引文格式：陈年松. 基础地理实体关系的知识图谱构建方法研究[J]. 地理空间信息, 2024, 22(6): 20-24. doi:10. 3969/j.issn. 1672-4623.2024.06.005

基础地理实体关系的知识图谱构建方法研究

陈年松[1]
（1. 江苏省测绘工程院，江苏 南京 210013）

摘　要：针对新型基础测绘中构建地理实体关联关系的问题，提出一种基础地理实体关系的知识图谱构建方法，利用江苏省基础地理实体库作为实验数据，结合本体构建、知识抽取、知识图谱存储和语义网表示等关键技术，实现了基础地理实体知识图谱的构建与应用。实验结果表明：与传统的库表属性关联方法相比，该方法降低了地理实体关联关系构建的复杂度，串联地理实体多源异构数据，实现结构化、半结构化数据的高效存储与快速索引，满足了新型基础测绘“一库多能，按需组装”的需求，为实景三维中国的建设提供了可靠、高效的服务基础。

关键词：基础地理实体；知识图谱；RDF三元组模型；本体构建
中图分类号：P205　　**文献标志码：**B　　**文章编号：**1672-4623（2024）06-0020-05

Research on Knowledge Graph Construction Method of Basic Geo-entities Relationships

CHEN Niansong[1]
(1. Jiangsu Province Surveying and Mapping Engineering Institute, Nanjing 210013, China)

Abstract: Aiming at the problem of constructing the association relationship of geo-entities in the new basic surveying and mapping, we proposed a knowledge graph construction method of basic geo-entities relationship. Taking the geo-entity database of Jiangsu Province as experimental data, we used the technical methods of ontology construction, knowledge extraction, knowledge graph storage and semantic web representation to realize the construction and application of basic geo-entity knowledge graph. The experimental results show that compared with the traditional database table attribute association method, this method reduces the complexity of geo-entities association relationship construction, connects multi-source heterogeneous data of geo-entities in series, realizes efficient storage and rapid indexing of structured and semi-structured data, meets the needs of “one library with multiple functions, and assembles on demand”, and provides a service foundation for the real scene 3D China.

Key words: basic geo-entity, knowledge graph, RDF triplet model, ontology construction

地理实体能够实现地理信息与社会、经济、自然资源等专题信息的挂接，是各类信息聚合载体[1]。地理实体概念被提出以来，很多国家都已开展了其关键技术的研究和应用。2008年，美国地理信息数据委员会发布了“地理信息框架数据内容系列标准”，对地籍、行政管理单元、水系、交通均采用面向实体的数据模型FGDC[2]。国家基础地理信息中心于2011年推出了《地理信息公共服务平台地理实体与地名地址数据规范》，对院落、房屋、政区、道路、铁路、河流等重要对象构建基础地理实体[3]。时空大数据平台建设大纲也将基础地理实体数据纳入到基础时空数据系列中，并扩充相应的地理实体类API及服务[4]。

知识图谱（Knowledge Graph）是一种大型的语义网络，用于描述客观世界中概念、实体及其相互关系[5-6]。知识图谱以更加接近人的思维形式，构建在计算机世界中表示信息和知识的有效载体，成为人工智能应用的重要基础设施。知识图谱是大数据环境下信息服务向知识服务拓展的关键，在语义搜索、智能问答、预测决策等方面显示出越来越重要的应用价值[7]。针对地理实体的知识图谱构建的相关技术研究，主要有地理实体抽取[8]、拓扑、方位关系的抽取[9]，以及地理知识图谱存储[10]；在地理知识图谱应用方面，典型的有地理知识语义共享网络系统[11]和基于地理知识的地名词典[12]。

本文依托测绘部门现有的基础地理实体数据库，分析了基础地理实体数据的分类情况、基本属性、专题属性和实体关系，构建基础地理实体数据本体概念，设计并实现了基础地理实体关系的知识抽取流

