引文格式：陈超, 李晖, 袁莹. 面向国土调查成果的知识图谱构建[J]. 地理空间信息, 2025, 23(9): 13-17. doi:10. 3969/j.issn. 1672-4623.2025.09.004

面向国土调查成果的知识图谱构建

陈　超[1,2]，李　晖[1,2*]，袁　莹[1,2]
（1. 厦门理工学院 计算机与信息工程学院，福建 厦门 361024；2. 厦门理工学院 空间信息技术研究所，福建 厦门 361024）

摘　要：为了更好地利用国土调查成果、提高知识利用率，以第三次全国国土调查成果为例，首先针对国土知识空间特征进行空间关联分析，并结合国土知识属性特征，确定国土知识特征关系；再基于国土调查成果属性和时空特征提取国土知识语义关系，构建本体模型，建立模式层；最后融合国土调查数据进行国土知识实体提取、属性抽取，丰富数据层，构建国土调查成果知识图谱。结果表明，该国土调查成果知识图谱扩充了国土知识实体间的关联性表达，顾及了国土知识实体的空间特征表示，实现了国土知识实体间的空间查询功能，可为国土管理和空间规划提供一定的技术支持和土地知识服务。

关键词：国土调查成果；知识图谱；本体；空间特征；关联性
中图分类号：P208　　**文献标志码：**B　　**文章编号：**1672-4623（2025）09-0013-05

Knowledge Graph Construction for Territory Survey Achievements

CHEN Chao[1,2], LI Hui[1,2], YUAN Ying[1,2]
(1. College of Computer & Information Engineering, Xiamen University of Technology, Xiamen 361024, China;
2. Institute of Spatial Information Technology, Xiamen University of Technology, Xiamen 361024, China)

Abstract: In order to make better use of the territory survey achievements and improve the utilization rate of knowledge, taking the third territory survey achievements for example, we carried out the spatial correlation analysis according to the spatial characteristics of territorial knowledge, and determined the relationship of territorial knowledge according to the attributes of territorial knowledge at first. Then, based on the attributes and spatio-temporal characteristics of territory survey achievements, we extracted the semantic relation of territorial knowledge, constructed an ontology model, and established the pattern layer. Finally, we fused the territory survey data to extract the territorial knowledge entity and attribute, enrich the data layer, and realize the knowledge graph construction of territory survey achievements. The results show that the knowledge graph of territory survey achievements constructed in this paper expands the relevance expression of territorial knowledge entities, takes into account the spatial characteristic representation of territorial knowledge entities, and realizes the spatial query function between territorial knowledge entities, providing certain knowledge services and technical support for territorial spatial planning and management services.

Key words: territory survey achievement, knowledge graph, ontology, spatial characteristic, relevance

第三次全国国土调查（以下简称三调）获取的权威国土数据，是国家制定经济社会发展重大战略规划和政策举措的根本依据[1]。国土调查成果中蕴含着数量多、价值高、拥有时空特征的地理知识。目前地类数据的存储方式以表格、GIS数据为主，可展现国土知识属性特征和空间特征的可视化；但在语义关系表达国土知识特征方面略显不足。为进一步提高国土调查成果的知识利用率，辅助研究者挖掘更深层的国土知识关系，本文提出了知识图谱技术与国土调查成果相结合的研究思路。

知识图谱于2012年被提出，是一种以节点和边的结构形式将不同类型的实体、概念、关系和属性组合在一起的大型语义网络[2-3]。实体是可搜索的对象，关系是实体之间的关联，属性是对实体的具体描述。知识图谱在智能问答、搜索推荐、预测决策[4]等领域具有广泛应用，是人工智能时代的重要基础之一。具有代表性的通用知识图谱包括Freebase[5]、WikiData[6]和Yago[7]等；针对各行业领域知识构建的领域知识图谱包括水利领域知识图谱[8]、电力领域知识图谱[9]等；地学领域也出现了面向遥感数据的地学知识图谱[10]、面向时空地理和多源异构数据的地理知识图谱[11-13]，然而知识图谱在利用国土调查成果进行知识获取、管理和关联表达方面还有待进一步研究。自然资源部发布了TD/T 1055-2019《第三次全国国土调查技术规

