

DOI: 10.16108/j.issn1006-7493.2021107

引用格式: 冉一早, 董少春, 王汝成, 胡欢, 方思源, 常丰年, 孔嘉. 2023. 铌钽矿床知识图谱的构建及应用实践[J]. 高校地质学报, 29(3): 359–371

铌钽矿床知识图谱的构建及应用实践

冉一早¹, 董少春^{1*}, 王汝成^{1,2}, 胡欢^{1,2}, 方思源¹, 常丰年¹, 孔嘉³

1. 南京大学地球科学与工程学院, 南京 210023;

2. 内生金属矿床成矿机制研究国家重点实验室, 南京大学地球科学与工程学院, 南京 210023;

3. 南京大学信息管理学院, 南京 210023

摘要: 关键金属作为未来新兴产业的核心资源保障, 近年来逐渐成为国内外学者研究的热点, 其研究存在难示踪、难辨识、难分离等瓶颈。知识图谱作为大数据和人工智能的重要基础构建, 能够为这些科学问题的解决提供新思路。与关键金属矿床相关的矿物学、年代学、地球化学等研究成果大多以文献的形式发表在国内外期刊杂志上, 这类文献通常以自然语言的形式进行描述, 难于为机器所理解和直接利用。文章以铌钽矿床为例, 对铌钽矿床重要的概念和描述性属性进行了整理, 构建了铌钽矿床知识图谱的本体层, 并从大量文献资料中提取相关信息, 建立了铌钽矿床知识图谱的数据层, 初步完成了Nb-Ta矿床知识图谱的建设。在此基础上, 从大数据分析的角度对中国铌钽矿床空间—时间—构造背景—成因类型多维分布特征以及矿石矿物共生组合特征与铌钽矿床多维分布之间的关联进行了分析, 为进一步揭示中国铌钽矿床时空演化规律提供了新的视角。

关键词: Nb-Ta矿床; 知识图谱; 大数据; 时空分布

中图分类号: P628+4 **文献标识码:** A **文章编号:** 1006-7493 (2023) 03-359-13

Construction of Nb-Ta Deposit Knowledge Graph and Its Application

RAN Yizao¹, DONG Shaochun^{1*}, WANG Rucheng^{1,2}, HU Huan^{1,2}, FANG Siyuan¹,
CHANG Fengnian¹, KONG Jia³

1. School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China;

2. State key laboratory for mineral deposits research, School of Earth Sciences and Engineering, Nanjing University, Nanjing 210023, China;

3. School of Information Management, Nanjing University, Nanjing 210023, China

Abstract: As an important resource for emerging industries, critical metals have gradually gain attention both domestically and overseas in recent years. However there are still some bottlenecks existing in the critical metal research due to its low abundance, difficulty to trace, identify and separate. As an important infrastructure of big data and artificial intelligence, knowledge graph provides new insights to tackle these issues. Most of the research achievements related to critical metal deposits are published in academic journals, which are written in natural language and is difficult to be understood and directly used by machines. In this paper, the ontology layer of Nb-Ta deposit knowledge graph was constructed and data extracted from literatures consists of the fact layer of the Nb-Ta deposit knowledge graph. Based on the Nb-Ta deposit knowledge graph, the features and relationships among the metallogenic periods, types of Nb-Ta deposits and the co-occurrence of niobium and tantalum related minerals were

收稿日期: 2021-09-27; 修回日期: 2021-12-13

基金项目: 国家自然科学基金面上项目 (42072320; 42230809) 资助

作者简介: 冉一早, 女, 1997年生, 硕士研究生, 矿物学、岩石学、矿床学专业; E-mail: MG1929017@smail.nju.edu.cn

*通讯作者: 董少春, 女, 1976年生, 博士, 副教授, 主要从事遥感与地理信息系统应用研究以及地球科学大数据技术和应用;

E-mail: dsc@nju.edu.cn

analyzed. It will help to further reveal the spatio-temporal distribution and evolution characteristics of China Nb-Ta deposits.

Key words: Nb-Ta deposit; knowledge graph; big data; spatio-temporal distribution

Corresponding author: DONG Shaochun, Associate Professor; E-mail: dsc@nju.edu.cn

知识图谱这一名词由谷歌于2012年正式提出 (Singhal et al., 2012)。它的概念实际上却是在20世纪七十年代开始的专家系统, 20世纪九十年代兴起的语义网络以及本体的基础上发展而来的 (Wang et al., 2014; Nickel et al., 2015; Gutierrez and Sequeda, 2021)。知识图谱通过图的方式对语义知识进行形式化的描述, 形成可以被人和机器所理解的大规模知识库, 是实现语义翻译、数据融合和知识建模的关键技术之一 (曹倩等, 2015; 徐增林等, 2016; 黄恒琪等, 2019)。

作为大数据和人工智能技术的重要基础设施, 知识图谱在跨学科数据交叉融合、数据与知识的整合与共享, 以及数据驱动下的知识推理与发现等方面具有重要的作用 (董少春等, 2019)。近年来, 地球科学知识图谱的建设受到了越来越多的重视, 也取得了长足的发展。例如 Ma 等 (2012) 建立了基于 SKOS 的多语言版本地质年代本体, 开发了基于地图的地质信息在线检索功能和 GIS 本体的可视化, 提供了机器可理解的地质历史时间的语义翻译。侯志伟等 (2018) 利用模块化的方法建立了以中国地质年代及地层为主的地质年代本体, 并基于地质年代本体设计了一个应用原型系统, 实现地质学数据的语义检索, 有效解决地学知识中的语义异构等问题。Wang 等 (2018) 利用分词规则对地质类文献进行解析, 建立了地质术语语料库, 分析了词语之间的语义关系, 并将词语及其相互关系可视化为知识图谱的节点和边, 直观清晰地表达了文献中的关键信息, 实现了从非结构化文档中提取关键信息自动构建知识图谱的目的。Osipova 等 (2020) 等基于煤矿和天然气的地质、水文、物理化学性质及采矿信息等知识构建了煤矿和天然气知识图谱, 并在此基础上利用数据挖掘的方法进一步分析煤层和天然气在突然爆发过程中的地球化学动力学和气体动力学的过程, 有效预测和防止由煤层和天然气突然爆发引起的地质灾害。Enkhsaikhan 等 (2021) 等利用自然语言处理技术和深度学习方法从高质量

矿床勘探报告中自动提取和正确标注了包括地理位置、地质年代、地层单元、岩石类型、矿石、矿床类型及所含矿物等地质术语, 构建了铁矿和金矿知识图谱, 建立了快速、准确的将文本数据转换为结构化数据的方法, 为地质知识挖掘领域提供了新的思路。周永章等 (2021) 基于斑岩铜矿床概念模型, 提取了钦杭成矿带斑岩铜矿床文本数据中的实体、属性及关系, 完成了斑岩铜矿床知识图谱的构建, 为进一步构建地球科学系统知识图谱提供了案例。

关键金属是近年来国际上新提出的资源概念, 主要指现今社会必需但安全供应存在风险的一类矿产的总称, 包括稀有、稀土、稀散和稀贵金属。由于关键金属元素地壳丰度低, 常以吸附、类质同象和极细小矿物形式存在于矿床中, 这些特征客观上决定了关键金属元素的研究存在难示踪、难辨识、难分离等瓶颈 (陈骏等, 2019; 毛景文等, 2019; 侯增谦等, 2020)。其中, 铌钽元素作为关键稀有金属元素, 有着顺磁性、极强延展性、低温下超导电性等特殊性质, 是新能源、超导技术、航空航天和国防军工等多个尖端领域核心材料之一, 因此对于其富集过程、地球化学行为、矿床形成、时空分布规律、资源勘探以及提高现存资源可利用等关键科学问题的研究, 具有重要的科学和现实意义 (胡瑞忠等, 2020; 王汝成等, 2020; 黎洁等, 2021)。

本文综合国内铌钽矿床文献, 构建了铌钽矿床知识图谱, 初步开展了基于知识图谱的铌钽矿床大数据共享平台建设和应用研究, 从大数据的角度对我国铌钽矿床空间-时间-构造背景-成因类型多维分布特征以及矿石矿物共生组合特征与铌钽矿床多维分布之间的关联进行了分析, 为进一步揭示铌钽矿床时空演化规律提供新的视角。