收稿日期：2023-02-01。
作者简介：陈年松（1979—），硕士研究生，高级工程师，主要研究方向为数据分析和地理信息系统应用，E-mail：chenniansong@163.com。

程，并将抽取后的实体关系存储到图数据库中，实现了基础地理实体库知识图谱的构建和应用。

1　RDF三元组模型简介

RDF三元组模型是当前知识图谱常用的一种数据模型[13]，用RDF描述的本体是最基本的本体[14]。知识图谱中最基本的语义单元用(节点1, 关系, 节点2）的RDF三元组结构表示，其中节点可以代表概念、实体和属性。例如“江苏省的省会是南京”这段文字信息，可存储为“（江苏, 省会, 南京）”的语义单元。“江苏”和“南京”为2个实体，“省会”表示关系, 每个实体还可以有人口、面积、经度和纬度等属性信息附着，如图1所示。

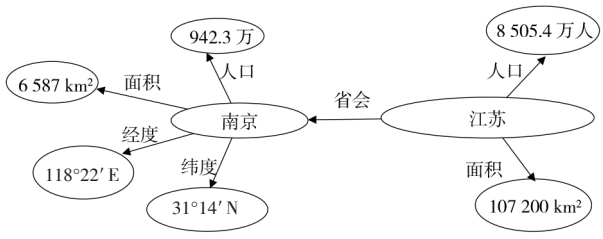


图1　RDF三元组模型示例

2　知识图谱构建的流程

本文采用江苏省新型基础测绘试点项目所生产的基础地理实体库作为实验数据，构建基础地理实体关系的知识图谱。

2.1　本体构建

本体（ontology）作为重要的知识组织系统，在知识管理、语义Web和人工智能等方面发挥着重要作用[9]。美国斯坦福大学Gruber教授给出了本体的定义：本体是概念化的规范说明[15]。基础地理实体数据库的分层结构，和它的本体定义之间并不能一一对应。数据库存储的空间数据，按照点、线、面、注记四类几何形态分层。例如河流中心线和河流面，本体的分类都属于河流，却分为2个图层。此外，基础地理实体间的拓扑关系在数据库中并没有存储，需要利用空间分析算法去构建。因此，需要进行基础地理实体的本体构建。本体构建主要分为分类构建、属性构建和关系构建三部分。

2.1.1　基础地理实体的分类构建

本文将基础地理实体分为自然实体、人工实体和地理单元三大类，每一类地理实体又分若干小类。由于本体分类并不与基础地理实体数据库的分层结构一致，所以在这一阶段需要手工标注两者间的映射匹配关系。

2.1.2　基础地理实体的属性构建

对基础地理实体的分类进行构建之后，需要对其属性进行构建。基础地理实体的属性又可分为基本属性和专题属性。基本属性是每一类地理实体图层所共有的属性，包括分类编码、实体编码、名称、其他名称、几何形态等；专题属性是各个基础地理实体图层所特有的属性。以河流为例，专题属性包含起点、终点、单双向、长度、河流等级、河流功能等属性。

在基础地理实体数据库，因为数据来源等原因，很多非必填字段存在空值。然而在本体的属性构建中，缺失一项属性意味着缺少一个三元组结构。因此这一阶段的主要工作是收集信息，尽可能多的进行基础地理实体的属性补全。

2.1.3　基础地理实体的关系构建

语义信息是地理实体的重要组成部分，在地理实体的管理应用中发挥着重要作用。本文将基础地理实体的语义关系归纳为拓扑关系、从属关系和时间关系三类关系。拓扑关系包括地理实体之间邻接、关联、包含和连通等空间关系；从属关系指地理实体之间的逻辑构成关系；时间关系包含了地理实体的产生时间、变更时间和消亡时间。这三类语义关系涉及的基础地理实体类别、语义描述以及三元组结构的表达式，如表1所示。

