



# 利用知识图谱的国土资源数据管理与检索研究

张永军<sup>1</sup> 程鑫<sup>1</sup> 李彦胜<sup>1</sup> 王飞<sup>1</sup> 刘晓健<sup>1</sup> 吴文品<sup>1</sup>

<sup>1</sup> 武汉大学遥感信息工程学院,湖北 武汉,430079

**摘要:**针对国土资源不同数据产品间难以进行有效管理与快速应用的问题,研究利用图数据库对GlobaLand30、FROM-GLC10\_2017、GLC\_FCS30\_2020等公开土地覆盖数据集进行语义层面的结构化存储,建立中国国土资源知识图谱。构建以行政区划为单位进行土地覆盖数据产品管理、知识提取以及数据获取与更新的新型应用框架,利用基于图的异常数据检索算法探究不同产品间的一致性,提出了一种基于知识图谱的感兴趣图节点快速检索算法。通过引入知识图谱,形成了具有447 817个节点、447 816条关系,且可动态更新的中国国土资源知识图谱,并发现了在覆盖全国的2 875个行政单元中有92个区域单元的产品数据一致性不足60%,区域产品精度可能存在较大误差。充分利用了多源土地覆盖数据产品间的信息,缩短了数据预处理的时间,为中国国土资源的知识化管理与应用提供了新思路。

**关键词:**国土资源管理;知识图谱;土地覆盖数据产品;异常数据检索;感兴趣图节点

**中图分类号:**P208

**文献标志码:**A

国土资源数据对于国家粮食安全、环境适应性评估、城市化进程等研究有着至关重要的作用<sup>[1-2]</sup>。近年来,随着遥感卫星影像数据量的快速增长以及深度学习在计算机视觉方面的发展,众多全球性的土地覆盖数据产品被研究者公开,是国土资源数据的重要来源之一。

当前主要利用网格化技术进行全球尺度土地覆盖数据产品的管理,在地学领域,网格化是一种对地理空间进行划分的方法<sup>[3]</sup>。以经纬度来划分地球网格是目前全球性土地覆盖数据产品最广泛采用的数据划分方式之一,它能够有效地描述地球表面的空间位置<sup>[4]</sup>,但不同数据产品之间所采用的网格大小不尽相同。GlobaLand30数据集在60°S~60°N区域内,按照5°(纬度)×6°(经度)的大小进行分幅,在南北半球纬度60°~85°区域内,按照5°(纬度)×12°(经度)的大小进行分幅<sup>[5]</sup>;1985—2020年全球30 m地表覆盖精细分类产品GLC\_FCS30\_1985—2020按照5°(纬度)×5°(经度)的大小进行分幅<sup>[6-8]</sup>;2017年全球10 m地表覆盖分类产品FROM-GLC10\_2017按照2°(纬度)×2°(经度)的大小进行分幅<sup>[9]</sup>。这种划分尺

度的不统一导致国土资源管理难以找到一个合适的网格大小来满足众多公开土地覆盖数据产品的管理。同时,不同尺度的网格化管理方式难以在多个数据产品之间建立关联,导致研究者难以充分利用多数据集的优势。

知识图谱是一种以节点和边的形式将不同类型的实体、概念有结构地组合在一起的巨大语义网络<sup>[10]</sup>,是迈向第三代人工智能的重要技术之一<sup>[11]</sup>。在地学领域,有关知识图谱的研究一方面集中在地学知识的提取,即从海量的地学文献、网页文本等非结构化数据中提取时间、地点、名胜古迹等地理实体及其关系<sup>[12-13]</sup>;另一方面集中在地学知识图谱的构建、管理、更新与利用知识图谱进行知识推理、知识计算、推荐等应用<sup>[14-16]</sup>。在大数据时代下,知识图谱在多源数据融合、知识推理、地理大数据分析等方面有着极大的优势与潜力<sup>[17]</sup>。然而知识图谱在利用遥感影像与遥感数据产品进行知识获取、管理与应用方面的研究还有待进一步展开。

本文基于传统网格化思想,以中国分级行政区划作为网格单元对现有部分公开土地覆盖数

收稿日期:2021-12-19

项目资助:国家自然科学基金(42030102,41971284);湖北省自然科学基金创新群体(2020CFA003)。

第一作者:张永军,博士,教授,主要从事航空航天摄影测量遥感一体化研究。zhangyj@whu.edu.cn

通讯作者:李彦胜,博士,副教授。yansheng.li@whu.edu.cn

据集进行网格化划分与管理,解决了不同数据集网格划分尺度不统一的问题。同时,根据人脑对于数据理解与记忆的方式以及学者们对于土地覆盖产品的利用方式<sup>[18-20]</sup>,本文构建了包含447 817个节点、447 816条关系的可动态更新的中国国土资源知识图谱,并在此基础上探究了基于该图谱的异常数据检索、感兴趣图节点(graph node of interest, GNOI)数据快速检索等方面的应用,使不同数据产品间的信息得到充分利用,发挥了知识化管理的优势,并缩短了数据预处理的时间,为知识图谱在地学领域的应用以及公开数据产品的知识化管理等方面提供了新思路。

# 1 中国国土资源知识图谱构建

## 1.1 国土资源本体建模

国土资源知识建模是指构建与国土资源相关的概念、属性以及概念之间关系的本体模型,对国土资源知识进行描述<sup>[21]</sup>。知识模型构建方式通常包括自顶向下和自底向上两种<sup>[22]</sup>,本文主要采用了自顶向下的方式,结合具体的应用需求,人工构建了国土资源的本体模型,如图1所示。与注重知识广度的通用知识图谱本体构建不同,该模型更加注重国土资源管理所涉及到的本体,虽然在本体模型的通用性上不及通用知识图谱,但是在知识的深度方面更具优势。