1 知识图谱逻辑架构

知识图谱是指以图的形式 (包括图数据库, 有向边标记图, 属性图等) 记录和表达现实世界的知

识(刘峤等,2016)。知识图谱将领域知识中的概念、对象和过程(或事件)等抽象成不同的实体,以三元组的形式对实体以及实体之间的相互关系进行形式化的描述,形成具有一定结构的语义网络,构建可以被人和机器理解的大规模知识库,为数据集集成、共享和重用提供服务(石顺中等,2020)。在知识图谱的三元组中,节点表示实体,边表示实体间的关系(Nickel et al., 2015)。三元组可以表示为<概念或实体 1, 关系, 概念或实体 2>和<概念或实体, 属性, 属性值>等不同形式。例如在岩石学知识图谱中,花岗岩和岩浆岩可以被定义为两个概念,它们之间的关系可以以三元组的形式表示为<花岗岩-属于-岩浆岩>。石英是另一个概念,化学式是石英的一个属性。石英的化学式写作 SiO_2 。这一关系用三元组表示即为<石英-化学式- SiO_2 >。

知识图谱从逻辑架构上可以分为两个层次(如图1所示):本体层和事实层(徐增林等,2016;石顺中等,2020;齐浩等,2020)。本体(Ontology)这一概念起源于哲学,表示对某种客观事物的系统描述(李善平等,2004)。Gruber等(1993)提出了信息科学领域中广泛认可的本体的正式定义,即本体是一个概念模型的规范化、标准化的说明。本体具有四个方面的特征(邓志鸿等,2002),即:(1)共享,本体中代表的知识是某个领域中共同认可的;(2)概念,本体是以概念的形式描述事物的;(3)明确,本体中的概念或属性都有明确定义;(4)形式,本体这一形式能够被计算机所识别。

本体层的构建也称为知识体系的构建或者知识建模(赵军,2018),即构建某一领域知识的通

用概念模型,对该领域的概念以及概念与概念之间的关系进行描述。本体层三元组的表达形式为<概念-关系-概念>或<概念-属性-属性值>。其中概念与概念之间的关系可细化为包含关系,从属关系,对等关系,相反关系、部分与整体的关系等。例如,<硅酸盐矿物-包括-岛状硅酸盐矿物>、<铌钽矿床-属于-金属矿床>、<sphene 等价于 titanite>。

本体层可以通过专家视角,采用自顶向下的方式人工构建,以保证知识描述的准确、清晰和完整性;也可通过文本信息抽取技术由机器自动从文献或数据库中抽取信息构建,或采用人工与机器构建相结合的方式实现。

事实层由从数据源中抽取的大量事实组成,以事实为单位对知识进行表达和存储。事实层三元组的表达形式为<实体-关系-实体>或<实体-属性-属性值>。其中实体指具体的实例,如实际的人名、地点、物品;属性用以说明属性值和实体间的关系。在事实层中,以节点表示实体,以边来表示属性和实体间的关系。事实层一般采用自底向上的方式建设,即从大量文献或者专题数据库中抽取实例及其属性的描述,形成“实体-属性-属性值”或者“实体-关系-实体”三元组,例如<癞子岭矿床-成矿年龄-155 Ma>和<癞子岭矿床-位于-湖南省>等。

本体层和事实层的构建不是孤立的,它们之间需要建立映射关系,并相互补充。事实层和本体层可以通过<实体-关系-概念>的三元组表示相互之间的映射关系。例如利用<“癞子岭矿床”-

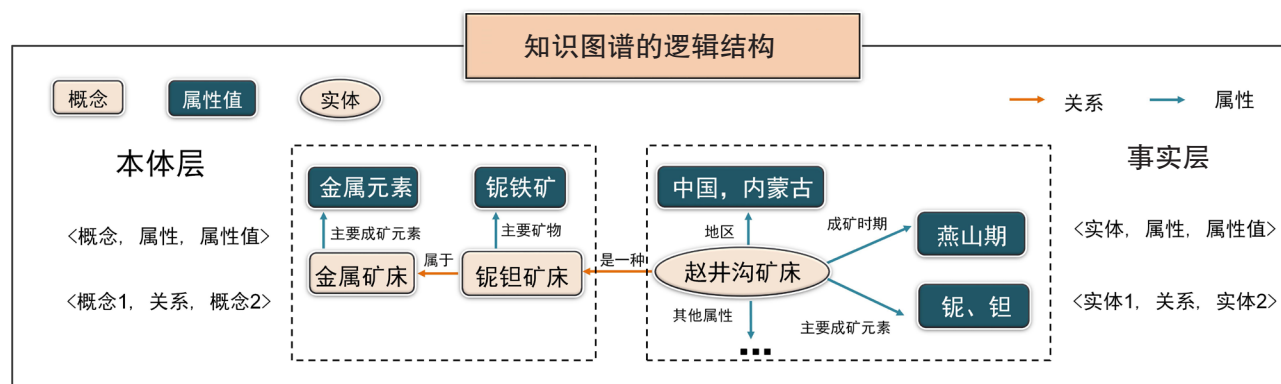


图1 矿床知识图谱的逻辑结构示意图

Fig.1 Framework of the Mineral Deposit Knowledge Graph

是一种 - “铌钽矿床” > 这一三元组即可将“癩子岭矿床”这一真实的矿床实体映射到“铌钽矿床”这一概念上。本体层中的概念可以管理和约束事实层中各实体及其属性之间的关系。事实层的数据需要映射到本体层中的相关概念和术语上, 才能使得数据之间具有明确、清晰的语义关系, 并且为机器所理解; 同时, 事实层描述的数据也可为本体层补充新的概念和实体, 从而完善本体层的构建。本体层和事实层相互连接才能在数据整合、语义翻译、知识推理和知识发现中充分发挥知识图谱的作用。

2 矿床知识图谱建设的国内外现状

知识图谱在矿床学研究中尚处于起步阶段, 还没有形成成熟的铌钽矿床知识图谱。由于知识图谱从结构上分为本体层和数据层, 因此我们分别调研了与矿床有关的本体层的建设现状, 以及与数据层有关的矿床专题数据库的建设现状, 为铌钽矿床知识图谱的建设提供借鉴。

2.1 与矿床有关的本体层建设现状

GEON (Geoscience Network) 项目由美国地质调查局及多家高校等科研机构于 2002 年启动, 目标是将异构的地球科学数据汇集在一起, 以促进信息集成和知识发现的新方式。该项目建立了一个集成本体, 包括地质年代、岩石分类、化学成分、岩石结构及有关的子本体。在此基础上, 消除了不同地质图间的内容异构问题 (Ludascher et al., 2003; Baru et al., 2009), 实现了对多源异构地质图的综合访问和查询功能。

GeoSciML (<https://www.seegrid.csiro.au/twiki/bin/view/CG/Model/GeoSciML>) 是一种描述地学数据的本体模型, 为地球科学专题数据和相关空间数据的应用提供了标准化的框架 (Sen and Duffy, 2005), GeoSciML 主要用于地质图数据、地质年代、岩石结构、化石、地球化学分析数据等地学信息的数据交换。

SWEET 是 NASA 于 2009 年创建的一个描述地球科学数据和知识的顶层领域本体库。该本体库的规模较大, 涉及面广, 对地球圈层、岩石、化学物质、生物、地质历史时间、空间位置、地质自然现象 (火山、地震等) 等多个地球领域的术语进行了规范化的描述 (马胜男等, 2010), 覆盖地球科学