在基础地理实体的关系构建中，拓扑关系最重要，因此这一阶段的主要工作是利用空间分析算法去构建实体间的拓扑关系。

2.2　知识抽取

知识图谱的质量与规模受知识抽取方式的控制[16]，针对结构化数据、半结构化数据和非结构化数据，分别有不同的知识抽取方法[17]。基础地理实体知识抽取是从结构化的关系数据库（RDB）抽取地理实体间的空间和语义关系，形成RDF数据结构的过程。江苏省基础地理实体采用PostGIS建库。在PostGIS中，基础地理实体均由二维表表示，每个表都由行和列组成，每一行为一个记录值，每一列为一个字段的属性。而RDF由三元组组成的，具有层次的数据模型。根据关系数据库模型与RDF模型之间基本要素的对应关系，可制定基础地理实体基本要素映射关系表，以此实现基础地理实体的知识抽取。

2.3　知识图谱存储

在探究地理实体关系时，关系型数据库以“表格化结构”的方式对地理实体中的关系建模，对结构化数据

表1 基础地理实体中主要的语义关系构建

关系类型	相关地理实体	语义描述	三元组表达式
拓扑关系	河段、岛屿	岛屿位于河段内	（岛屿A，Inside，河段B）
	河段、桥梁	桥梁横跨于河段上	（桥梁A，Crosses，河段B）
	河段、码头	码头在河段边	（河段A，Intersects，码头B）
	湖泊、岛屿	岛屿位于湖泊内	（岛屿A，Inside，湖泊B）
	铁路、火车站	铁路经过火车站	（铁路A，Crosses，火车站B）
	公路、收费站	公路经过收费站	（公路A，Crosses，收费站B）
	公路、服务区	公路经过服务区	（公路A，Crosses，服务区B）
	地铁、地铁站	地铁经过地铁站	（地铁A，Crosses，地铁站B）
	行政区、居民地	居民地位于行政区内	（居民地A，Inside，行政区B）
从属关系	河流、河段	河段从属于河流	（河段A，subjection，河流B）
	道路、路段	路段从属于道路	（路段A，subjection，道路B）
	行政区	区县从属于市，街道乡镇从属于区县，社区行政村从属于街道乡镇	（区县A，subjection，市B），（街道乡镇A，subjection，区县B），（社区行政村A，subjection，街道乡镇B）
时间关系	所有类别	描述地理实体的产生时间、变更时间和消亡时间	（地理实体A，Start，时间A），（地理实体B，Update，时间B），（地理实体C，End，时间C）

注：三元组的主语和宾语均为相应的地理实体编码。

的处理起到重要的作用。但其严格的建模约束，使其难以适应动态变化的数据结构关系，在处理实体间的关联关系上，关系型数据库需要通过“外键”连接的方式对表格进行操作，这使得其查询效率低下。而图数据库采用RDF三元组模型和“点－边”有向图的形式，可以更好地表达关系结构，在进行实时关系查询时更加高效。

以水系知识图谱为例，分别采用关系型数据库和图数据库的方式进行表现。要使用关系数据库创建水系知识图谱的存储结构，必须为河流、河段、岛屿、桥梁、行政区单独创建表结构，各个表结构之间通过外键约束相互关联，对于多对多的复杂关系还必须创建中间表，如图2所示。

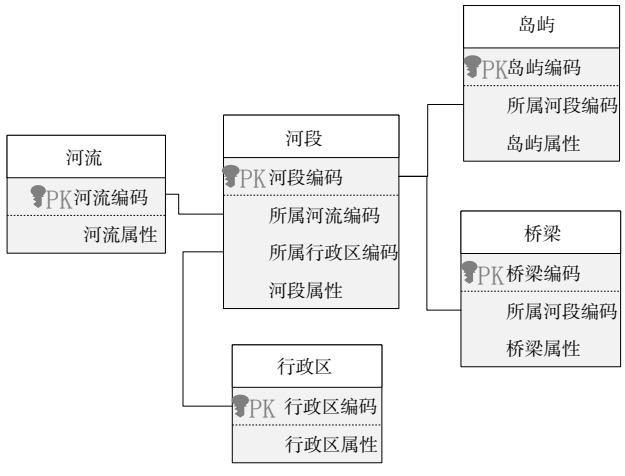


图2 关系型数据库的水系知识图谱

各个表结构中创建了大量的主键、外键列，并创建了中间表来维护复杂关系，这将消耗大量的系统资源。而采用图数据库创建水系知识图谱的存储结构则容易得多，只需要为河流、河段、岛屿、桥梁、行政区等地理实体类别创建节点，节点之间不需要主键、外键，也不需要中间表，只保留必要的属性即可，各个节点直接通过关系指向来表达节点之间的复杂关系，如图3所示。