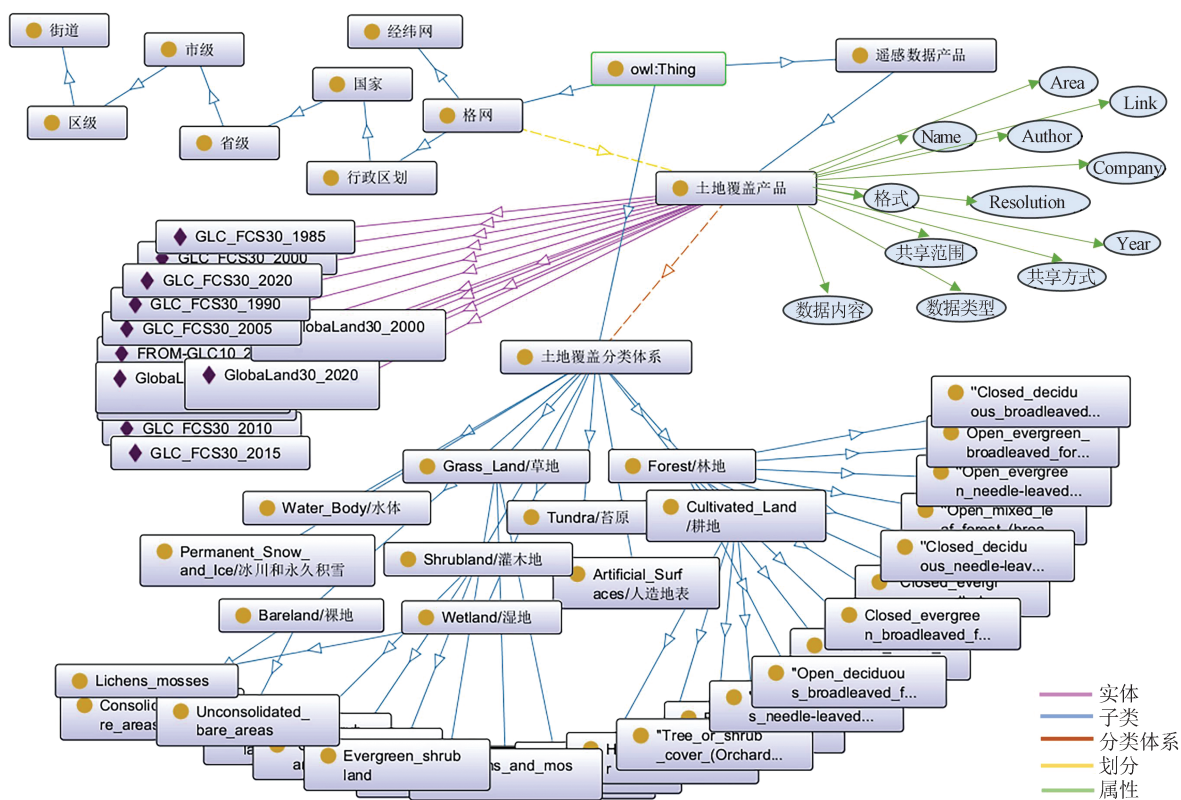


图1 国土资源知识图谱本体模型

Fig.1 Ontogeny Model of Land Resources Knowledge Graph

本文构建的国土资源知识图谱本体主要有网格、遥感数据产品与土地覆盖分类体系三大类。其中,遥感数据产品是核心,土地覆盖分类体系是对其内容的描述,网格是对其位置的描述。三者构成了一个能够描述遥感数据产品内容和位置的本体知识模型。

行政区划网格本体主要依据传统经纬度网格的划分进行,目前公开的土地覆盖产品大多数以经纬度为基础对数据进行划分与管理,但是经纬度与数据有关的知识联想十分有限,如2020年

GlobaLand30数据集中的n50\_30\_2020lc030.tif网格影像切片,在不进行数据下载与预处理的情况下,除了能够直接推算出影像范围在通用横墨卡托格网系统(universal transverse Mercator grid system, UTM)投影50带,30°N~35°N外,对影像内容难以有进一步认识。然而通过行政区划网格进行管理,该网格切片包含了武汉市武昌区,在不探究数据本身的情况下,研究者也能对产品的位置信息以及内容有一个常识性的认识,如武汉市武昌区占比较大的土地类型可能是不透水

面、水体等,而出现冰雪、裸地大概率是产品分类错误导致。此外,通过比较基于土地覆盖产品数据集进行的相关研究发现<sup>[23-26]</sup>,学者通常不直接利用以经纬度划分的数据,而是经过大量的数据拼接、裁剪等预处理工作,得到行政区划级别的数据。因此,以行政区划为网格的数据管理在数据利用效率上更具优势,本文将中国的行政区划划分为5个层级,分别对应本体模型图中的国家、省级、市级、区级、街道级5个本体节点,这样的层次结构相比经纬度网格有着更加丰富的语义关系。

遥感数据产品本体下包含土地覆盖产品本体,该本体与格网本体之间存在划分关系,并且为了满足应用的需求,添加了“Year”“Name”“Resolution”等12种属性。土地覆盖分类体系本体与土地覆盖产品本体之间存在分类体系关系,其子节点由土地覆盖产品所使用的分类体系决定。表1展示了部分国土资源知识图谱本体模型的三元组示例。整个国土资源知识图谱本体模型与应用需求紧密结合,并且可以随着数据以及应用需求的增加进行增量式扩充。

表 1 国土资源知识图谱本体模型三元组示例

Tab.1 Examples of Triples in the Ontology Model of Land Resources Knowledge Graph

节点	三元组
土地覆盖产品	(土地覆盖产品,SubClassOf,遥感数据产品)
	(土地覆盖产品,dataProperty,分辨率)
	(土地覆盖产品,分类体系,土地覆被分类体系)
行政区划	(行政区划,SubClassOf,网格)
	(行政区划,HasSubClass,国家)
	(行政区划,划分,土地覆被产品)
土地覆被分类体系	(土地覆盖分类体系,HasSubClass,耕地)
	(土地覆盖分类体系,HasSubClass,林地)
	(土地覆盖分类体系,HasSubClass,草地)
	(土地覆盖分类体系,HasSubClass,灌木地)
	(土地覆盖分类体系,HasSubClass,湿地)
	(土地覆盖分类体系,HasSubClass,水体)
	(土地覆盖分类体系,HasSubClass,苔原)
	(土地覆盖分类体系,HasSubClass,不透水面)
	(土地覆盖分类体系,HasSubClass,裸地)
(土地覆盖分类体系,HasSubClass,永久性冰雪)	

## 1.2 国土资源知识获取

### 1.2.1 行政区划矢量数据

本文所需行政区划数据来源于高德地图提供的行政区划应用程序接口(application programming interface, API),获取的行政区划数据按照

层级可划分为国家、省、市、区、街道,矢量数据最高可获取到区级。高德地图所采用的坐标系为火星坐标系 GCJ-02, 本文将其转化为 1984 世界大地坐标系(world geodetic system 1984 coordinate system, WGS-84),并在国家地理信息公共服务平台天地图系统上对数据准确性进行了验证,结果表明数据准确率为 100%。随着中国城市化进程的加速,中国行政区划每年都会有一定的调整。从动态接口获取数据可以动态更新知识图谱中的行政区划数据,并且灵活的图结构对于数据的更新与变动更具优势。

### 1.2.2 土地覆盖数据集

本文所用土地覆盖数据集包含了近年来发布的较为流行的全球性产品,产品信息如表 2 所示。

表 2 土地覆被产品信息表

Tab.2 Table of Land Cover Product Information

数据集名称	分辨率/m	年份	图幅数/幅	坐标系
GlobaLand30_2000	30	2000年	853	WGS-84坐标系;85°S~85°N: UTM投影;南北半球纬度
GlobaLand30_2010		2010年		85°~90°:极地
GlobaLand30_2020		2020年		966
GLC_FCS30_1985	30	1985年	961	WGS-84坐标系
GLC_FCS30_1990		1990年		
GLC_FCS30_1995		1995年		
GLC_FCS30_2000		2000年		
GLC_FCS30_2005		2005年		
GLC_FCS30_2010		2010年		
GLC_FCS30_2015		2015年		
GLC_FCS30_2020		2020年		
FROM-GLC10_2017		10		

