8.【x=1.1】针对实体关系查询的知识图谱划分算法设计

应用背景:

知识图谱的分析和挖掘，由于知识图谱体积巨大，例如Freebase 接近400GB，已经很难在单机系统上面实现。目前主要的方法是将其移植到分布式RDF存储管理系统中。知识图谱被,根据实体（entity)为单位，通过将实体id哈希后,随机均匀的分布在整个集群空间中,并且提供全局地址映射。

然而不同的数据划分方法对各种算法影响较大,某些情况下系统可能只需要针对两个实体之间关系查询(可达性查询）的算法。比如在警察系统中,经常在破案过程中需要对某些人物之间的特定关系进行检查。

1. 了解RDF 数据格式，SPARQL,知识图谱结构，以及图划分基本方法,针对知识图谱上的实体关系查询划分问题进行算法设计
2. 请使用 Hadoop/Spark系统实现该算法，并且使用LUBM生成合适大小的数据集，试该算法性能（例如,划分结果的通信开销,算法运行时间,等等)
3. 测试上述方法和该数据集在 Trinity 上的性能差距（也可以选择其他分布式 RDF 数据存储查询引整)

数据来源:http:// swat.cse.lehigh.edu/pro jects/lubm/