领域各类概念约 4600 个、实例约 2100 个、属性及关系约 360 个, 为地球科学领域本体的建设及应用提供了借鉴意义。

“EarthCube” (<https://www.earthcube.org/>) 是美国自然基金会地球科学局和计算机信息科学与工程局自 2012 年开始共同资助的项目, 旨在以开放、透明、包容的方式共享地球科学的数据和知识。在随后几年时间里, 陆续出现了与 EarthCube 相关的项目。2014 年, 美国加利福尼亚大学联合多所大学创建了 EarthCube-GeoLink (<http://www.geolink.org/>)。该项目通过关联数据实现了海洋地质、生物地球化学和古气候科学研究等地学数据的直接联系, 并进一步创建了 4500 万个 RDF 三元组以及本体数据可视化工具, 极大地方便了用户在地球科学领域的查询和推理。

GeoNames Ontology (<http://sws.geonames.org/3020251/about.rdf>) 是对覆盖全球地名的免费地名词典 GeoNames (<http://www.geonames.org/>) 中地名术语的本体描述, 具有本体层的功能。整个本体库以 RDF 的形式描述了全球大约 1200 万个地理实体的国家代码、经纬度等属性信息。之后, 意大利特伦托大学创建了 GeoWordNet 本体库, 集成了 WordNet 及 GeoNames 两个地名数据库中的实体, 并增加了地理实体的同义词集合 (陆锋等, 2017)。其中, GeoWordNet 的实体部分均来源于 GeoNames 的地名, 包含具体的经纬度坐标; 概念部分则分别来自 WordNet 和 GeoNames, 共计约 300 个概念, 360 万个概念及 180 个概念间关系, 为地理知识的获取与知识服务提供了基础。

德国莱比锡大学利用 OpenstreetMap 空间数据集创建了一个大型空间知识库 LinkedGeoData (<http://linkedgeodata.org/>)。该知识库数据量庞大、覆盖面广, 并且具有良好的数据开放性, 数据增长量也非常迅速。LinkedGeoData 建立地理实体节点间的相互关联并以 RDF 的形式存储数据, 共计约 30 亿个节点, 3 亿条边及产生 200 亿个 RDF 三元组, 在地理数据深层集成方面展示了巨大的潜力和作用。

上述本体库的建设可以服务于地球科学多个分支领域, 属于地球科学领域的顶层本体和通用本体的范畴。而中国学者则在矿床相关领域进行了一些

探索性工作。吴永亮等(2017)依靠机器学习方法开发了地质大数据发现与挖掘系统,实现了对地质找矿专题数据的自动采集和深层次的数据挖掘,完成了多数据源找矿模型库和知识库的构建。其次对模型知识库中的模型进行计算分类,建立了全面可靠的找矿预测模型,并将该模型应用到焦家金成矿带,验证了该模型的科学性和有效性。姚健鹏等(2017)等依据与铜矿相关的专著及文献,对我国各类型的铜矿床进行了成因分类和概念的描述,建立了铜矿床本体模型,添加了铜矿床实例进一步完善了铜矿床本体,并以可视化的方式构建了结构清晰、知识丰富的铜矿床本体,可用于铜矿床资料中的精准检索,为铜矿床信息的整合和找矿工作的进一步开展起到支撑作用。常立恒等(2018)在深入分析结构化与非结构化矿床数据特征的基础上,以钨矿为例建立了矿产知识库,在此基础上利用模糊匹配算法实现了矿床类型的分类,验证了该算法在区分预测矿床类型上的有效性,为大数据的整合与计算提供了服务,也对钨矿成矿规律及找矿工作等研究起到了一定的指导意义。叶帅(2019)依据煤矿领域的知识结构整理了煤矿实体及相互间的关系,构建了煤矿知识图谱的框架,并利用网络结构模型提取了煤矿实体及关系,随后基于图数据库构建了煤矿知识图谱,实现了基于知识图谱的智能检索和可视化展示,并在知识管理和知识查询等应用上做了积极的探索。

2.2 与矿床有关的数据层建设现状

与矿床有关的数据层的建设主要来源于专题数据库和文献资源库。专题数据库收集整理了以矿床基本信息为核心的实地勘探数据,如矿床类型、矿体规模、成矿时代、矿区历史以及各类解释型数据。这类信息通常保存在数据库中,虽然也会存在一定的格式异构和语义异构问题,但经过数据清洗和整合,可以转化为直接被机器所利用的结构化数据,具有较高的共享和复用程度。期刊文献中也包含矿床的基本信息,同时还有大量实验室分析测试数据,如地球化学数据等。这一类数据通常以自然语言描述,以文件形式存储。虽然以数据形式发表,但在PDF文件中无法被直接识别读取,属于非结构化的数据,无法被机器直接理解,需要经过信息提取转换成结构化的数据,才能实现数据的共享和复用。

2.2.1 矿床专题数据库建设情况

矿床数据库是矿床工作者基于计算机技术将非结构化的矿床数据进行数字化,转变成结构化的数据,将结构化的数据进一步整合,最终集合而成的专业数据库。矿床数据库可以提供集成和共享的矿床资源数据。

矿床资源是一个国家重要的自然资源,各个国家都非常重视矿床数据库的建设。美国地调局(USGS)自1973年就开发并建设了全球性的矿床资源数据系统GSMRDS,收集整理了全世界171个国家共304,632条矿床和矿点的地质背景、储量等信息(Selner et al., 1992a, b, c)。自2011年起,USGS停止了对GSMRDS的系统更新,并着力创建一个新的数据库USMIN,主要侧重于整合美国本土重要的矿床和矿产区的权威信息源。该数据库尤其关注美国列出的多种关键金属,数据覆盖范围涵盖美国31个州,624个矿区,639条记录,数据可以以Geodatabase、Shapefiles和csv等格式被用户下载。

澳大利亚矿山地图集(<https://portal.ga.gov.au/persona/minesatlas>)是一个矿床资源地图交互式查询系统,提供澳大利亚矿床地图集资料的检索和下载服务。该地图集记录了澳大利亚421个矿区的开采现状、经纬度、矿石类型、基本描述等矿床信息。地图集的在线显示功能标出了澳大利亚正在开采的矿床位置,底图为澳大利亚地质图。但这些地图集只能以PDF或JPG的形式被用户下载。

英国地质调查局(BGS)着力建设了BRITPITS数据库(<https://www2.bgs.ac.uk/mineralsuk/mines/dmq.html>)。该数据库收集了英国在产和已停产矿床的地理位置、开发商、地质背景、生产的主要矿石及其用途等信息。目前数据库中有超过23万条记录。数据覆盖范围为英国,北爱尔兰,曼岛和海峡群岛,下载格式为GIS点数据或Excel形式。

中国地质调查局也开发了地质云矿产资源共享平台([http://geocloud.cgs.gov.cn/#/portal/geologicalDatabase/Areal Geology?type=dzsjk&code=矿产资源数据库](http://geocloud.cgs.gov.cn/#/portal/geologicalDatabase/Areal%20Geology?type=dzsjk&code=矿产资源数据库))。这是一个地图交互式查询系统,可以在线显示中国煤层气、稀土矿床、镍铜铂岩浆硫化物矿床、能源矿产、黑金属、成矿带、有色金属、贵金属、稀有金属、水汽矿产等各类矿床的全

国分布图,提供各类矿床的浏览和检索、矿区面积及长度测量等服务。

这些数据库的建设包含了大量与矿床有关的数据,有助于矿床资源的潜力分析和预测。但是有些数据库已停止更新,同时大部分数据库都不同程度地存在矿床资源数据不全面、重要信息缺失(如关键金属的矿床位置未明确标注、矿床的地球化学数据严重缺失等)以及数据不完全共享等问题,限制了矿床数据库的发展和利用。

2.2.2 文献中矿产数据现状

每年收录在各大数字出版商中的期刊论文是矿床数据的重要来源。这类数据规模巨大,而且增长迅速,仅中国知网(截止2020年)就收录了超过30万篇与矿床有关的中文文献。在英文文献库Elsevier ScienceDirect中检索关键字“deposit”,仅1998年至2020年就有超过150万篇相关文献。这些文献中的矿床基本描述(如研究地区、地理坐标、构造背景、矿床名称、矿床类型、矿石矿物、脉石矿物、成矿时代、围岩蚀变、矿体规模、开采状态)和相关数据(如全岩、微区主量、微量元素含量等)以自然语言的形式(包括文字、图片和表格等不同形式)出现,属于非结构化的数据,不易转化为结构化信息被机器直接利用。这部分信息被称为科学数据中的“暗数据”和“长尾数据”(Heidorn et al., 2008),是矿床大数据的重要组成,也是知识图谱事实层数据的重要来源。