土地覆盖数据集处理的核心在于计算出行政区划单元所涉及的栅格图幅,进行栅格数据的镶嵌与裁剪,从而得到行政区划大小的数据,并统计所获数据中的语义信息。本文将统计信息作为一种知识节点,由于统计信息使用频率较高,并且人们通常不会记忆像素信息,而是根据土地分类影像中的地物种类和占比等语义信息决定是否采用该数据,这些信息可以通过统计获得。由于行政区划存在层级包含关系,因此本文只利用具有矢量数据且没有包含关系的最小行政区划单元对栅格数据进行处理,共划分为不重叠的 2 875 个行政区划单元,其中省级 1 个,市级

34个,区级2840个。

### 1.3 国土资源知识图谱构建与存储

构建国土资源知识图谱的核心是确定本体模型与知识图谱的映射关系,实现本体概念层级结构到知识图谱语义网络的转换<sup>[27]</sup>,转化规则如图2所示。本体模型中的国家、省级、市级、区级、

街道节点分别对应知识图谱 country、province、city、district、street 标签下的节点,土地覆盖产品和土地覆盖分类体系对应知识图谱中 statistics 标签下的节点和节点间关系,这样的映射关系有利于后续应用中的检索算法以及数据的扩展,但是并不唯一。

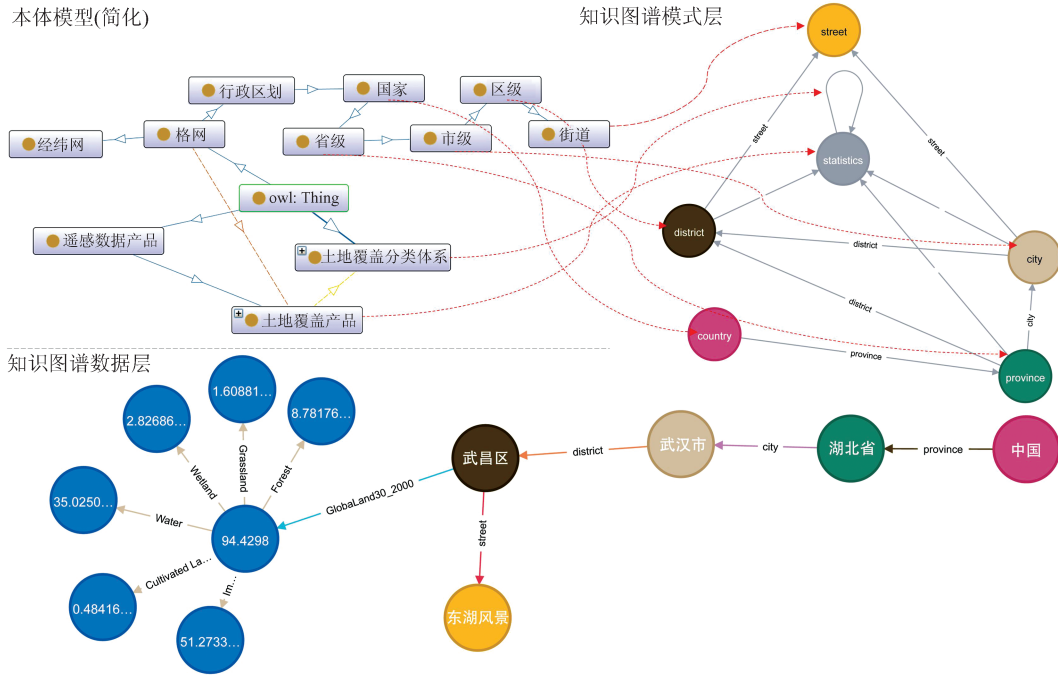


图2 本体与知识图谱映射图

Fig.2 Mapping Figure from Ontology to Knowledge Graph

三元组是知识图谱的一种通用表示方式,即  $G=(E, R, S)$ ,其中  $E=\{e_1, e_2 \dots e_{n_E}\}$  是知识库中的实体集合,共包含  $n_E$  种不同实体; $R=\{r_1, r_2 \dots r_{n_R}\}$  是知识库中的关系集合,共包含  $n_R$  种不同关系; $S \subseteq E \times R \times E$  表示知识库中的三元组集合<sup>[28]</sup>。三元组的存储方式灵活,基于效率考虑,本文采用了更易于数据管理的图数据库 Neo4j 进行存储。Neo4j 是一个开源的原生图数据库,带有查询语言以及丰富的第三方驱动程序。本文将土地覆盖数据集按照设计的映射规则,构建了包含 447 817 个节点、447 816 条关系的可动态更新的中国国土资源知识图谱,并将其存储到 Neo4j 中,如图 3 所示。

## 2 基于知识图谱的中国国土资源数据获取与更新

### 2.1 国土资源数据获取

传统数据获取与处理通常要经过研究区域数据筛选、栅格投影变换、栅格镶嵌与裁剪等步

骤,且对于不同数据产品,处理算法不具有通用性。与传统从土地覆盖数据集中进行数据获取相比,利用知识图谱管理的国土资源数据获取更加语义化、高效、简单和符合人的思维,只需要明确研究区域,就可以通过通用的数据处理流程获取知识图谱管理下的国土资源数据,流程如图 4 所示。

基于知识图谱的国土资源数据获取可分为 3 条处理流:(1)区级行政区划数据可以直接利用图检索算法获取相关知识节点,并与栅格数据链接,以武汉市各区国土资源数据获取为例,数据获取结果如图 5 所示;(2)区级以上行政区划数据由于没有直接对原始数据进行知识存储,无法直接获取,但是知识图谱具有丰富的语义关系,可以利用图节点计算得到所需数据,以武汉市国土资源数据获取为例,通过武汉市关联的区级节点动态计算武汉市国土资源数据节点,并与武汉市进行连接,结果如图 6 所示;(3)自定义区域的数据获取与前两种情况的区别在于需要通过自定义区域矢量数据与中国行政区划矢量进行简单