项目来源：福建省自然科学基金资助项目（2023J011425）；厦门市科技计划资助项目（2022CXY0417）。
第一作者简介：陈超（1993—），硕士在读，研究方向为知识图谱，E-mail：xiangyaoganjue2022@163.com。
通信作者：李晖（1985—），博士，副教授，研究方向为地学知识图谱，E-mail：lih666@xmut.edu.cn。

程》[14]、《国土调查数据库标准（试行修订稿）》[15]等一系列技术规程和标准文件，确定了三调的基本原则和要求、农村和城镇各类土地使用调查情况[14]，明确了土地数据库建设标准[15]，建立的国土调查成果可为国土管理提供服务。本文采用知识图谱构建技术，结合三调成果，构建了面向国土调查成果的知识图谱；并基于知识图谱，对三调中的土地图斑知识实现了土地图斑空间查询、特征表示和空间关联性表示，为三调成果的知识服务提供了理论和技术支持。

1 国土知识特征

本文选取福建省某县国土调查数据，包括各类土地图斑、土地权属、基础特征等地理矢量数据，其中地类图斑是本文的主要研究对象。根据《第三次全国国土调查技术规程》[14]，地类图斑即为被行政区、城镇村庄等调查界线或土地权属界线分割的单一地类地块，其中城镇村庄内部同一地类的相邻宗地合并为一个图斑。整合地理对象之间的空间连接，在图谱中实现地理对象的空间关联。图是一种对一组对象（节点或顶点）及其交互作用（边）进行建模的数据结构[16]。本文中图数据结构是知识图谱底层存储数据的基本结构，每个地理对象对应一个节点，地理对象之间的空间关系对应边。选取福建省某县部分地类图斑矢量数据图，以地类图斑为地理对象，展示地理对象的空间特征关联（图1），颜色代表不同地类图斑的地理对象，数字代表不同地理对象。针对国土调查成果，所有的地类图斑以邻接关系为主，包含或被包含（岛屿等）可以看作特殊的邻接关系。在空间特征关联图中，以邻接作为地理对象之间的空间连接，若两个地理对象相互接触（共享其边界的一部分），则它们在空间上连接并成为彼此的“邻居”[16]。

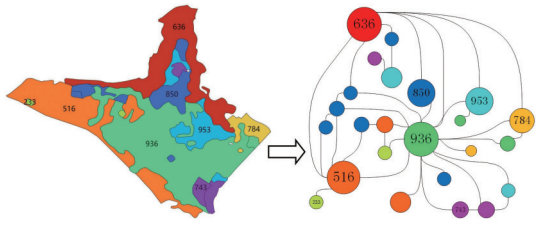


图1 地理对象空间特征关联图

2 国土调查成果知识图谱构建

2.1 国土调查成果知识图谱构建流程

国土调查成果存储在三调建立的标准数据库中，成果为GIS矢量数据，可转化为表格式的结构化数据。首先采用自顶向下的图谱构建方式，从各地类数据中抽取土地知识概念、属性和关系，分别建立概念之间以及概念与属性之间的语义关联，形成国土知识的本体库，建立知识图谱的模式层；然后基于国土知识本体库，从空间特征数据、时间属性数据和专题属性数据[17]中抽取国土知识实体、属性和关系，完成模式层和数据层之间的映射关系；最后按照图数据库中节点和边的数据存储形式，实现实体到节点、关系到边的转化和存储，构建国土调查成果知识图谱。

2.2 模式层构建

根据技术规程，以湿地、林地、园地、草地、耕地、交通运输用地、水域及水利设施用地、城镇村及工矿用地8类土地图斑为基础[1]，基于知识本体的思想，构建国土调查成果知识本体四元组[18]，即国土={概念集合，关系集合，属性集合，具体实例}，其中概念集合用于表示国土中的土地概念；关系集合用于表示对象和概念、概念和概念、对象和对象之间的关联关系；属性集合表示对象的属性；具体实例表示概念所映射的特定对象。