3 铌钽矿床知识图谱的设计与构建

铌钽矿床知识图谱的构建主要包括本体层构建(知识体系构建)、事实层构建(知识获取)以及知识融合等部分。下文重点阐述铌钽(Nb-Ta)矿床知识图谱中本体层和事实层的设计与构建。

3.1 本体层设计

本体层是经过高度概括和提炼的领域知识的描述,是一类具有共性的概念或对象的描述。矿床知识图谱的本体层主要描述矿床基本概念及其相互关系,是对矿床基本信息的知识建模,以三元组的形式进行表达。例如我们将所有矿床抽象为概念A或类A,所有金属矿床抽象为概念B,所有铌钽矿床抽象为概念C,那么概念A、B和C之间就存在包含关系,用三元组表示即为“概念B-包含于-

概念A”,“概念C-包含于-概念B”,而根据知识推理规则可以推断出“概念C-包含于-概念A”。包含关系还具有互逆性,即根据“概念B-包含于-概念A”可推断出“概念A-包含-概念B”。“矿床”概念具有成矿时代、成矿区域、主要成矿元素、构造背景等描述性属性。这些属性和矿床概念之间也可以用三元组形式表示,即“矿床(概念)-具有(关系)-成矿时代(属性)”。由于概念之间具有上下位的包含关系,其属性可以自上而下继承,但是居于下位的概念创建的属性却不能为上位的概念所继承,即继承只能自上而下,不能自下而上。因此当我们为“金属矿床”概念B创建了一系列属性后,那么作为“金属矿床”概念B的下位“铌钽矿床”概念C,也具有这一系列属性。但“铌钽矿床”概念C可能还有一些特殊的属性,例如,“铌钽矿床”概念C富含铌钽副矿物,作为其父类的金属矿床B却不具有这些属性。

按照这一思路,本文将“铌钽矿床”和“金属矿床”创建为两个类(class),它们之间具有包含关系,即“铌钽矿床”是“金属矿床”的子类,记作<“铌钽矿床”-属于(is a subclass of)-“金属矿床”>。这一关系可以根据互逆性表示成<“金属矿床”-包含(contains/is the super class of)-“铌钽矿床”>。类“铌钽矿床”具有成矿时代、地理位置、构造背景、储量、矿石矿物、指示性矿物、成矿母岩等描述性属性。类及其属性构成了铌钽矿床知识图谱本体层(图2a)的概念模型。本体层需要利用本体描述性语言进行形式化描述才能更好地为机器所理解。本文采用万维网联盟W3C(World Wide Web Consortium)推荐的本体描述语言OWL(Ontology Web Language)对本体层进行了描述,部分代码如下图2b所示。

3.2 事实层设计

事实层主要根据本体层的模式设计,从数据源中抽取相应的信息,作为实体对象的实例,即每个具体矿床的详细描述信息。数据源主要包括矿床专题数据库和文献资料。由于矿床专题数据库存在开放共享程度不统一的问题,难以获得完整的铌钽矿床基本信息,因此文献资料成为构建事实层的主要数据来源。这部分资料丰富、可信度高,但属于自然语言文本数据,因此需要利用自然语言处理技术

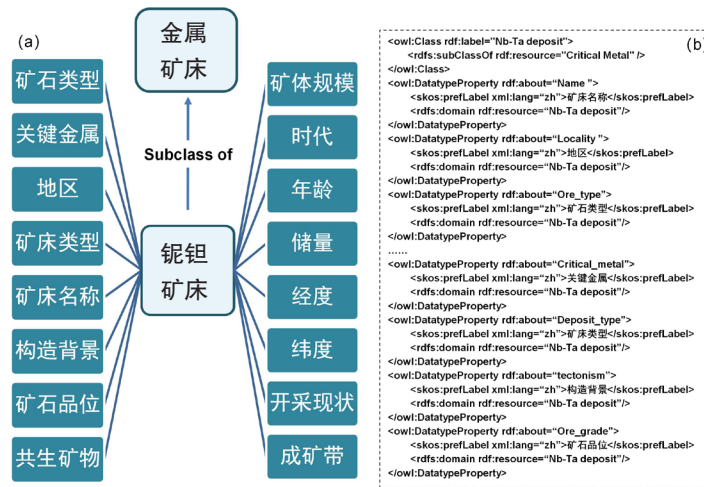


图2 矿床知识图谱本体层概念模型及部分OWL代码描述

Fig. 2 Simple model of the Mineral Deposit Ontology layer (a) and its OWL code (b)

或人工方式对文本中的实例和相互之间的关系进行识别和抽取。例如，文献（聂凤军等，2013）中关于研究对象的描述“内蒙古武川县赵井沟矿床是近年来在内蒙古中部地区找到的一处大型铌钽矿床”，可根据本体层的模式提取出实例名称“赵井沟矿床”，是一种“铌钽矿床”（关系），位于“内蒙古武川县”（属性），组织成三元组即为<“赵井沟矿床”-是一种-“铌钽矿床”>以及<“赵井沟矿床”-地区-“内蒙古武川县”>。

事实层的存储可以基于图数据库的方式，也可以参照关系型数据库的形式以表结构的方式进行存储（图3）。本文根据铌钽矿床本体层的模式采用表结构的形式设计了事实层的概念模式。其中，表1描述了铌钽矿床的基本信息，包括每一个铌钽矿床的地理位置、构造背景、矿床类型、品位等。每

一个矿床具有一个唯一标识符作为主键码；表2为提供矿床基本信息和分析测试数据的文献列表，包括文献的名称、作者、期刊名、发表年份、起止页码等。表3和表4保存了每个矿床岩石采样全岩的主量和微量元素地球化学分析数据，表5和表6保存了矿床样品中矿物的主要化学成分和微量元素含量分析数据，表7保存了矿床样品所含特征副矿物的同位素数据，文献信息表和地球化学数据等关系表通过矿床唯一标识与矿床基本信息表进行连接。年龄测试分析数据通过样品唯一标识符与表5连接。其表结构和相互关系如图3所示。

事实层的数据可以通过文本分词、信息提取和文本挖掘等自然语言处理技术自动建设。尽管基于自然语言处理技术的信息提取方式具有快捷的特点，但是仍然需要本体层提供概念和语义关系，形

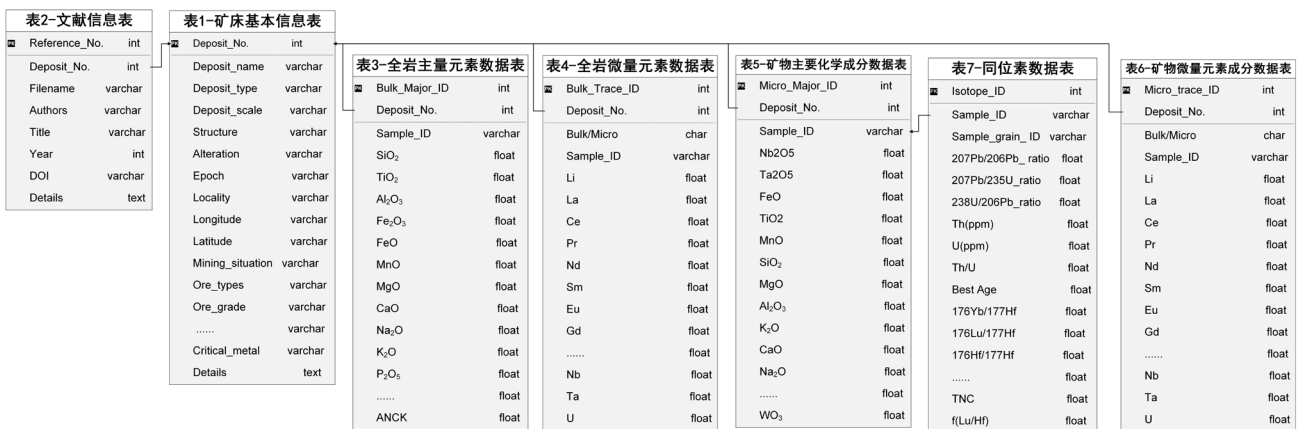


图3 矿床数据库结构图

Fig. 3 Data Structure of the mineral deposit database

成语料库,以增加命名实体提取和语义关系建立的准确性(图4),并通过人工标注和机器学习等方式相结合不断训练以提高提取准确度。

3.3 知识融合

虽然本体层定义了铌钽矿床的知识描述框架,但仍然需要进一步补充完善新的属性描述或者对原有语义关系进行修订。同时,需要融合不同来源的数据对事实层进行更新。此外,矿床研究还需要与其他领域知识(例如岩石学、矿物学、构造等)进行融合,因此也存在不同学科知识图谱之间的跨领域融合的需求。不同用户构建的知识图谱在融合的时候容易出现实体重叠、歧义和关系冲突等问题。这些问题的解决都是需要在知识图谱的构建过程中注意的。