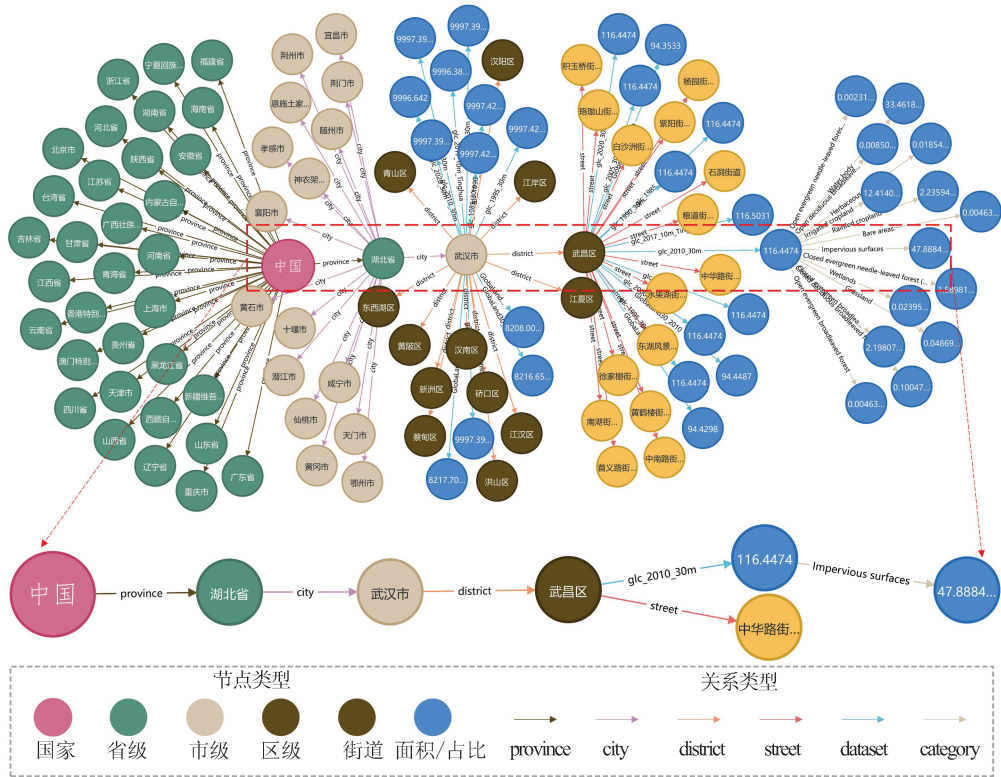


图 3 国土资源知识图谱(局部)

Fig.3 Land and Resources Knowledge Graph (Part)

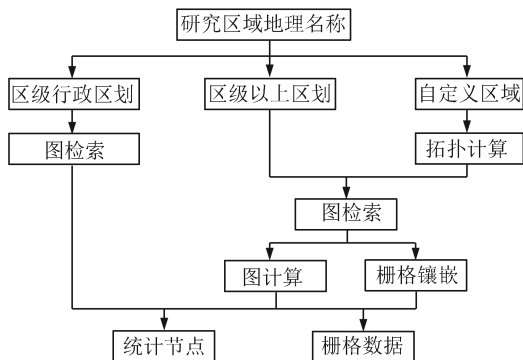


图 4 基于知识图谱的国土资源数据获取流程

Fig.4 Flowchart of Land Resources Data Acquisition Based on Knowledge Graph

的拓扑计算,以确定自定义区域所涉及的行政区划节点。

基于知识图谱管理的国土资源数据获取方法减少了大量数据预处理工作,其数据的返回形式包括知识图谱和栅格数据。对于大多数不涉及空间格局的研究,知识图谱中的统计节点数据能够满足研究者的数据需求。利用知识图谱辅助管理对于数据的服务者和使用者都是高效的,并给数据服务者提供了一种创新性的数据管理与发布形式。

### 2.2 国土资源数据更新

随着遥感技术的快速发展,每年有大量的遥

感产品产生,并且大量科研工作者将高质量的研究数据共享,这些数据对于遥感应用领域有着重要作用。其中大量国土资源相关数据需要及时更新到中国国土资源知识图谱中,确保研究者能够获取对于研究最有价值的知识。Neo4j 利用原生图结构存储知识,只需将新数据按照 §1.2 的流程进行数据预处理和知识提取,再以新的节点和关系存储到中国国土资源知识图谱中即可实现更新。

本文使用的中国行政区划数据通过动态接口获得,已经考虑了更新的问题。但是受经济、地理位置等因素影响,中国还划分出来很多具有特别含义的区域,如华东地区、东三省区域、长江经济带、京津冀都市圈、长江三角洲地区等。这些数据的更新只需将区域节点与知识图谱中相关行政区划节点相关联即可。由于行政区划相连的数据节点未受到任何影响,因此,区域数据的获取与区级以上行政区划数据的获取完全相同,本文利用长江三角洲地区矢量数据更新国土资源知识图谱,更新流程及结果如图 7 所示。

知识图谱对于多源数据有较强的融合能力,不仅可以用栅格数据和矢量数据来更新知识图谱,还可以利用文本、表格、数据库等数据形式更新知识图谱,这是知识不断迭代进化的过程。由

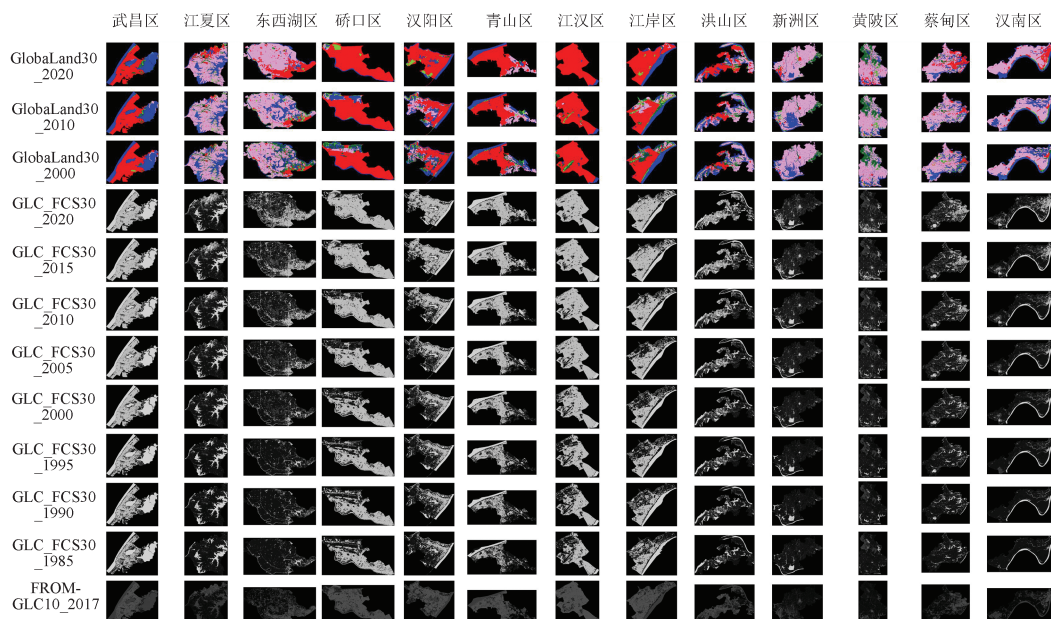


图5 武汉市各区级国土资源数据

Fig.5 Data of Land Resources in All Districts of Wuhan City

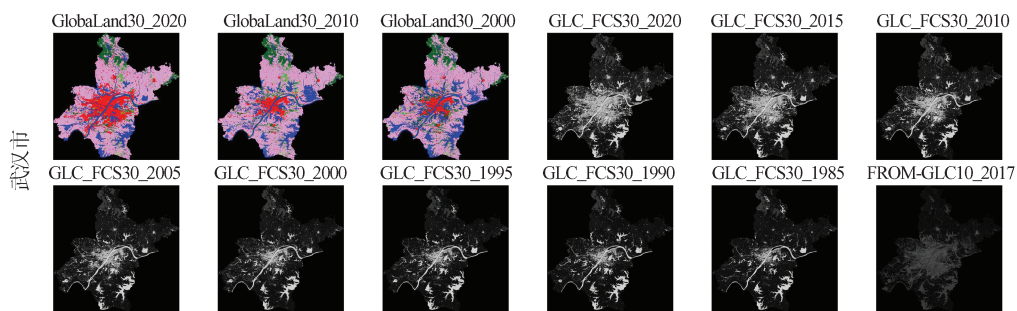


图6 知识图谱生成的武汉市国土资源数据获取结果

Fig.6 Results of Wuhan Land Resources Data Achievement Generated by Knowledge Graph