1）概念集合描述了耕地等8个国土本体的概念分类体系，包含土地图斑、权属单位、坐落单位等不同概念，涵盖国土知识中的空间要素、土地基本要素、土地利用要素等性质[15]。概念集合中的国土知识以节点的形式独立存在，不同概念之间存在语义关系。

2）具体实例表示国土知识概念的具体对象，如某块地类图斑的权属单位是概念，而具体的国家单位则是具体对象。单个概念可包含多个对象。

3）属性集合描述国土知识对象的属性。属性关系中的数据关系，具体实例是对象的属性，表示为实例—值的关系[19]，如某块土地对象拥有图斑面积，面积是对象属性，而面积值则是属性值。

4）关系集合。国土调查成果包括地类数据以及地类数据属性、时间和空间关系。属性关系中的对象关系是描述概念之间、概念与对象之间、对象之间的关联关系，包括等同关系、相似关系等。时间关系是获取地理数据的时间，包含公历时间、中国传统时间[20]。空间关系包括拓扑关系、方位关系、度量关系3个子类：拓扑关系表示地理实体点、线、面间的邻近和关联程度，包括包含、邻接和相离等关系[21-22]；方位关系表达地理实体在空间维度上的相对方位，包括东、南、西、北、东北、西北、东南、西南8个方

位；度量关系包括距离关系和尺寸位置关系，利用远、近、大、小等进行形式化表达[23]。模式层的关系类型见表1。

表1　模式层的关系类型

关系	关系子类	具体实例
空间关系	拓扑关系	相离、包含、邻接、覆盖
	方位关系	东、南、西、北、东南、东北、西南、西北
	度量关系	定性描述：远、近、非常近、非常远
		定量表达：距离500 m、东偏北300 m
时间关系	公历时间	年、月、日、min、h、s、星期
	传统时间	年号、节气、时辰、干支
属性关系	数据关系	名称、面积、宽度、编号、坐标、长度
	对象关系	父子关系、相似关系、等同关系、实例关系

整理《第三次全国国土调查技术规程》[14]、《国土调查数据库标准》[15]等相关文件，得到耕地、园地、林地、草地、湿地等本体对象，根据本体四元组模型，利用Protégé本体编辑工具构建国土调查知识本体库，部分本体展示见图2。以土地图斑耕地为例，知识图谱本体设计框架见表2。

表2　知识图谱本体设计框架

本体	对象	关系	概念	关系	属性
耕地	水田、水浇地、旱地	包含	地类图斑	拥有	图斑编号、图斑面积、地类编码、图斑细化代码、图斑地类面积、图斑细化名称
		拥有	耕地性质	拥有	耕地类型、耕地种植属性代码、耕地种植属性名称、耕地等别、耕地坡度级别
		属于	城镇村	拥有	城镇村用地类型、城镇村代码、城镇村名称、城镇村面积
		拥有	扣除地类	拥有	扣除地类编码、扣除地类系数、扣除地类面积
		隶属于	权属单位	拥有	权属单位代码、权属单位名称、权属性质
		隶属于	坐落单位	拥有	坐落单位名称、坐落单位代码

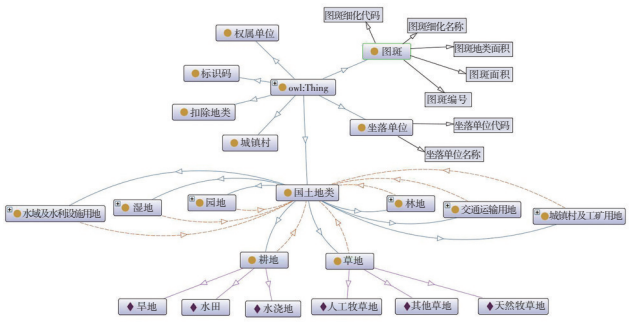


图2　部分国土调查知识本体

2.3　数据层构建

从国土调查数据出发，根据模式层的知识本体库，结合知识三元组结构[24]，分别将数据存入数据层，概念知识存入模式层。根据国土调查成果的特点，数据层构建流程可分为知识获取、知识抽取、知识整理和知识存储4个部分。