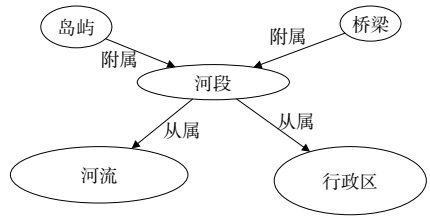


图3 图数据库的水系知识图谱

因此，使用图数据库可以更加简洁、明确地描述数据间的复杂关系，在存储基础地理实体知识图谱的性能上有很大的优势。

3 知识图谱构建的过程

基础地理实体的知识图谱构建过程如图4所示。

3.1 本体构建的实现

当前本体构建的工具有Protege、WebOneto、KA-ON、WebODE等[18]。其中，Protege具有支持中文、通过图形化界面构建、支持推理及软件的扩展等多种优势。本文选择Protege作为构建地理实体本体的工具，在Protege中将基础地理实体的类别、属性和关系做了定义。设计完成后，将本体文件以owl格式导出。

3.2 知识抽取的实现

本文使用D2RQ实现地理实体的知识抽取。D2RQ

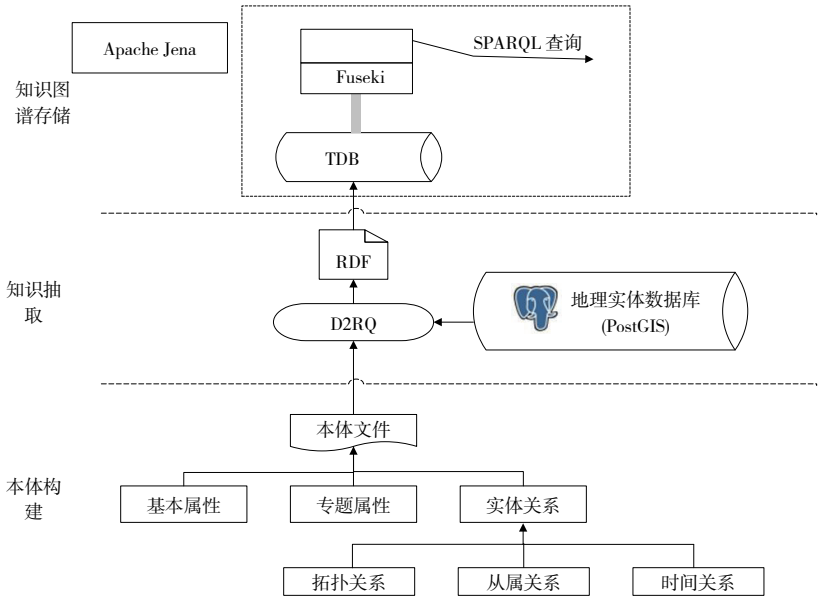


图4　基础地理实体知识图谱构建过程

可将关系型数据库中的内容快速发布为关联数据，并提供一个通用机制可将关系数据库的表和列进行映射[19]，实现知识抽取。

使用D2RQ将RDB转换为RDF，主要步骤如下：

1）使用D2RQ连接PostGIS数据库后，基于owl格式本体文件，生成ttl格式的映射文件。

2）利用dump-rdf命令，将ttl文件作为参数代入，抽取出RDF数据。

3.3　知识图谱存储的实现

本文使用图数据库Apache Jena来存储RDF数据。Apache Jena提供TDB、Fuseki等组件，其中TDB是Jena用于存储RDF类型数据的组件，Fuseki是Jena提供的SPARQL服务器，支持SPARQL语言进行检索，可在单机和服务器端高效运行[20]。实现知识图谱存储的主要步骤如下：

