旨在利用大语言模型本身的语义理解能力或者通过**微调、提示工程**等方法将输入问题直接转换为逻辑查询语句以进行后续的知识图谱问答。

chatKBQA,大模型进行问题到数据库查询语言的转换，首先利用微调的大预言模型生成逻辑模式，而后通过无监督检索方法检索和替换实体和关系。

UniOQA，包含翻译器和搜索器，翻译器核心为大语言模型，将自然语言转换为CQL

还有基于提示工程的方法

直接生成的方法目的在于提升知识检索的效率，然而，在没有相关内容参考的情况下直接将问题转换为对应的逻辑查询语句，过度依赖于大语言模型的能力，准确率不能得到很好的保障。

检索生成：