4 矿床知识图谱的功能

4.1 消除语义异构

多源异构数据在数据共享和整合的时候最明显的问题就是语义异构,主要表现为以下两种情况:

(1) 不同数据集用同一术语表示不同的含义

例如不同的实体都具有“名称”这一属性。在岩体表中可能表示的是岩体的名称,而在矿床表中表示的是矿床的名称。机器并不能直接区分这两个同名属性的不同。事实层中的实例和本体层中的概念、对象或者类之间建立联系后,就可以消除歧义,同名属性就不会产生歧义了。

(2) 不同术语在不同数据集中表示相同的含义

例如属性“岩体年龄”,有的数据集用“ $^{207}\text{Pb}/^{206}\text{Pb}$ age”表示,有的用“ $t_{207/206}$ Ma”

表示,因此需要在这些不同名称的属性之间建立对等关系,并标注分析测试年龄的矿物和方法以消除语义异构问题。再如,在一些文献中“成矿母岩”又叫做“赋矿岩体”。在机器自动识别中,“成矿母岩”和“赋矿岩体”都比较容易识别成实体,但是两者之间存在的“共指”关系却很难由机器自动判断出来,因此可以通过实体共指方式对这些实体建立语义对等关系,实现语义翻译。

4.2 隐性知识显性化

知识图谱中明确、清晰的语义关系可以为知识推理提供基础,使得隐含在显性知识中的隐性知识得以挖掘出来。因此可以根据已有的知识图谱,利用语义关系或者推理规则推断出新的实体间关系或者对新的实体关系进行逻辑冲突检测。通过推断隐性知识、检查知识库的不一致性达到知识扩展、知识清洗、甚至产生新知识的目的。例如根据“赵井沟矿床-属于-铌钽矿床”和“铌钽矿床-属于-金属矿床”,可以推断出“赵井沟矿床-属于-金属矿床”;根据“硅酸盐矿物-包含-层状硅酸盐矿物”和“层状硅酸盐矿物-包含-黑云母”,可以推断出“层状硅酸盐矿物-包含-黑云母”;根据“赵井沟矿床-属于-铌钽矿床”和“赵井沟矿床-矿床类型-花岗岩型”,可以推断出“赵井沟矿床-属于-花岗岩型铌钽矿床”。知识推理的简单示意如图5所示。

4.3 提供结构化的知识表达,为新的专题数据库建设提供规范和标准

知识图谱的本体层对其领域内的知识进行了形

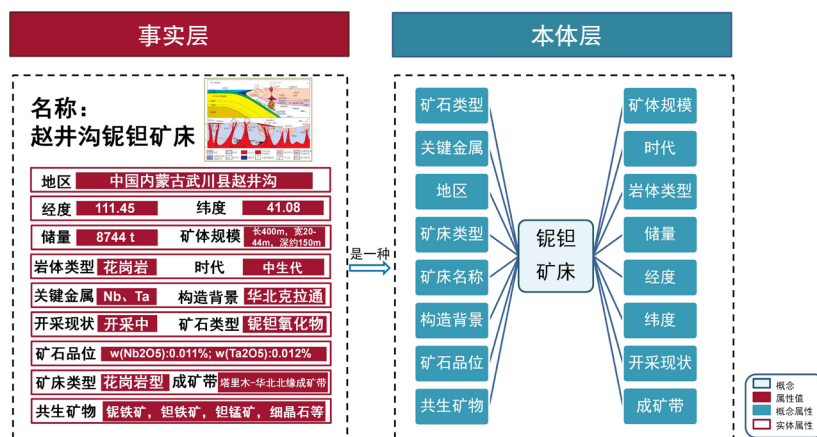


图4 依据本体层模式从文献中提取出的事实层示例

Fig. 4 Construction of the fact layer with information extracted from literature based on the ontology layer

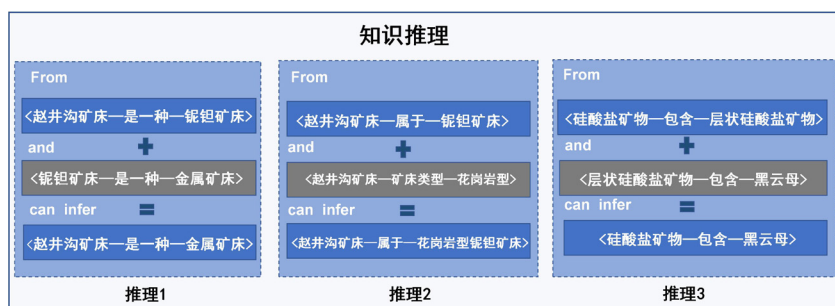


图5 知识推理示意

Fig. 5 Examples of knowledge reasoning

式化的描述,形成了结构化的知识组织网络,提供了领域知识表达的标准,有利于在基于人工智能的模型训练中提供知识建模的基础。同时,对于新建的铌钽矿床专题数据库,知识图谱可以提供统一、规范的数据库逻辑结构,减少甚至避免新建数据库的语义异构问题,使其符合国际 FAIR (Findable, Accessible, Interpretable, Reusable) 准则 (Adam-Blondon et al., 2016),为数据整合、共享与智能检索提供便利。

4.4 提供领域内公用的术语库/语料库

铌钽矿床知识图谱提供了描述铌钽矿床的各类专门术语,包括年代术语、矿物术语、岩石术语、地球化学术语等大量的语料数据,可以为基于自然语言处理技术的文献信息自动提取提供丰富的语料库,提高实体的提取精度和准确度,为机器学习的模型训练等提供准确度高训练样本集,为基于大数据和人工智能的数据挖掘和知识发现提供服务。

5 基于矿床知识图谱的铌钽矿床大数据共享平台

根据上文建立的铌钽矿床知识图谱基本框架,我们在收集整理中国铌钽矿床已发表的中英文文献基础上,提取了相关信息,初步完成了铌钽矿床知识图谱事实层的建设,包括我国铌钽矿床的基本描述信息(矿床名称、地理位置、矿床类型、品位、成矿时期、成矿母岩等),以及公开发表的我国各铌钽矿床岩石地球化学数据和矿物学数据等,形成了覆盖我国近 110 个铌钽矿床(矿化点)的大数据集。

基于铌钽矿床知识图谱,我们在 ArcGIS 中建立了空间数据库,设计了铌钽矿床资源大数据共享

平台。该共享平台具有铌钽矿床大数据的导入导出、浏览、语义检索与数据分析等功能(图 6)。

5.1 数据的导入导出

该平台支持数据的单条或批量的导入导出。随着文献资料的不断更新,数据的导入工作可以持续补充完善大数据集的建设,使得共享平台具有更加全面、系统的铌钽矿床相关数据,实现数据集的持续更新。同时,用户也可根据自己的需要导出表格或是地图集形式的数据,实现铌钽矿床大数据集集成、共享和网络发布。

5.2 数据浏览形式

数据的浏览形式分为表格和地图形式,用户可以根据不同的需要选择不同的浏览形式。表格形式的数据包含矿区的地理位置、类型、时代等特征,也可以在地图上直接显示符合条件的检索结果,并叠加不同主题的图层生成专题地图,方便用户浏览。

