

基于 Neo4j 图数据库的孔子人物关系图谱的设计与实现

李宇

(长春财经学院, 吉林 长春 130122)

摘要: 人物关系数据中隐含着丰富的信息, 图数据库在数据挖掘和可视化方面具有优势。本文基于图数据模型, 采用 Neo4j 技术, 定义了 7 类实体、6 类关系及其属性, 存储节点、关系及属性数据, 构建了孔子人物关系图谱。通过 Neo4j 查询实现了人物关系图谱的 Web 可视化, 为进一步研究孔子及儒家文化提供有力支撑。

关键词: 孔子; Neo4j 图数据库; 数据库设计; 可视化

DOI: 10.12184/wspcyxcx2WSP2516-415502.20220607

引言

知识图谱 (Knowledge Graph) 的概念由谷歌公司在 2012 年 5 月正式提出, 旨在实现更智能的搜索引擎, 于 2013 年之后开始在学术界和业界普及^[1]。知识图谱本质上是一种叫做语义网络的知识库, 即具有有向图结构的知识库, 其中图的结点代表实体 (entity) 或者概念 (concept), 图的边代表实体或概念之间的各种语义关系^[1]。图数据库的数据存储结构和数据查询方式都以图论为基础, 适用于含有大量联系的人物关系数据的查询^[2]。

孔子是我国古代伟大的思想家、政治家、教育家, 儒家思想的创始人, 是儒家学派的代表人物之一, 在中国历史上乃至世界文化史上有着深远的影响和重要的地位。对以孔子为中心的历史人物关系进行挖掘, 对于弘扬儒家文化, 具有重要的意义和历史价值。目前, 对孔子人物关系的研究主要集中在孔子世家谱^[3-4], 对孔子与其他历史人物关系的研究的资料不多, 为了更好地以可视化方式展示孔子与家人、弟子及儒家学派等历史人物关系, 基于 Neo4j 图数据库构建了以孔子为中心的人物关系图谱, 并使用 Neo4j CQL^[5-7] 对人物关系进行查询。

一、Neo4j 图数据库

(一) Neo4j 图数据库基本概念

Neo4j 是目前最热门的图数据库系统, 遵循属性图模型来存储和管理数据, 包含节点和关系两种基本的数据类型, 节点和关系包含 key-value 形式的属性。

节点 (node) 是图数据库的基本元素, 每个节点代表一个实体, 在图数据库中用圆圈表示。

关系 (Relationship) 用以表示节点之间

的联系, 在图数据库中用箭头表示。关系具有方向性, 可以是单向的也可以是双向的。

属性 (Property) 用以表示节点或关系所具有的特征, 节点和关系都可以有多个属性。属性以 key-value 对的形式存在 [8], Key 是一个字符串, Value 是基本类型或基本类型数组。

(二) Neo4j 图数据库的优势

传统的关系型数据库不能很好地表现数据的联系, 相较于传统的关系型数据库, 图数据库强化了“关系”的概念, 以图结构来存储和管理数据, 存储的就是关联数据本身, 能够直接表现数据的关联特性。在可扩展性方面, 关系型数据库的模式约束使之很难扩展, 而 Neo4j 可以动态增删节点和关系, 且不需要重构数据库, 因此, 相比于关系型数据库, Neo4j 具有良好的可扩展性^[9]。Neo4j 图数据库用可视化的方式对数据进行管理, 可以更加直观地与数据进行交互。

Neo4j 是一个高性能的图数据库, 支持查询效率高、扩展性强的声明式图查询语言。本文选用 Neo4j 存储人物关系图谱, 将人物实例存储为图数据库中的节点, 将人物关系存储为连接节点的边 (edge)。

二、孔子人物关系图数据库设计

(一) 节点定义

孔子人物关系图谱结构比较单一, 适用于基于规则的实体识别方法。本文定义了 7 类实体及其属性, 构建了以孔子为中心人物的图数据库, 7 类节点分别是: “center” 节点、“家人” 节点、“老师” 节点、“弟子” 节点、“先秦三儒” 节点、“国学三圣” 节点及“孔门十哲” 节点。