此,本文总结了耦合知识图谱的国土数据管理完整工作流程:数据源、数据处理、知识建模与存储、关键算法、应用构建与迭代更新。数据源是知识的来源,数据处理部分的核心在于将数据源中的知识以结构化的方式提取出来,知识建模与存储的核心在于将提取的知识高效地进行存储管理,关键算法为应用提供算法支持,并且随着应用的丰富以及数据源的不断增加,能够重复上述过程对知识图谱不断地进行知识迭代进化,使知识图谱能够不断地扩展其知识的广度和深度。

### 3 潜在异常数据发现与GNOI数据快速检索

#### 3.1 基于图的潜在异常数据发现算法

尽管全球性的土地覆盖数据产品的精度在不断提高,GlobaLand30\_2020总体精度达到85.72%,但是小区域的精度仍然无法保证能够超

过80%,尤其是景观和地形复杂的地区<sup>[29]</sup>。在利用土地覆盖产品的相关研究中,大多数研究者只选择一套数据集作为数据源,在没有同类型数据对照的情况下常常会忽略可能存在的区域精度问题,导致研究结论有所偏差<sup>[30]</sup>。同时,不同产品之间由于采用的数据源、模型以及分类体系的不同,难免会产生数据不一致问题,且差异的大小与区域有关<sup>[31-33]</sup>。

本文创建的国土资源知识图谱包含12套土地覆盖产品,尽管本研究无法衡量数据集精度的好坏,但是能够充分挖掘知识图谱的优势,在不同产品之间一致性存在较大差异的行政区划节点给予数据使用者提示,由使用者综合判断以选取最适合其研究的数据,使数据集能够得到充分的利用。

本文比较了GlobaLand30\_2020和GLC\_FCS30\_2020数据集中国区域的分类结果,由于两套数据集采用了不同的分类标准,在计算前需

进行分类系统的统一。通过目视比较,对上述产品的分类体系进行聚合,最终形成9个类,分别为

林地、灌木地、耕地、草地、裸地、内陆水体、不透水面、永久性冰雪和湿地,如表3所示。

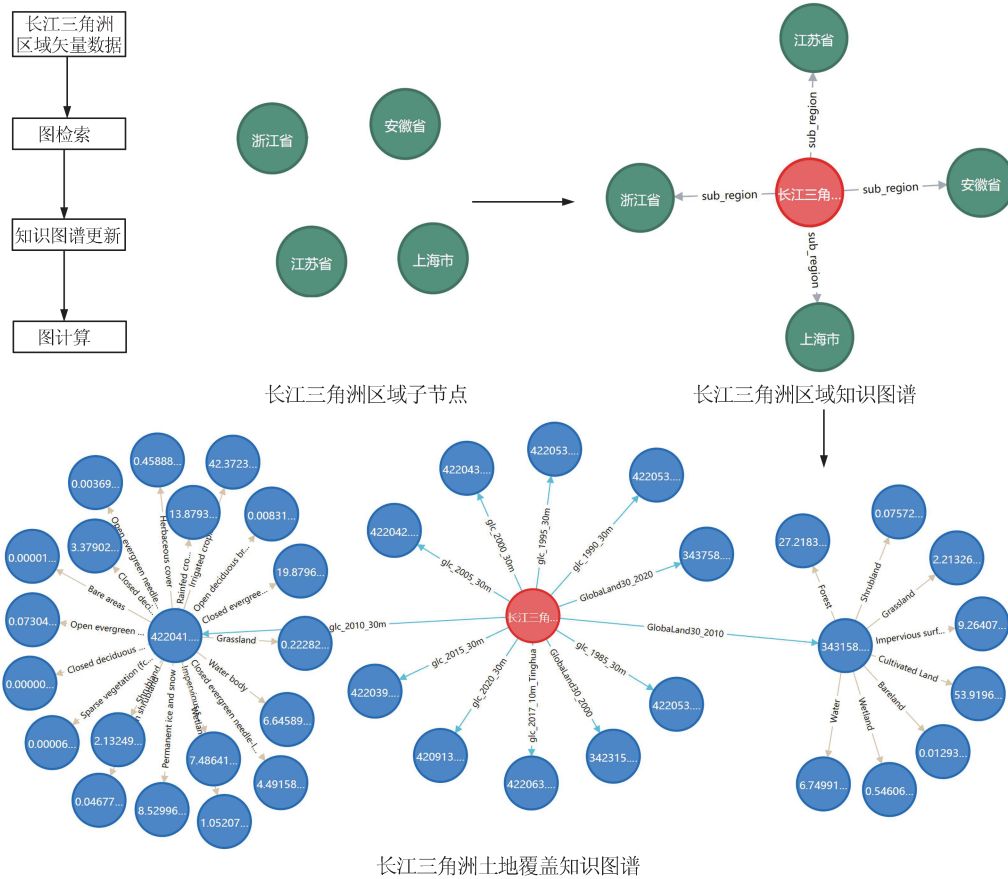


图7 长江三角洲地区的知识图谱更新流程

Fig.7 Flowchart of Knowledge Graph Update Using the Yangtze River Delta

表3 产品类别聚合表

Tab.3 Table of Category Aggregated for Products

产品名称	1耕地	2林地	3草地	4灌木地	5湿地	6内陆水体	7裸地	8不透水面	9永久性冰雪
GlobaLand30_2020	10	20	30,70	40	50	60	90	80	100
GLC_FCS30_2020	10~20	50~100	130~160	120;121;122	180	210	200;201;202	190	220

注:表中的数字1~9代表不同遥感产品原始分类体系类别代码

数据的一致性评价分为2个方面:(1)面积一致性,采用的指标为皮尔森相关系数R和百分比不一致性P;(2)空间一致性,采用的指标为不同地类的一致性系数Mi和总体一致性系数N。4个指标的计算如下:

$$R = \frac{\sum_{k=1}^n (X_k - \bar{X})(Y_k - \bar{Y})}{\sqrt{\sum_{k=1}^n (X_k - \bar{X})^2 \sum_{k=1}^n (Y_k - \bar{Y})^2}} \times 100\% \quad (1)$$

$$P = \frac{|X_k - Y_k|}{X_k + Y_k} \times 100\% \quad (2)$$

$$M_i = \frac{P_i}{(X_i + Y_i)/2} \times 100\% \quad (3)$$

$$N = \frac{\sum_{i=1}^n P_i}{T} \times 100\% \quad (4)$$

式中,n为类别数目;X<sub>k</sub>和Y<sub>k</sub>分别为GloLand30\_2020和GLC\_FCS30\_2020数据集中k类别的总面积;X̄和Ȳ分别为两数据集所有土地覆盖类别的平均面积;X<sub>i</sub>、Y<sub>i</sub>分别为两数据集中第i种类别的像元数;P<sub>i</sub>为在同一位置两种产品都为类别i的像元数;T为总像元数。