1）知识获取。2021年8月26日发布的《第三次全国国土调查主要数据公报》[1]对全国国土主要地类数据进行了统计公布，其中耕地、园地、林地、草地、湿地、城镇村及工矿用地、交通运输用地、水域及水利设施用地面积分别为12 786.19、2 017.16、28 412.59、26 453.01、2 346.93、3 530.64、955.31、3 628.79万hm^2。此次国土调查数据覆盖国家、省、地、县4级国土调查数据库，内容详细且结构清晰[1]。已获得的国土调查成果为GIS矢量数据，可转化为表格式的结构化数据，从中直接进行知识抽取，也能保证知识图谱的质量。

2）知识抽取。数据结构不同，其知识抽取的方式也不同。由于三调数据已实现结构化、并存储在表格中，本文直接从三调数据中提取实体、关系和属性。根据规程，基于地理先验知识和模式层的知识本体库，从数据中抽取旱地、水浇地等地理实体。以地理实体为空间对象，抽取实体间的空间关系、时间关系以及包含地理实体特征的属性关系。

3）知识整理。基于获得的土地图斑实体、属性和关系，根据地理专家知识构建知识三元组，实现“实体—关系—实体”[24]、“实体{属性：属性值}”的数据样式，如（旱地{地类编码：xxx}）—[坐落于]→（铺透村{坐落单位代码：xxx}）。将构建的知识三元组存储在表格中，并转化为csv文件，为后续的知识存储奠定数据基础。

4）知识存储。本文采用Neo4j作为实现知识图谱的数据库。Neo4j是图数据库中功能较强大的数据库之一，基于高性能、高可靠性、可扩展性强的开源图数据库，完全兼容原子性、一致性、隔离性和持久性[25]。Neo4j的基础数据存储形式为节点和边，节点存储实体、边存储关系，属性既可存储在节点中，也可存储在边中。两个节点之间用边进行连接，形成关系型网络结构[26]。

本文将存储国土知识三元组的csv文件导入Neo4j，利用Cypher语言导入csv文件中的实体、关系和属性，可查找Neo4j图数据库中某个节点的关系边以及关系边连接的其他节点。依据不同的Cypher语言可实现节点和边的增删改查，为国土调查成果知识的分析奠定技术基础。

3 实验和结果分析

基于上述知识图谱构建流程，结合福建省某县的三调数据构建国土调查成果知识图谱，统计得到558 259个实体数目，形成了1 753 170个关系数目（含重复），实体类型和关系类型统计见表3。

表3 国土调查知识图谱实体、关系类型及其数量

实体类型	实体类型数量	实体类型	实体类型数量	关系类型	关系类型数量
国土地类实体	102	海岛实体	7	实体属性关联	425 952
图斑属性实体	102 013	扣除地类实体	6 251	数值计算关联	19 221
城镇村实体	11 151	种植属性实体	14	发生时间	1
坐落单位实体	693	线状地物宽度	13	发生地点	319
权属单位实体	747	飞入地实体	2	包含关系	319
标识码实体	98 596	其他实体	99 528	从属关系	677

3.1 国土调查成果关联性表示

本文采用图数据库Neo4j存储国土调查成果，能存储、处理和表达数据间的复杂关系[27]。在Neo4j中采用Cypher语言对“标识码为350……749的地类图斑”进行关联性查询的结果见图3，可以看出：①通过点击查询到的节点，可显示显性或隐性关联的其他知识节点；②通过图数据库中边的结构来连接不同知识节点，边的标签可填写起始端和终端知识节点的关系和关系属性，直观反映了不同知识节点的关联特征；③通过图数据库节点和边的结构特点，结合国土知识的关联性可查询到隐性的关联知识，如由于“梅市村”这一权属单位，还查询到了权属单位相同，标识码为350……330的另一地类图斑及相关知识，扩充了国土知识范围。由此可见，国土调查成果知识图谱可通过语义关系查询国土知识之间的关联特征，为研究国土知识关联特征和知识延伸提供帮助。