Neo4j CQL 是 Neo4j 图形数据库的查询语

言。它是一种声明性模式匹配语言,用户不必编写图形结构的遍历代码,就可以对图形数据进行高效的查询,具有简单可读、易于操作的特点。Neo4j CQL 使用 CREATE 命令创建带有属性的节点。它创建一个具有一些属性(键值对)的节点来存储数据。以孔子节点的定义为例,用 CREATE 命令创建一个节点,节点包含“center”、“先秦三儒”、“国学三圣”及“世界十大文化名人”等标签,包含“名”、“字”、“性别”、“出生日期”、“死亡日期”、“学派”及“代表作”等属性。

```
CREATE
(
    p:`center`:`先秦三儒`:`国学三圣`:`
    世界十大文化名人`
    {
        name:`孔子`, Given_Name:`丘`,
        Courtesy_Name:`仲尼`, Gender:`男`,
        Date_of_Birth:`公元前 551 年 9 月 28 日`,
        Date_of_Death:`公元前 479 年 4 月 11 日`,
        Place_of_Birth:`鲁国陬邑(今山东曲阜)`,
        dynasty:`春秋末期`, Country:`中国`,
        School:`儒家`, Major_Works:`六经`,
        Identity:`思想家、教育家`
    }
)
```

(二) 关系定义

根据对孔子与相关人物关系的分类,本文定义了 6 种关系,分别是“家人”、“师傅”、“弟子”、“先秦三儒之一”、“国学三圣之一”和“孔门十哲之一”。以“孔子”与“家人”之间关系为例,关系属性描述如表 1 所示。

表 1 “孔子”与“家人”关系表

姓名	关系
叔梁纥	父亲
颜徵在	母亲
孟皮	哥哥
亓官氏	妻子
孔鲤	儿子
伯夏	祖父
孔伋	孙子
防叔	曾祖
孔白	曾孙

下面以“孔子”与“家人”之间的关系为例,用 Neo4j CQL 命令创建“孔子”与“家人”之间的关系:

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'颜徵在'}) merge(m)-[:母
```

亲]->(n)

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'叔梁纥'}) merge(m)-[:父亲]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'孟皮'}) merge(m)-[:哥哥]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'亓官氏'}) merge(m)-[:妻子]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'孔鲤'}) merge(m)-[:儿子]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'伯夏'}) merge(m)-[:祖父]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'孔伋'}) merge(m)-[:孙子]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'防叔'}) merge(m)-[:曾祖]->(n)
```

```
MATCH (m:center{name:'孔子'}),(n:`
    家人`{name:'孔白'}) merge(m)-[:曾孙]->(n)
```

三、孔子人物关系图数据库的实现

将 7 类实体、6 类关系及其属性录入完成后,在 Neo4j 图数据库中生成孔子人物关系图谱,如图 1 所示。

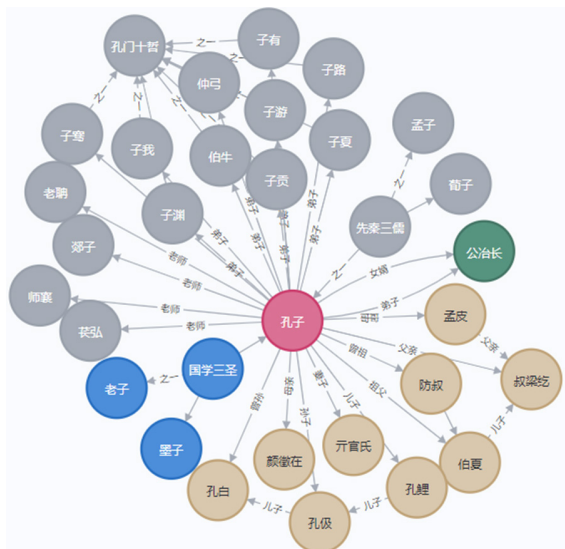


图 1 孔子人物关系图谱

(一) 人物关系查询

(1) 孔子与家人关系

“孔子”与“家人”关系查询命令: MATCH

(m{name:'孔子'})-[]-(n:`家人`) RETURN m, n

查询结果如图 2 所示。

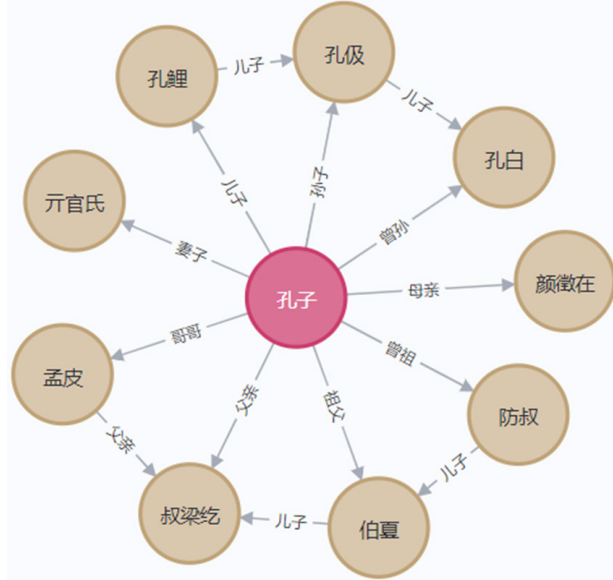


图 2 孔子与家人关系图谱

(2) 孔子与部分弟子关系

孔子与部分弟子 (“孔门十哲”及“孔门七十二贤”之一的公冶长) 关系的查询命令:

MATCH (m{name:'孔子'})-[]-(n:`弟子`)-[]-(p) RETURN m, n, p

查询结果如图 3 所示。

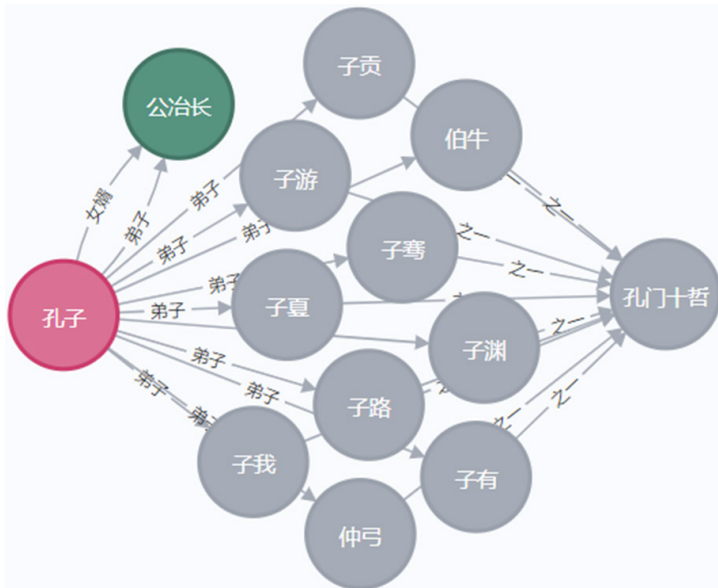


图 3 孔子与弟子关系图

四、结语

本文利用 Neo4j 图数据库查询性能高、易于操作的特点, 将数据存储在 Neo4j 图数据库中, 构建了孔子人物关系图数据库, 通过 Neo4j 查询功能实现了孔子与各类人物关查询。

孔子人物关系图谱化有利于直观展示孔子与各类人物关系, 帮助人们更好理解和梳理孔子人物关系, 便于挖掘孔子人物关系图谱中的隐藏关联关系, 有助于儒家文化的研究与发展。

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(01): 4-25.
- [2] 于娟, 黄恒琪, 席运江, 朱正祥. 基于图数据库的人物关系知识图谱推理方法研究 [J]. 情报科学, 2019, 37(10): 8-12. DOI:10.13833/j.issn.1007-7634.2019.10.002.
- [3] 张政平, 倪建成. 基于本体的孔子世家谱知识图谱 [J]. 通信技术, 2021, 54(06): 1370-1377.
- [4] 李平, 侯贺良. 人类最庞大的家谱:《孔子世家谱》[J]. 走向世界, 2008(23): 38-43.
- [5] Vukotic A, Watt N, Abedrabbo T, et al. Neo4j in Action [M]. US: Manning Publications Co, 2014.
- [6] DRAKOPOULOS G, BAROUTIADI A, MEGALOOIKONOMO V. Higher order graph centrality measures for Neo4j [C] // International Conference on Information, 2016: 6-8.
- [7] DRAKOPOULOS G, KANAVOS A, MYLONAS P, et al. Defining and evaluating Twitter influence metrics: a higher-order approach in Neo4j [J]. Social Network Analysis & Mining, 2017, 7(1): 52.
- [8] 王红, 张青青, 蔡伟伟, 姜洋. 基于 Neo4j 的领域本体存储方法研究 [J]. 计算机应用研究, 2017, 34(08): 2404-07.
- [9] 王余蓝. 图形数据库 NEO4J 与关系数据库的比较研究 [J]. 现代电子技术, 2012, 35(20): 77-79. DOI: 10.16652/j.issn.1004-373x.2012.20.045.

(作者简介: 李宇(1984-09), 女, 汉族, 辽宁锦州人, 长春财经学院, 工程师, 研究生学历, 硕士学位, 研究方向: 大数据、知识图谱。)