基于图的潜在异常数据发现算法主要由栅格数据空间总体一致性计算、面积一致性节点计算和图检索算法构成。本文首先在图上动态计算出各区域节点的空间总体一致性节点和面积

一致性节点,并连接到图上,然后在整个图上运行检索算法。当两套数据集在同一区域的空间一致性和皮尔森相关性都不足60%时,表明该区域分类精度可能较差。检索发现2875个行政区划中有92个区域数据质量可能较差,782个行政区划数据的空间总体一致性和皮尔森相关性同时超过了80%,数据较为可靠。

对于两数据产品空间总体一致性和皮尔森相关性都表现较好的区域,本文所提国土资源知识图谱仍然能够挖掘出一些有价值的信息,使数

据能够得到合理的利用。以武昌区为例,数据一致性评价指标计算结果如图8所示,两数据产品在该区域空间总体一致性和皮尔森相关性分别达到85.37%和97.92%(蓝色框),但是GlobalLand30\_2020在该区域不存在湿地和裸地类型,导致该类别的空间一致性和百分比不一致性分别为0%和100%(红色框),空间一致性和百分比不一致性较好的类别是水体和不透水面,说明两数据集在武昌区比较适合做与水体和不透水面相关的研究,而不适合做与裸地、湿地相关的研究。

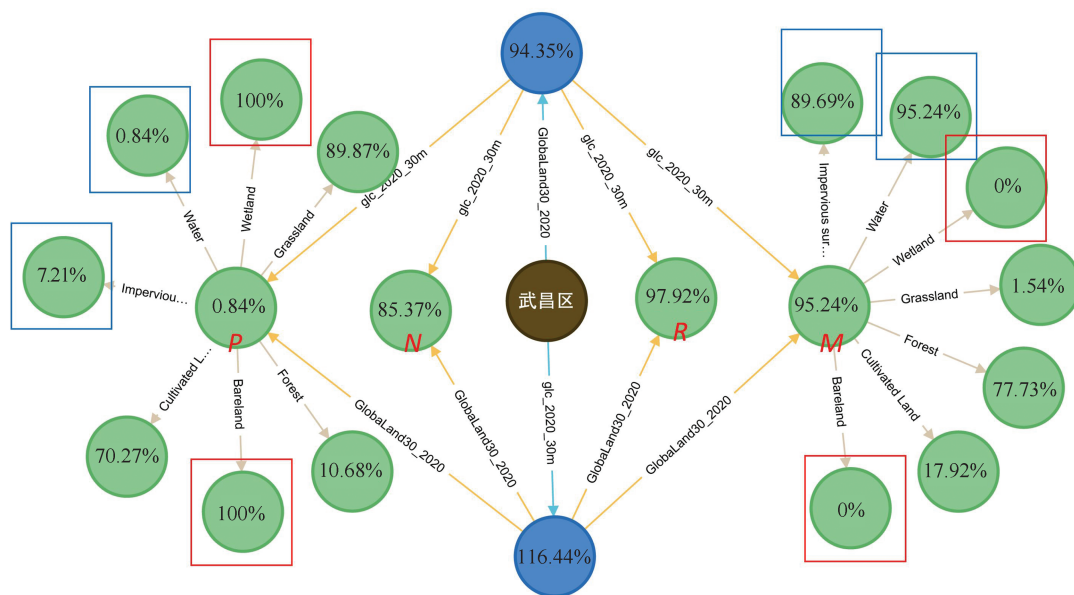


图8 武昌区知识图谱评价指标结果

Fig.8 Results of Evaluation Indicators of Knowledge Graph of Wuchang District

### 3.2 基于知识图谱的GNOI数据快速检索

在地学领域,感兴趣点(point of interest, POI)是一种与用户相关的常用数据。在知识图谱管理的国土资源数据中,使用者可能会对一些包含知识的节点感兴趣,例如根据土地覆盖类型评价的宜居性区域节点、城市化进程较快的区域节点等。因此,本文提出GNOI,并以中国城市化进程为例演示基于知识图谱的GNOI数据快速检索算法。

本文使用GLC\_FCS30\_1985和GLC\_FCS30\_2020土地覆盖数据集,两数据集的数据源选取和处理模型使用具有一致性,适合时序方面的研究。所提国土资源知识图谱存储了相关数据,并且相关节点附有年份信息,因此,利用GNOI检索算法可以快速筛选出1985—2020年中国城市化最快的10个区域的GNOI,并且该算法可以扩展到其他GNOI的快速检索,算法流程如图9所示。其中,GNOI衡量指标为城市增长

率,需要检索区级节点以及GLC\_FCS30\_1985、GLC\_FCS30\_2020数据不透水面类型节点,经过求差和排序处理后即可得到GNOI检索结果,如表4所示。在搭载Intel i5-10500 CPU的Windows 10系统机器上,利用Neo4j 4.3.2社区版软件,进行10次相同检索,从447 817个节点、447 816条关系的中国国土资源知识图谱中获得实验所需的10个GNOI,平均耗时为208 ms,大大缩短了目标数据获取的时间。

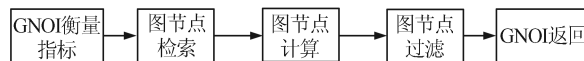


图9 GNOI快速检索算法流程图

Fig.9 Flowchart of Fast Retrieval Algorithm for GNOI

## 4 结语

针对多源国土资源数据难以管理和充分利用的问题,本文提出了利用行政区划网格进行数据划分,根据知识图谱对划分数据进行知识化管



表 4 区域城市面积增长率统计表/%

Tab.4 Statistical Table of Regional Urban Area Growth Rate/%

城市	区域	1985年面积占比	2020年面积占比	增长占比
成都市	武侯区	24.17	92.49	68.32
西安市	未央区	16.23	79.24	63.01
长沙市	芙蓉区	24.63	86.11	61.48
西安市	雁塔区	28.69	87.08	58.39
成都市	金牛区	29.28	86.61	57.33
郑州市	管城回族区	15.53	72.04	56.51
福州市	仓山区	12.80	68.47	55.67
杭州市	滨江区	6.59	61.29	54.70
深圳市	龙华区	1.56	55.94	54.38
石家庄市	裕华区	30.26	82.97	52.71