1）RDF转为TDB。Apache Jena需要TDB类型的数据，所以需要将已得到的RDF类型数据转换成TDB类型数据，转换方法可通过Apache Jena提供的工具进行实现。

2）SPARQL知识检索。RDF数据转换成TDB类型数据之后，需要通过Fuseki查看和检索数据。完成Fuseki配置，启动成功之后，即可查看所有数据集、增加数据集，和进行SPARQL查询。

4　实例应用

基于WebGIS的相关技术和基础地理实体知识图谱构建方法，本文开发了“基础地理实体数据共享应用平台”，用于基础地理实体数据的管理和应用。在平台中，当用户在前端查询“长江”的时候，系统查询图数据库，得到“长江”地理实体以及其周围相关的地理实体，返回结果到网页前端，以树状图的形式展现。

用户单击树状图中“长江支线八卦洲段”的叶子节点，系统可以查询到与该地理实体相关的知识图谱，并以“点-边”有向图的形式在网页前段展示八卦洲段相关的河流、桥梁、道路、行政区的情况。

5　结　语

本文探索了在新型基础测绘中构建地理实体关系的新思路，提出了一种基础地理实体知识图谱的构建方法，利用基础地理实体库作为试验数据，设计了基础地理实体本体构建、知识抽取、知识图谱存储和语义网表示等关键技术，实现了基础地理实体知识图谱的建设与应用。与传统的库表属性连接关联方法进行了比较分析，该方法降低了地理实体关联关系构建的复杂度，串联地理实体多源异构数据，实现结构化、半结构化数据的高效存储与快速索引，具备“一库多能，按需组装”的定制化能力，为实景三维中国的建设提供了可靠、高效的服务基础。

参考文献

[1]　闾国年,俞肇元,周良辰,等. 地理实体分类与编码体系的构建[J]. 现代测绘, 2019, 42(1):1-6

[2]　FGDC. Geographic Information Framework Data Content and Standard （FGDCSTD-014.X-2008） [EB/OL]. （2008-08-09） [2022-02-11].https://www. fgdc.gov/stan-

dards/projects/framework-data-standard,2008
[3] 卫东. 面向服务的山西省地理信息公共服务平台建设[J]. 地理信息世界,2012,10(1):76-79
[4] 郝利娟,刘冬枝. 智慧城市时空大数据云平台建设技术大纲研究[J]. 地理空间信息,2019,17(6):33-35
[5] Paulie, H. Marciano P. J. S. W. Knowledge Graph Refinement: A survey of Approaches and Evaluation Methods[J]. Semantic Web,2017,8(3):489-508
[6] 赵军,刘康,何世柱,等. 知识图谱[M]. 北京:高等教育出版社,2018
[7] Marrero M, Urbano J, S Nchez–cuadrado S, et al. Named Entity Recognition: Fallacies, Challenges and Opportunities[J]. Computer Standards and Interfaces, 2013, 35(5): 482-489
[8] 余丽,陆锋,刘希亮. 开放式地理实体关系抽取的Bootstrapping方法[J]. 测绘学报,2016,45(5):616-622
[9] 段红伟,孟令奎,黄长青,等. 面向SPARQL查询的地理语义空间索引构建方法[J]. 测绘学报,2014,43(2):193-199
[10] Gao Y, Gao S, Li R Q, et al. A Semantic Geographical Knowledge Wiki System Mashed up with Google Maps[J]. Science China Technological Sciences, 2010, 53(增刊): 52-60
[11] Liu Y, Li R Q, Chen K C, et al. Kidgs: A Geographical Knowledge-informed Digital Gazetteer Service[C]// Proceedings of the 17th International Confeerence on Geoinformatics. Fairfax,VA:IEEE,2009:1-6
[12] 李臻,刘彦林,马小雯,等. 基于标准知识图谱构建前瞻性研究[J]. 标准科学,2021(12):63-68
[13] 赵天忠,苗壮,张亚非,等. 基于图模型的RDF/RDFS本体抽取方法[J]. 解放军理工大学学报(自然科学版),2007(2):123-126
[14] 王向前,张宝隆,李慧宗. 本体研究综述[J]. 情报杂志,2016,35(6):163-170
[15] Gruber T. A Translation Approach to Portable Ontology Specifications[J]. Knowledge Acquisition,1993(5):199-220
[16] 张雪英,张春菊,吴明光,等. 顾及时空特征的地理知识图谱构建方法[J]. 中国科学(信息科学),2020,50(7):1 019-1 032
[17] 漆桂林,高桓,吴天星. 知识图谱研究进展[J]. 情报工程,2017,3(1):4-25
[18] 樊小辉,石晨光. 本体构建研究综述[J]. 舰船电子工程,2011,31(6):15-18
[19] 杨王黎,郑雪芸,袁满. 关系模型向RDF模型转换研究[J]. 微型电脑应用,2017,33(9):3-8
[20] 杨夏柏,杨明,杨德强,等. 基于Jena系统的知识融合三元组存储结构研究[J]. 价值工程,2018,37(8):126-129