5.3 数据的检索

铌钽矿床知识图谱提供了铌钽矿床的详细信息描述,因此可以开展多种信息组合的检索方式。例如可以按照地理位置、矿区、成矿类型、成矿时代等进行组合检索,也可以基于地球化学元素分析结果进行检索。检索结果和数据浏览形式同样可以以表格和地图的形式呈现。尽管从文献资料中提取的各类铌钽矿床信息具有语义异构的问题,由于知识图谱的本体层具有清晰明确的语义描述,能够避免和消除语义异构,实现语义翻译的功能。因此对于用户的数据检索请求,可以通过知识图谱的语义翻译,映射到数据层的数据,使得具有同名异义或者同义异名的相关数据都能得到正确的语义翻译,从而使得用户的检索需求能够正确定位到相应数据集中,实现数据的语义检索和发现。

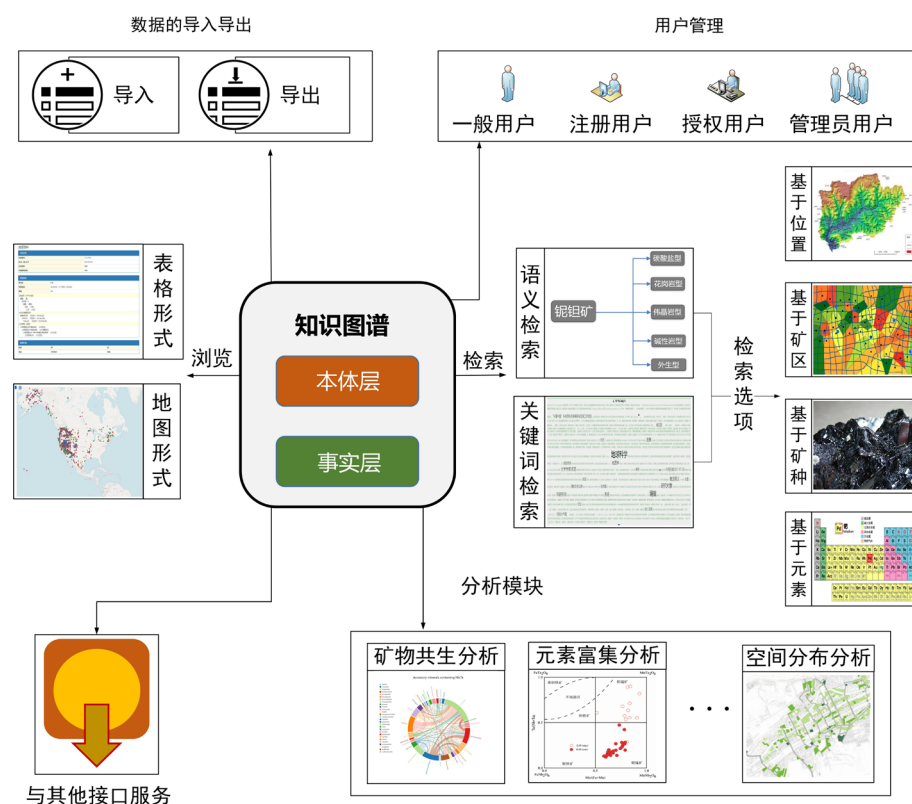


图6 矿床大数据共享平台架构图
Fig. 6 Framework of the mineral deposit big data platform

5.4 数据分析功能

基于铌钽矿床知识图谱和大数据共享平台,本文开展了铌钽矿床时空演化、地球化学和矿物数据特征的挖掘和相关性分析,尤其是我国铌钽矿床空间-时间-构造背景-成因类型多维分布特征以及矿石矿物共生组合特征与铌钽矿床多维分布之间的关联,为进一步揭示铌钽矿床时空演化规律提供新的视角。

5.4.1 铌钽矿床多维特征分布图

图7为根据构建的Nb-Ta矿床知识图谱生成的我国铌钽矿床空间-时间-构造背景-矿床成因类型多维分布图。从图中可以看出,中国铌钽矿床空间上主要分布于华南、新疆北部、四川西部、西藏等地,内蒙古、陕西、河南等地有零星分布。根据铌钽矿床赋存的岩石类型,铌钽矿床可分为花岗岩型、伟晶岩型、碱性岩型、碳酸岩型和外生型等成因类型,我国重要的铌钽矿床类型是花岗岩型和伟晶岩型,其中花岗岩型铌钽矿床主要分布于华南扬子克拉通和华夏板块,成矿时代集中于燕山期,著名矿床有江西宜春、江西松树岗和广西栗木等铌

钽矿床;伟晶岩型矿床则主要位于中亚造山带、松潘甘孜造山带和喜马拉雅造山带,成矿时代为印支期和喜马拉雅期,代表性矿床有新疆可可托海、四川甲基卡、西藏错那洞等;碱性和碳酸岩型矿床分布相对较少,多位于造山带和克拉通交界处,如湖北竹山土地岭碱性岩铌钽矿床位于昆仑-祁连-秦岭-大别造山带东段和扬子克拉通拼接处,碳酸岩型矿床以内蒙古白云鄂博铌铁矿床为代表,两者成矿时代均以加里东期为主;外生型铌钽矿床在我国基本不发育,目前研究程度相对较低。

5.4.2 矿石矿物共生组合关系与铌钽矿床多维特征的关联分析

图8为根据Nb-Ta矿床知识图谱生成的含铌钽副矿物共生组合与铌钽矿床多维特征关联分析图。图中矿床构造背景用方形表示,成因类型用三角形表示,铌钽矿物种类则用圆形表示,而图形(圆形、方形和三角形)的大小则代表着该铌钽矿物种类、矿床构造背景和成因类型出现的频率。图形越大,表明该铌钽矿物、矿床构造背景和成因类型越常见,图形越小,说明该铌钽矿物,矿床构造背景

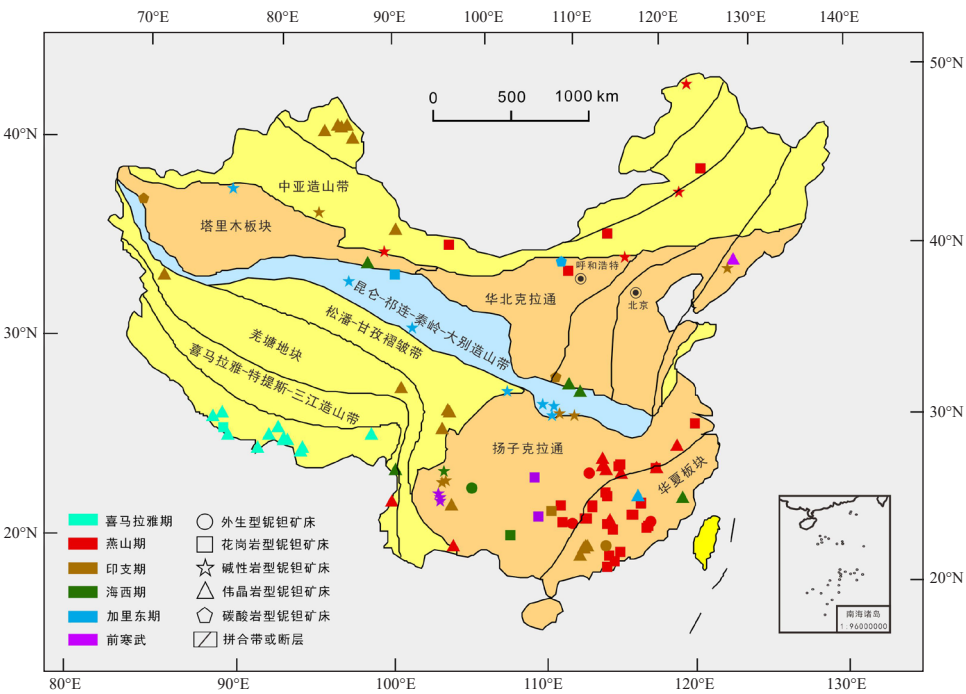


图7 中国铌钽矿床多维特征分布图（底图据毛景文等，2019修改）
Fig. 7 Multidimensional China Nb-Ta deposits distribution map

