地球表层系统科学数据叙词表

陈 锦[1]，王 曙[2,3*]，诸云强[2,3]，段福洲[1]，王 斌[4*]

1. 首都师范大学资源环境与旅游学院，北京 100048；
2. 中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室，北京 100101；
3. 江苏省地理信息资源开发与利用协同创新中心，南京 210023；
4. 中国地质调查局自然资源综合调查指挥中心，北京 100055

摘 要：地球表层系统科学数据叙词表是以规范化形式描述地球表层系统各圈层术语及其间基本语义关系的词汇表。作为基础性的数据资源，高质量的地球表层系统科学数据叙词表有助于概念的辨析对比、资源的组织检索、数据的标准化与共享，从而推动地球表层系统跨学科的研究。本文基于领域主题词词典（如全球变化主目录）、领域专著（如地理辞典）、领域本体（如地球与环境术语语义网）及在线资源（如维基百科）等主题词数据，它们涵盖全球变化、地理环境、资源体系等领域，明确了地球表层系统的定义与范围，构建了高质量地球表层系统科学数据叙词表。其中具体包含基础空间、圈层、系统、子系统、对象、要素、属性七层概念，总计 3,463 个主题词，以及主题词之间的等同关系、层级关系和相关关系，共计 4.454 个。研究表明，叙词表在词表规模与词表性能两方面表现良好，有望为地球表层系统数据网络构建、信息关联对齐、信息资源检索、知识服务与知识发现等方面提供数据支撑。数据集以.xlsx 格式存储，由 3 个数据文件组成，数据量为 1.84 MB（压缩为 1 个文件，1.78 MB）。

关键词：地球表层系统；科学数据；主题词；叙词表；本体模型；知识服务

DOI: https://doi.org/10.3974/geodp.2024.02.01

CSTR: https://cstr.escience.org.cn/CSTR:20146.14.2024.02.01

数据可用性声明：

本文关联实体数据集已在《全球变化数据仓储电子杂志（中英文）》出版，可获取：https://doi.org/10.3974/geodb.2024.07.10.V1 或 https://cstr.escience.org.cn/CSTR:20146.11.2024.07.10.V1.

1 前言

地球表层（Earth Surface）不仅关注地球表面空间地理特征，还聚焦生物与自然环境的相互关系，是地理学研究的核心对象[1, 2]。随着地理学的不断发展，学者们逐渐认识到地球表层是一个开放且具有物质和能量交换的复杂巨系统[3]。地球表层系统（Earth Surface

收稿日期：2024-04-08；**修订日期：**2024-06-10；**出版日期：**2024-06-25

基金项目：中华人民共和国科学技术部（2022YFF0711601，2022YFB3904201）；国家自然科学基金（42101467）；资源与环境信息系统国家重点实验室创新项目（KPI009）

***通讯作者：**王曙，中国科学院地理科学与资源研究所，wangshu@igsnrr.ac.cn；王斌，中国地质调查局自然资源综合调查指挥中心，wangbincgs@mail.cgs.gov.cn

数据引用方式：[1] 陈锦，王曙，诸云强等. 地球表层系统科学数据叙词表[J]. 全球变化数据学报，2024，8(2): 111−124. https://doi.org/10.3974/geodp.2024.02.01. https://cstr.escience.org.cn/CSTR:20146.14.2024.02.01.

[2] 陈锦，王曙，诸云强等. 地球表层系统科学数据叙词表[J/DB/OL]. 全球变化数据仓储电子杂志，2024. https://doi.org/10.3974/geodb.2024.07.10.V1. https://cstr.escience.org.cn/CSTR:20146.11.2024.07.10.V1.

System，简称地表系统）揭示了气候、生物、水体、地质、土壤等基本要素之间的相互作用与影响及其在时空上的演化和发展。与此同时，地球表层巨系统内部各要素的演化和发展所产生的多源、异构、海量、复杂的时空知识推动着地表系统科学向数据密集型科学发展[4, 5]。如何管理和利用诸如气候变化度量、极端灾害事件预报、生态环境监测等地表系统科学数据对更好地管理地球资源、维护环境可持续性以及预测自然灾害等方面至关重要[6, 7]。