(上接第9页)

[11] 李敏,杨艳,王海刚,等. 北京房山区城市化进程与土地储备管理[J].上海国土资源,2011,32(4):54-57
[12] 北京市统计局. 北京市房山区统计年鉴2020[M]. 北京:中国统计出版社,2020
[13] 北京市房山区统计局. 北京市房山区第七次全国人口普查公报,2021
[14] 北京市统计局. 2015年北京市1%人口抽样调查资料[M]. 北京:中国统计出版社,2015
[15] 河北省统计局. 2015年河北省1%人口抽样调查资料[M]. 北京:中国统计出版社,2015
[16] 天津市统计局. 2015年天津市1%人口抽样调查资料[M]. 北京:中国统计出版社,2015
[17] Bhaduri B, Bright E, Urban C M L. LandScan USA: A High-resolution Geospatial and Temporal Modeling Approach for Population Distribution and Dynamics[J]. Geojournal,2007,69(1-2):103-117
[18] 中国地震局.GB/T18306-2015[S]. 北京:中国标准出版社,2015
[19] 中国地震局. GB/T 18208.4-2011 [S]. 北京:中国标准出版社,2011
[20] 南师大智慧城市感知与模拟实验室. 中国90座城市建筑物屋顶矢量数据集[Z].
[21] Unisdr Terminology on Disaster Risk Reduction[R]. Unisdr,2009
[22] Wu J, Wang C, He X, et al. Spatiotemporal Changes in Both Asset Value and GDP Associated with Seismic Exposure in China in the Context of Rapid Economic Growth from 1990 to 2010[J]. Environmental Research Letters,2017,12(3):034002
[23] 孙柏涛,胡少卿. 基于已有震害矩阵模拟的群体震害预测方法研究[J]. 地震工程与工程振动,2005,25(6):102-108
[24] 尹之潜. 结构易损性分类和未来地震灾害估计[J]. 中国地震,1996,12(1):49-55
[25] 李华玥,聂高众,邓砚,等. 晋冀蒙交界区农村房屋建筑震害矩阵研究[J]. 自然灾害学报,2017,26(5):174-182
[26] 张合,吕国军,孙丽娜. 邢台市重要建筑中砖混结构震害预测[J]. 华北地震科学,2017,35(1):73-77
[27] 张孝奎. 极军遇地震在防灾减灾规划中的应用研究:以唐山市建筑抗震易损性分析为例[J]. 四川建筑科学研究,2018,44(5):8-14
[28] 聂永安. 天津市滨海新区(天津经济技术开发区)震害预测和防震减灾对策示范工程[Z]. 天津市:天津市地震局,2002-01-09