和成因类型越少见。构造背景、矿床类型和铌钽矿物种类各图形之间关系连线的颜色深浅代表两个特征之间相关性大小。线段颜色越深，说明二者相关性越高，联系越紧密；颜色越浅，说明二者相关性越低，联系越疏远。

自然界铌钽矿物约有 70 多种，从图 8 中看

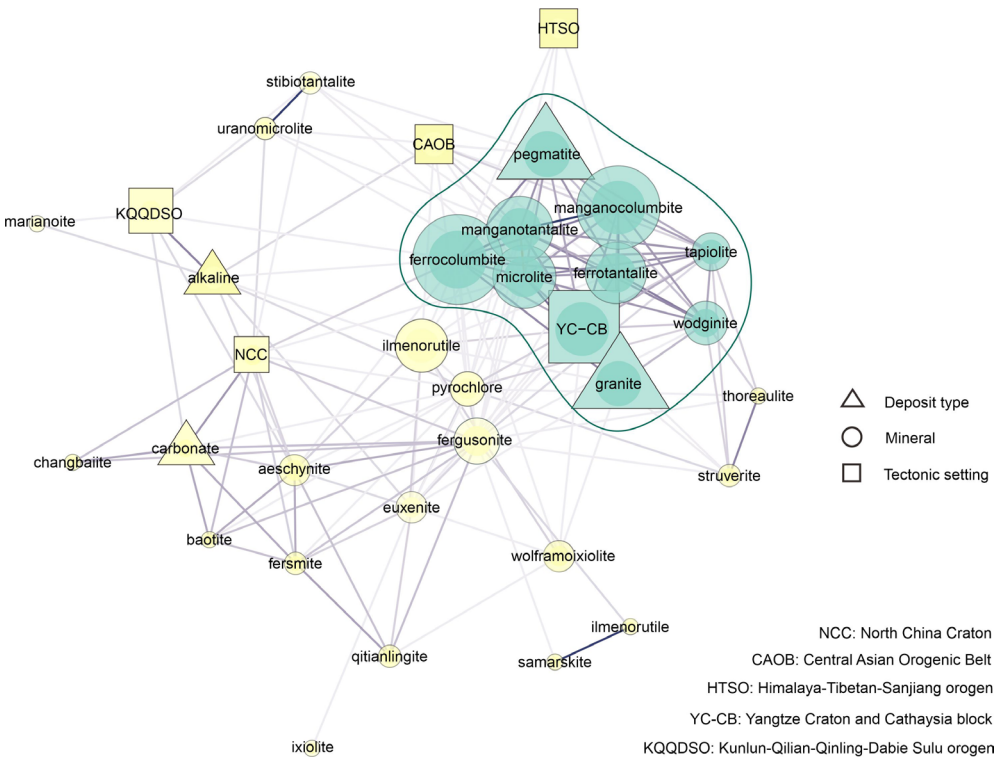


图8 含铌钽副矿物的共现图
Fig. 8 Co-occurrence diagram of Niobium and Tantalum rich accessory minerals

出,我国主要产出的铌钽矿物有25种,其中最常见的是有铌锰矿(manganocolumbite)、铌铁矿(ferrocolumbite)、钽铁矿(ferrotantalite)、钽锰矿(manganotantalite)、细晶石(microlite)、铌铁金红石(ilmenorutile)、重钽铁矿(tapiolite)、锡锰钽矿(wodginite)等,它们和花岗岩型、伟晶岩型铌钽矿床关系相关性高(或紧密程度高),连线颜色深又密集,且多集中分布于扬子克拉通和华夏板块,其次见于中亚造山带和喜马拉雅—三江造山带;与碱性岩型、碳酸岩型矿床特征铌钽副矿物主要有褐钇铌矿(fergusonite)、易解石(aeschnite)、铌钙矿(fermite)、包头矿(baotite)等,多见于华北克拉通和昆仑—祁连山—秦岭—造山带,它们之间的连线相对较密且关系密切。还有一些铌钽副矿物,如铌铁金红石(ilmenorutile)、烧绿石(pyrochlore)、黑稀金矿(euxenite)在四个类型矿床中均可见,但从图8中可注意铌铁金红石和烧绿石更靠近花岗岩型矿床和扬子克拉通,关系密切程度明显高于碳酸岩型和华北克拉通。另外,我国还有较罕见的铌钽矿物产出,如长白矿(changbaiite)和硅钨铌钙钠石(marianoite)。

上述铌钽矿床空间—时间—构造背景—成因类型多维分布特征,以及矿石矿物共生组合与铌钽矿床类型、构造背景之间多维关联分析结果与前人文献(Zhang et al., 2004; 郝雪峰等, 2015; Zhu et al., 2015; 王汝成等, 2017; 王盘喜等, 2017; 康清清等, 2018; Che et al., 2019; 鲁显松等, 2019; Xie et al., 2019; 马万伟等, 2020; 王汝成等, 2020; 秦玉芳等, 2021)基本一致。说明基于大数据的分析方法在揭示铌钽矿床时空分布特征方面具有可靠性和可行性。但以上仅仅是基于Nb-Ta矿床知识图谱获得的一些初步认识,更深入的讨论和分析有待于知识图谱的进一步完善,以及与岩石学、矿物学、地球化学、年代学等知识图谱的横向融合和数据驱动下的知识发现。

5.5 数据的管理服务

数据的管理服务主要分为矿床数据的导入、导出、删除、修改等基本操作。这些功能仅对授权用户开放。用户被设置为四类,分别是一般用户、注册用户、授权用户和管理员用户。管理员用户具有数据管理的全部功能,以及用户权限的授予和收回。

授权用户可在管理员授权后,具有对授权范围内的数据的管理权。注册用户在注册之后可以进行基本的浏览、分析、导出、检索等操作,而一般用户则只能使用浏览和检索两项功能。

6 总结

知识图谱是大数据时代数据共享、整合以及复用的重要基础。作为大数据和人工智能应用的基础设施之一,知识图谱在知识建模和知识发现中的作用收到了越来越广泛的关注。本文以铌钽矿床知识图谱的建设为例,阐述了知识图谱的逻辑结构和基本框架,构建了铌钽矿床知识图谱的本体层和事实层,并对铌钽矿床知识图谱的应用做了一些探索。

知识图谱的建设不是一蹴而就的,需要不断的更新和完善。矿床知识图谱的不断成熟和完善,以及与其他学科知识图谱的深度交叉融合,将更加有利于基于知识图谱的矿床大数据分析和深度数据挖掘,为基于大数据和人工智能技术的关键金属核心科学问题的求解提供新的思路。

参考文献(References):

- 曹倩, 赵一鸣. 2015. 知识图谱的技术实现流程及相关应用[J]. 情报理论与实践, 38(12): 127-132.
- 常力恒, 朱月琴, 汪新庆, 等. 2018. 大数据环境下的矿产知识库构建:以钨矿为例[J]. 中国矿业, 27(9): 93-96+108.
- 陈骏. 2019. 关键金属超常富集成矿和高效利用[J]. 科技导报, 37(24): 3.
- 邓志鸿, 唐世渭, 张铭, 等. 2002. Ontology研究综述[J]. 北京大学学报(自然科学版), 5: 730-738.
- 董少春, 齐浩, 胡欢. 2019. 地球科学大数据的现状与发展[J]. 科学技术与工程, 19(20): 1-11.
- 郝雪峰, 付小方, 梁斌, 等. 2015. 川西甲基卡花岗岩和新三号矿脉的形成时代及意义[J]. 矿床地质, 34(6): 1199-1208.
- 侯增谦, 杨志明, 王瑞, 等. 2020. 再论中国大陆斑岩Cu-Mo-Au矿床成矿作用[J]. 地学前缘, 27(2): 20-44.
- 侯志伟, 诸云强, 高楹, 等. 2018. 地质年代本体及其在语义检索中的应用[J]. 地球信息科学学报, 20(1): 17.
- 胡瑞忠, 温汉捷, 叶霖, 等. 2020. 扬子地块西南部关键金属元素成矿作用[J]. 科学通报, 65(33): 3700-3714.
- 黄恒琪, 于娟, 廖晓, 等. 2019. 知识图谱研究综述[J]. 计算机系统应用, 28(6): 1-12.
- 康清清, 江宏君, 李鹏, 等. 2018. 陕西华阳川钨钼矿床矿石矿物学特征[J]. 东华理工大学学报(自然科学版), 41(2): 14-26.
- 黎洁, 谢贤, 吕晋芳, 等. 2021. 铌矿资源概述及选矿技术研究进展[J]. 金属矿山, 536(2): 120-126.
- 李善平, 尹奇麟, 胡玉杰, 等. 2004. 本体论研究综述[J]. 计算机研究与发展, 7: 1041-1052.
- 刘峤, 李杨, 段宏, 等. 2016. 知识图谱构建技术综述[J]. 计算机研