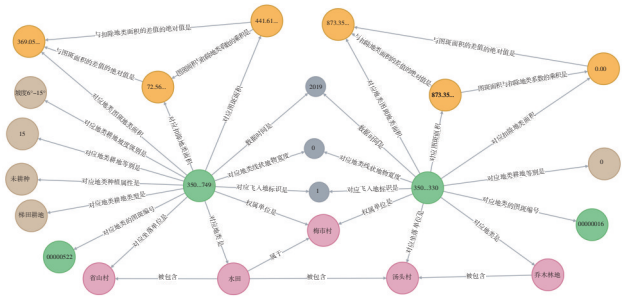


图3 国土调查成果知识图谱部分节点和关联性表示

3.2 国土调查成果空间特征表示

国土调查成果具有空间特征，本文构建的国土调查成果知识图谱利用图数据库节点和边的数据结构来表达国土知识中的空间特征。部分地类图斑和坐落单位的空间特征见图4。国土调查成果知识图谱不仅顾及了国土知识的空间特征，也弥补了GIS软件在数据空间特征语义关系表达方面的不足。

3.3 国土调查成果空间查询

Neo4j-Spatial是专门为实现Neo4j图数据库空间数据处理的一个插件，通过集成GIS能力，使Neo4j图数据库能存储、索引和查询空间数据，如可实现空间位置查询、探求两个空间点位的距离等。为验证构建的知识图谱的空间查询功能，本文基于Neo4j-Spatial空间插件，以某县的村级调查区数据为基础，以“以河福村为中心，通过空间查询找到距河福村不到3 km的其他村级调查区坐落单位点位”为例，对国土调查知识图谱进行空间查询。结果表明，共查询到11个距河福村不到3 km的点位，结果与ArcGIS等软件的结果相同，说明知识图谱可根据国土数据的空间特点，实现不同国土单位间的空间距离查询，为国土管理提供了一定的空间服务。

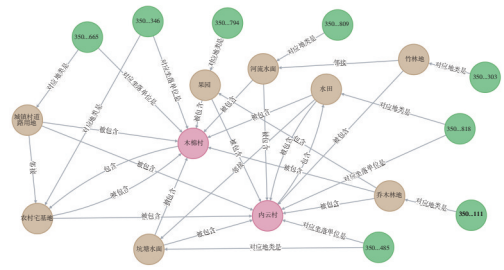


图4 国土调查成果知识图谱部分节点空间特征表示

4 结 语

随着大数据时代的到来，能清晰简明表达地理时空关联关系的地理知识图谱成为地理信息科学领域的研究热点。国土调查成果之间具有复杂的关系，从关联数据中挖掘含有时空特征变化、关联关系复杂的信息难度较大。为进一步利用国土调查成果，本文以三调成果为研究对象，梳理了国土调查成果的数据特征，分析了国土调查数据中用于构建图谱的本体、关系和属性；再根据领域型知识图谱构建方式，初步构建了面向国土调查成果的知识图谱；最后采用部分国土数据进行实验验证，分析了国土调查成果知识图谱的实际应用价值。结果表明，本文构建的国土调查成

果知识图谱通过国土知识的概念、关系和属性的语义关联，能够较完整地表达国土知识实体间丰富的关联关系和空间特征。

在知识图谱构建过程中还存在不足之处，如数据来源多样性不足；尚未涉及知识融合、知识更新等知识处理技术。在后续研究中将根据本体知识增加知识图谱数据来源的多样性，丰富知识图谱内容，补充地学规律信息，增加非结构化数据、地类遥感图像数据、地理矢量数据等。同时，面向国土调查成果权属溯源、年度变更等任务，将深入研究利用知识推理、时空查询、知识溯源等实现图谱的进化迭代，解决图谱数据实时更新问题并实现知识图谱在地理信息服务中的辅助作用，促进信息服务向知识服务转化，更好地为三调成果共享应用转化、推进国家治理体系和治理能力现代化服务。

参考文献

[1] 新华社.第三次全国国土调查主要数据成果发布[EB/OL].(2021-08-26) [2023-03-21]. http://www.gov.cn/xinwen/2021-08/26/content_5633497. htm