理的方法。基于国土领域知识和土地覆盖数据,构建了国土资源知识图谱本体模型以及包含 447 817 个节点、447 816 条关系,且可动态更新的国土资源知识图谱。在数据获取方面,基于知识图谱的国土资源数据获取与传统方式相比,具有更加语义化、高效、符合人的思维,并且可以同时获取知识图谱语义化数据以及栅格产品数据等优势。在数据挖掘方面,根据国土资源知识图谱丰富的语义关系以及图结构管理的灵活性,本文设计了潜在异常数据检索算法,发现 Global-land30\_2020 和 GLC\_FCS30\_2020 数据集在覆盖全国的 2 875 个行政单元中有 92 个区域单元的数据一致性不足 60%,显示产品在该区域数据分类精度上可能存在较大误差。同时首次提出了 GNOI,并以中国 30 年来城市增长最快的 10 个区域 GNOI 检索为例,验证了 GNOI 快速检索算法的有效性。综上所述,知识图谱的引入,给遥感数据产品的管理提供了一种全新的思路。

## 参 考 文 献

- [1] Liu Jiyuan, Kuang Wenhui, Zhang Zengxiang, et al. Spatiotemporal Characteristics, Patterns and Causes of Land Use Changes in China Since the Late 1980s[J]. *Acta Geographica Sinica*, 2014, 69(1): 3-14 (刘纪远, 匡文慧, 张增祥, 等. 20 世纪 80 年代末以来中国土地利用变化的基本特征与空间格局[J]. *地理学报*, 2014, 69(1): 3-14)
- [2] Liu Fang, Yan Huimin, Liu Jiyuan, et al. Spatial Pattern of Land Use Intensity in China in 2000[J]. *Acta Geographica Sinica*, 2016, 71(7): 1130-1143 (刘芳, 闫慧敏, 刘纪远, 等. 21 世纪初中国土地利用强度的空间分布格局[J]. *地理学报*, 2016, 71(7): 1130-1143)
- [3] Li Deren, Bin Hongchao, Shao Zhenfeng. Design and Implementation of Land & Resources Grid Management and Service System [J]. *Geomatics and Information Science of Wuhan University*, 2008, 33(1): 1-6 (李德仁, 宾洪超, 邵振峰. 国土资源网格化管理与服务系统的设计与实现[J]. *武汉大学学报·信息科学版*, 2008, 33(1): 1-6)
- [4] Sahr K, White D, Kimerling A J. Geodesic Discrete Global Grid Systems [J]. *Cartography and Geographic Information Science*, 2003, 30(2): 121-134
- [5] Jun C, Ban Y F, Li S N. Open Access to Earth Land-Cover Map[J]. *Nature*, 2014, 514(7523): 434
- [6] Zhang X, Liu L Y, Chen X D, et al. GLC\_FCS30: Global Land-Cover Product with Fine Classification System at 30 m Using Time-Series Landsat Imagery [J]. *Earth System Science Data*, 2021, 13(6): 2753-2776
- [7] Zhang X, Liu L Y, Wu C S, et al. Development of a Global 30 m Impervious Surface Map Using Multi-source and Multitemporal Remote Sensing Datasets with the Google Earth Engine Platform [J]. *Earth System Science Data*, 2020, 12(3): 1625-1648
- [8] Liu L Y, Zhang X, Gao Y, et al. Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects [J]. *Journal of Remote Sensing*, 2021, 2021: 1-38
- [9] Gong P, Liu H, Zhang M N, et al. Stable Classification with Limited Sample: Transferring a 30 m Resolution Sample Set Collected in 2015 to Mapping 10 m Resolution Global Land Cover in 2017 [J]. *Science Bulletin*, 2019, 64(6): 370-373
- [10] Wang Haofen, Ding Jun, Hu Fanghui, et al. Survey on Large Scale Enterprise-Level Knowledge Graph Practices[J]. *Computer Engineering*, 2020, 46(7): 1-13 (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)
- [11] Zhang Bo, Zhu Jun, Su Hang. Toward the Third Generation of Artificial Intelligence [J]. *Scientia Sinica (Informationis)*, 2020, 50(9): 1281-1302 (张钲, 朱军, 苏航. 迈向第三代人工智能[J]. *中国科学: 信息科学*, 2020, 50(9): 1281-1302)
- [12] Jiang Bingchuan, Wan Gang, Xu Jian, et al. Geographic Knowledge Graph Building Extracted from Multi-sourced Heterogeneous Data [J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(8): 1051-1061 (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): 1051-1061)
- [13] Liu Liu. Construction of Knowledge Graph of Jilin