- 究与发展, 53(3): 582–600.
- 鲁显松, 黄景孟, 熊意林, 等. 2019. 南秦岭土地岭铌钽矿床火山岩地球化学、锆石U–Pb年代学特征及地质意义[J]. 地质科技情报, 38(3): 40–51.
- 陆锋, 余丽, 仇培元. 2017. 论地理知识图谱[J]. 地球信息科学学报, 19(6): 723–734.
- 马胜男, 孙翊, 郭明明. 2010. SWEET本体研究述评[J]. 标准科学, 9: 38–43.
- 马万伟, 黄小龙, 于洋, 等. 2020. 栗木花岗岩中云母的特征: 对锡成矿热液作用过程的指示[J]. 大地构造与成矿学, 44(6): 17.
- 毛景文, 曾载淋, 李通国, 等. 2019. 21世纪以来中国关键金属矿产找矿勘查与研究新进展[J]. 矿床地质, 38(5): 935–969.
- 聂凤军, 王丰翔, 赵宇安, 等. 2013. 内蒙古赵井沟大型铌钽矿床地质特征及成因[J]. 矿床地质, 32(4): 730–743.
- 齐浩, 董少春, 张丽丽, 等. 2020. 地球科学知识图谱的构建与展望[J]. 高校地质学报, 26(1): 2–10.
- 秦玉芳, 王其伟, 金海龙, 等. 2021. 白云鄂博矿霓石型稀土矿石中铌的赋存状态与分布规律研究[J]. 矿产保护与利用, 41(3): 144–149.
- Selner G I, Taylor R B, 陆玲. 1992a. GSMRDS数据库的管理[J]. 华东地质学院学报, 15(4): 15–25.
- Selner G I, Taylor R B, 陆玲. 1992b. GSMRDS数据库与GSMAP图件, MRDS数据库的关系[J]. 华东地质学院学报, 15(4): 324–336.
- Selner G I, Taylor R B, 陆玲. 1992c. 矿产资源数据库—GSMRDS数据库[J]. 华东地质学院学报, 15(4): 301–312.
- 石顺中, 闫海荣, 董少春, 等. 2020. 地球科学知识体系编辑平台[J]. 高校地质学报, 26(4): 384–394.
- 王盘喜, 朱黎宽, 刘璐, 等. 2017. 河南官坡花岗岩伟晶岩地质与地球化学特征[J]. 中国地质调查, 4(6): 40–49.
- 王汝成, 车旭东, 郭斌, 等. 2020. 中国铌钽锆铅资源[J]. 科学通报, 65(33): 119–133.
- 王汝成, 吴福元, 谢磊, 等. 2017. 藏南喜马拉雅淡色花岗岩稀有金属成矿作用初步研究[J]. 中国科学: 地球科学, 47(8): 871–880.
- 吴永亮, 贾志杰, 陈建平, 等. 2017. 基于大数据智能的找矿模型构建与预测[J]. 中国矿业, 26(9): 79–84.
- 徐增林, 盛泳潘, 贺丽荣, 等. 2016. 知识图谱技术综述[J]. 电子科技大学学报, 45(4): 589–606.
- 姚健鹏, 郭艳军, 潘懋, 等. 2017. 铜矿床领域本体的构建方法研究[J]. 中国矿业, 26(8): 140–145.
- 叶帅. 2019. 基于Neo4j的煤矿领域知识图谱构建及查询方法研究[D]. 徐州: 中国矿业大学.
- 赵军. 2018. 知识图谱. 人工智能丛书[M]. 北京: 高等教育出版社.
- 周永章, 张前龙, 黄永健, 等. 2021. 钦杭成矿带斑岩铜矿知识图谱构建及应用展望[J]. 地学前缘, 28(3): 67–75.
- Adam-Blondon A F, Alaux M, Pommier C, et al. 2016. Towards an open grapevine information system [J]. Horticulture Research, 3: 16056.
- Baru C, Chandra S, Lin K, et al. 2009. The GEON service-oriented architecture for Earth science applications [J]. International Journal of Digital Earth, 2(S1): 62.
- Che X D, Wang R C, Wu F Y, et al. 2019. Episodic Nb–Ta mineralisation in South China: Constraints from in situ LA–ICP–MS columbite–tantalite U–Pb dating [J]. Ore Geology Reviews, 105: 71–85.
- Enkhsaikhan M, Holden E J, Duuring P, et al. 2021. Understanding ore-forming conditions using machine reading of text [J]. Ore Geology Reviews, 135(2): 104200.
- Gruber T R. 1993. A translation approach to portable ontology specifications [J]. Knowledge Acquisition, 5(2): 199–220.
- Gutierrez C and Sequeda J F. 2021. Knowledge graphs [J]. Communications of the ACM, 64(3): 96–104.
- Heidorn P B. 2008. Shedding light on the dark data in the long tail of science [J]. Library Trends, 57(2): 280–299.
- Ludascher B, Lin K, Brodaric B, et al. 2003. GEON: toward a cyberinfrastructure for the geosciences—A prototype for geologic map integration via domain ontologies [R]. United States Geological Survey.
- Ma X G, Carranza E J M, Wu C, et al. 2012. Ontology-aided annotation, visualization, and generalization of geological time-scale information from online geological map services [J]. Computers & Geosciences, 40: 107–119.
- Nickel M, Murphy K, Tresp V, et al. 2015. A review of relational machine learning for knowledge graphs [J]. Proceedings of the IEEE, 104(1): 11–33.
- Osipova I and Gospodinova V. 2020. Representation of the process of sudden outbursts of coal and gas using a knowledge graph [J]. E3S Web of Conferences, 192(83): m04022.
- Sen M and Duffy T. 2005. GeoSciML: Development of a generic GeoScience Markup Language [J]. Computers & Geosciences, 31(9): 1095–1103.
- Singhal A. 2012. Introducing the knowledge graph: things, not strings [R]. America: Official Blog of Google, <https://www.blog.google/products/search/introducing-knowledge-graph-things-not/>.
- Wang C B, Ma X G, Chen J G, et al. 2018. Information extraction and knowledge graph construction from geoscience literature [J]. Computers & Geosciences, 112: 112–120.
- Wang Z, Zhang J W, Feng J L, et al. 2014. Knowledge graph embedding by translating on hyperplanes [C] // Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence: 1112–1119.
- Xie L, Tao X Y, Wang R C, et al. 2019. Highly fractionated leucogranites in the eastern Himalayan Cuonadong dome and related magmatic Be–Nb–Ta and hydrothermal Be–W–Sn mineralization[J]. Lithos, 354–355(7): 105286.
- Zhang A C, Wang R C, Hu H, et al. 2004. Chemical evolution of Nb–Ta oxides and zircon from the Koktokay No. 3 granitic pegmatite, Altai, northwestern China [J]. Mineralogical Magazine, 68(5): 739–756.
- Zhu Z Y, Wang R C, Che X D, et al. 2015. Aegirite-hydrothermal rare-element mineralization in the Songshugang granite (northeastern Jiangxi, China): Insights from an electron-microprobe study of Nb–Ta–Zr minerals [J]. Ore Geology Reviews, 64: 749–760.