[2] 王昊奋,丁军,胡芳槐,等. 大规模企业级知识图谱实践综述[J]. 计算机工程,2020,46(7):1-13

[3] 张永军,程鑫,李彦胜,等. 利用知识图谱的国土资源数据管理与检索研究[J]. 武汉大学学报(信息科学版),2022,47(8):1 165-1 175

[4] 赵军,刘康,何世柱,等. 知识图谱[M]. 北京:高等教育出版,2018

[5] Bollacker K,Evans C,Paritosh P,et al. Freebase: A Collaboratively Created Graph Database for Structuring Human Knowledge[C]. International Conference on Management of Data,Vancouver,2008

[6] Erxleben F, Günther M, Krötzsch M, et al. Introducing Wikidata to the Linked Data Web[C]. Springer International Publishing,Cham,2014

[7] Hoffart J, Suchanek F M, Berberich K, et al. YAGO2: A Spatially and Temporally Enhanced Knowledge Base from Wikipedia[J]. Artificial Intelligence,2013,194:28-61

[8] 段浩,韩昆,赵红莉,等. 水利综合知识图谱构建研究[J]. 水利学报,2021,52(8):948-958

[9] 蒲天骄,谈元鹏,彭国政,等. 电力领域知识图谱的构建与应用[J]. 电网技术,2021,45(6):2 080-2 091

[10] 王志华,杨晓梅,周成虎. 面向遥感大数据的地学知识图谱构想[J]. 地球信息科学学报,2021,23(1):16-28

[11] 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

[12] 刘俊楠,刘海砚,陈晓慧,等. 面向多源地理空间数据的知识图谱构建[J]. 地球信息科学学报,2020,22(7):1 476-1 486

[13] 蒋秉川,万刚,许剑,等. 多源异构数据的大规模地理知识图谱构建[J]. 测绘学报,2018,47(8):1 051-1 061

[14] 自然资源部. 自然资源部关于发布《第三次全国国土调查技术规程》《县级国土调查生产成本定额》2项行业标准的公 告 [EB/OL]. (2019-10-15) [2023-07-17]. https://www.gov.cn/zhengce/zhengceku/2019-10/15/content_5439991.htm

[15] 自然资源部. 国务院第三次全国国土调查领导小组办公室关于发布《国土调查数据库标准(试行修订稿)》的通知[EB/OL]. (2019-04-12)[2023-07-17]. https://www. gov. cn/xinwen/2019-04/12/content_5381961. htm

[16] Zou X, Liu X, Liu M, et al. Spatio-temporal Graph-based Analysis of Land Cover Evolution Using Remote Sensing Time Series Data[J]. International Journal of Geographical Information Science,2023,37(5):1 009-1 040

[17] 李国斌,汤永利. 空间数据库技术[M]. 北京:电子工业出版社,2010

[18] 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

[19] 王东旭,诸云强,潘鹏,等. 地理数据空间本体构建及其在数据检索中的应用[J]. 地球信息科学学报,2016,18(4):443-452

[20] 侯志伟,诸云强,高星,等. 时间本体及其在地学数据检索中的应用[J]. 地球信息科学学报,2015,17(4):379-390

[21] 张雪英,张春菊,杜超利. 空间关系词汇与地理实体要素类型的语义约束关系构建方法[J]. 武汉大学学报(信息科学版),2012,37(11):1 266-1 270

[22] Clementini E, Sharma J, Egenhofer M J. Modelling Topological Spatial Relations: Strategies for Query Processing[J]. Computers and Graphics,1994,18(6):815-822

[23] 张斌,李宏伟,崔铁军. 全息地理空间本体模型构建[J]. 地理空间信息,2021,19(12):1-6

[24] 董佳琳,张宇航,徐永康,等. 基于知识图谱的新冠疫情智能问答系统[J]. 信息技术与信息化,2021(6):258-261

[25] 寿增,狄跃斌,马骁,等. 基于知识图谱的网络威胁行为检测系统设计[J]. 无线电工程,2022,52(8):1 330-1 337

[26] 王红,张青青,蔡伟伟,等. 基于Neo4j的领域本体存储方法研究[J]. 计算机应用研究,2017,34(8):2 404-2 407

[27] 庞国明,叶伟民,宋建栋,等. 精通Neo4j[M]. 北京:清华大学出版社,2022