- Based on Geographic Ontology [D]. Beijing: Beijing Jiaotong University, 2017 (刘鑾. 基于地理本体的吉林地域知识图谱的构建[D]. 北京: 北京交通大学, 2017)
- [14] Lü Gang. Research and Application of Personalized Recommendation System Based on Knowledge Graph [D]. Harbin: Harbin Normal University, 2020 (吕刚. 基于知识图谱的个性化推荐系统研究与应用[D]. 哈尔滨: 哈尔滨师范大学, 2020)
- [15] Mao Wenshan. Research on Water Pollution Event Tracing Method Based on Knowledge Graph Relational Reasoning [D]. Lanzhou: Lanzhou Jiatong University, 2020 (毛文山. 基于知识图谱关系推理的水污染事件追踪溯源方法研究[D]. 兰州: 兰州交通大学, 2020)
- [16] Yu Helong, Shen Jinmeng, Bi Chunguang, et al. Intelligent Diagnostic System for Rice Diseases and Pests Based on Knowledge Graph [J]. *Journal of South China Agricultural University*, 2021, 42(5): 105-116 (于合龙, 沈金梦, 毕春光, 等. 基于知识图谱的水稻病虫害智能诊断系统[J]. 华南农业大学学报, 2021, 42(5): 105-116)
- [17] Zhou Chenghu, Wang Hua, Wang Chengshan, et al. Geoscience Knowledge Graph in the Big Data Era [J]. *Scientia Sinica (Terrae)*, 2021, 51(7): 1070-1079 (周成虎, 王华, 王成善, 等. 大数据时代的地质知识图谱研究[J]. 中国科学: 地球科学, 2021, 51(7): 1070-1079)
- [18] Shi X, Nie S, Ju W, et al. Application and Impacts of the GlobeLand30 Land Cover Dataset on the Beijing Climate Center Climate Model [J]. *IOP Conference Series: Earth and Environmental Science*, 2016, 34: 012032
- [19] Shi X L, Nie S P, Ju W M, et al. Climate Effects of the GlobeLand30 Land Cover Dataset on the Beijing Climate Center Climate Model Simulations [J]. *Science China Earth Sciences*, 2016, 59(9): 1754-1764
- [20] Cao X, Chen J, Chen L J, et al. Preliminary Analysis of Spatiotemporal Pattern of Global Land Surface Water [J]. *Science China Earth Sciences*, 2014, 57(10): 2330-2339
- [21] Yang Y, Kong C, Liu C, et al. Research on Hierarchy Structure Generation Method of Ontology Knowledge Pan-Concept in Agriculture [J]. *Advances in Robotics & Automation*, 2017, 6(3): 1-6
- [22] Yu Fan. Methodology and Empirical Research on Domain Ontology [M]. Wuhan: Wuhan University Press, 2015 (余凡. 领域本体构建方法及实证研究: 以测绘学领域为例[M]. 武汉: 武汉大学出版社, 2015)
- [23] He Haishan, Zhao Yuhao, Wu Jiansheng. Simulation of Urban Landscape Pattern Under the Influence of Low Carbon: A Case Study of Shenzhen [J]. *Acta Ecologica Sinica*, 2021, 41(21): 8352-8363 (何海珊, 赵宇豪, 吴健生. 低碳导向下土地覆被演变模拟: 以深圳市为例[J]. 生态学报, 2021, 41(21): 8352-8363)
- [24] Pan Ruyi, Shen Lijuan, Xie Deti, et al. Land Cover Transformation of the Old City District and Landscape Pattern Change—A Case Study of Beibei District of Chongqing [J]. *Economic Geography*, 2015, 35(1): 155-161 (潘如玉, 申丽娟, 谢德体, 等. 重庆旧城改造区土地覆被与景观格局变化: 以北碚区为例[J]. 经济地理, 2015, 35(1): 155-161)
- [25] Jiang Dongdong, Ma Weibo, Zou Fengli, et al. Spatiotemporal Changes of Ecosystem Service Value in Dalou Mountain Area at Township Scale [J]. *Research of Environmental Sciences*, 2020, 33(12): 2713-2723 (姜栋栋, 马伟波, 邹凤丽, 等. 乡镇尺度大娄山区生态系统服务价值时空变化研究[J]. 环境科学研究, 2020, 33(12): 2713-2723)
- [26] Zhang Yang, Liu Yanfang, Gu Jianping, et al. Land Use/Land Cover Change and Its Environmental Effects in Wuhan City [J]. *Scientia Geographica Sinica*, 2011, 31(10): 1280-1285 (张杨, 刘艳芳, 顾渐萍, 等. 武汉市土地利用覆被变化与生态环境效应研究[J]. 地理科学, 2011, 31(10): 1280-1285)
- [27] Aydin S, Aydin M N. Ontology-Based Data Acquisition Model Development for Agricultural Open Data Platforms and Implementation of OWL2MVC Tool [J]. *Computers and Electronics in Agriculture*, 2020, 175: 105589
- [28] Xu Zenglin, Sheng Yongpan, He Lirong, et al. Review on Knowledge Graph Techniques [J]. *Journal of University of Electronic Science and Technology of China*, 2016, 45(4): 589-606 (徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述[J]. 电子科技大学学报, 2016, 45(4): 589-606)
- [29] Shao Mingchao, Song Hongli, Shang Ming, et al. Multisource Land-Cover Data Category Accuracy Evaluation and the Difference of Landscape Patterns: A Case Study of Beijing-Tianjin-Hebei Region [J]. *Chinese Journal of Ecology*, 2020, 39(11): 3881-3889 (邵明超, 宋宏利, 尚明, 等. 多源土地覆被数据类别一致性及景观格局差异性: 以京津冀区域为例[J]. 生态学杂志, 2020, 39(11): 3881-3889)
- [30] Ma Liqun, Li Li, Liu Junjie, et al. Land Cover Classification of the Loess Plateau Based on GLAS

- Full Waveform Data[J]. *Science of Surveying and Mapping*, 2021, 46(3): 80-86 (马利群, 李理, 刘俊杰, 等. 大光斑波形数据在土地覆盖分类的适用性分析[J]. 测绘科学, 2021, 46(3): 80-86)
- [31] Meng Wen, Tong Xiaohua, Xie Huan, et al. Accuracy Assessment for Regional Land Cover Remote Sensing Mapping Product Based on Spatial Sampling: A Case Study of Shaanxi Province, China[J]. *Journal of Geo-Information Science*, 2015, 17(6): 742-749 (孟雯, 童小华, 谢欢, 等. 基于空间抽样的区域地表覆盖遥感制图产品精度评估: 以中国陕西省为例[J]. 地球信息科学学报, 2015, 17(6): 742-749)
- [32] Huang Yabo, Liao Shunbao. Regional Accuracy Assessments of the First Global Land Cover Dataset at 30-Meter Resolution: A Case Study of Henan Province[J]. *Geographical Research*, 2016, 35(8): 1433-1446 (黄亚博, 廖顺宝. 首套全球 30m 分辨率土地覆被产品区域尺度精度评价: 以河南省为例[J]. 地理研究, 2016, 35(8): 1433-1446)
- [33] Song Hongli, Zhang Xiaonan. Precision Analysis and Validation of Multi-sources Landcover Products Derived from Remote Sensing in China[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2012, 28(22): 207-214 (宋宏利, 张晓楠. 中国区域多源土地覆被遥感产品精度分析与验证[J]. 农业工程学报, 2012, 28(22): 207-214)

## Research on Land and Resources Management and Retrieval Using Knowledge Graph

ZHANG Yongjun<sup>1</sup> CHENG Xin<sup>1</sup> LI Yansheng<sup>1</sup> WANG Fei<sup>1</sup> LIU Xiaojian<sup>1</sup> WU Wenpin<sup>1</sup>

<sup>1</sup> School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China

**Abstract: Objectives:** Aiming at the problem of difficult effective management and rapid application between different data products of land and resources, the study uses the graph database to store the public land cover datasets, including GlobaLand30, FROM-GLC10\_2017, GLC\_FCS30\_2020, etc., on the semantic level to establish a knowledge graph of land resources. It provides a new processing framework for the management, rapid application, and data quality assessment of land and resources data. **Methods:** A new application framework for land cover data product management, knowledge extraction, and data acquisition and update based on administrative divisions is proposed. Anomaly data retrieval algorithms based on graphs are used to explore the consistency of different products, and a knowledge-based fast retrieval algorithm for graph nodes of interest (GNOI) in the graph. **Results:** Through the introduction of the knowledge graph, a dynamically updateable nationwide land resource knowledge graph containing 447 817 nodes and 447 816 relationships has been formed, and it is found that the data accuracy of 92 units may have large errors in the 2 875 administrative units covering the whole country. **Conclusions:** The research has greatly improved the utilization rate of multi-source land cover data products, shortened the time of data preprocessing for researchers, and provided new ideas for the knowledge management and application of land resources.

**Key words:** land and resources management; knowledge graph; land cover data product; abnormal data retrieval; graph nodes of interest (GNOI)

**First author:** ZHANG Yongjun, PhD, professor, majors in photogrammetric and remote sensing integrated processing of aerial and space data. E-mail: zhangyj@whu.edu.cn

**Corresponding author:** LI Yansheng, PhD, associate professor. E-mail: yansheng.li@whu.edu.cn

**Foundation support:** The National Natural Science Foundation of China(42030102, 41971284); the Foundation for Innovative Research Groups of the Natural Science Foundation of Hubei Province(2020CFA003).

**引文格式:** ZHANG Yongjun, CHENG Xin, LI Yansheng, et al. Research on Land and Resources Management and Retrieval Using Knowledge Graph[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(8): 1165-1175. DOI:10.13203/j.whugis20210714 (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): 1165-1175. DOI:10.13203/j.whugis20210714)