叙词表，又称主题词表，是信息管理中重要的组织与检索工具，用于规范化地描述和分类领域内特定的概念或术语[8]。在地球表层相关学科的研究领域内，叙词表的构建在单一学科或交叉学科的研究中已经积累了一定的基础。例如,《地理科学叙词表》[9]覆盖自然、人文、区域地理等领域的专业术语;《地质学汉语叙词表》[10, 11]专注于岩石矿物、地质构造等方面的主题词;《环境科学叙词表》[12]则囊括了环境科学领域检索的专用术语。同时，也存在一些涵盖地学主题词的综合性叙词表，如《中国分类主题词表》[13]涉及自然科学各领域学科和主题概念;《NASA 叙词表》[14]聚焦自然空间科学领域，同时兼顾地球科学。这些叙词表涵盖基础地理、资源环境、地质地貌等地学涉及领域。然而，这些词表单独难以完全覆盖地表系统所研究的核心主题，不同词表对于相同概念的内涵可能存在一定分歧，词表间的数据难以实现共享，尚未形成专注于地表系统领域的统一化、标准化的知识体系。综上所述，单一学科或交叉学科形成的叙词表存在概念定义不统一、难以完全覆盖地表系统领域的核心概念等问题，当前业界缺乏全面、完整、准确的地球表层系统科学数据叙词表。

因此，构建《地球表层系统科学数据叙词表（Earth Surface System Scientific Data Thesaurus）》（简称《地表科学数据叙词表》），能够更好地梳理地表系统领域所涵盖的关键对象、概念及其相互关联，为组织、存储和利用地表系统科学数据提供便捷途径。针对上述问题，本文手工构建了高质量的《地表科学数据叙词表》，以期为地球表层系统数据网络构建、信息关联对齐、信息资源检索、知识服务与知识发现等方面提供数据支撑。

2 数据集元数据简介

《地表科学数据叙词表（V1.0）》[15]的元数据信息见表 1。

3 数据研发方法

《地表科学数据叙词表》采用自顶向下和自底向上相结合的策略，通过综合利用领域权威词典、专著、本体和在线资源等多源数据，确保了词表的全面性和专业性。同时，通过设计层次化的结构框架和语义关系，实现了地表科学数据的有效组织和标准化，以支持数据的分析、应用和共享。本节将详细阐述《地表科学数据叙词表》的构建方法。

3.1 地球表层系统科学数据叙词表构建技术路线

《地表科学数据叙词表》采用“自顶向下”和“自底向上”的相结合方式进行构建，其总体构建技术路线如图 1 所示。首先，结合叙词表的基础及通用应用需求，如地表科学数据共享服务、应急灾害知识服务等，明确地表系统科学数据的内涵与范围。其次，通过

表 1 《地球表层系统科学数据叙词表》元数据简表

条　目	描　述
数据集名称	地球表层系统科学数据叙词表
数据集短名	ESSSD_Thesaurus
作者信息	陈锦，首都师范大学，cj15160172956@163.com 王曙，中国科学院地理科学与资源研究所，wangshu@igsnrr.ac.cn 诸云强，中国科学院地理科学与资源研究所，zhuyq@igsnrr.ac.cn 段福洲，首都师范大学，duanfuzhou@263.net 王斌，中国地质调查局自然资源综合调查指挥中心，wangbincgs@mail.cgs.gov.cn
数据格式	.xlsx
数据量	1.84 MB，压缩后 1.78 MB
数据集组成	主题词中英文名称、中英文描述、主题词关系、主题词分类、数据源
基金项目	中华人民共和国科学技术部（2022YFF0711601，2022YFB3904201）；国家自然科学基金（42101467）；资源与环境信息系统国家重点实验室创新项目（KPI009）
出版与共享服务平台	全球变化科学研究数据出版系统 http://www.geodoi.ac.cn
地址	北京市朝阳区大屯路甲 11 号 100101，中国科学院地理科学与资源研究所
数据共享政策	（1）“数据”以最便利的方式通过互联网系统免费向全社会开放，用户免费浏览、免费下载；（2）最终用户使用“数据”需要按照引用格式在参考文献或适当的位置标注数据来源；（3）增值服务用户或以任何形式散发和传播（包括通过计算机服务器）“数据”的用户需要与《全球变化数据学报（中英文）》编辑部签署书面协议，获得许可；（4）摘取“数据”中的部分记录创作新数据的作者需要遵循 10%引用原则，即从本数据集中摘取的数据记录少于新数据集总记录量的 10%，同时需要对摘取的数据记录标注数据来源[16]
数据和论文检索系统	DOI，CSTR，Crossref，DCI，CSCD，CNKI，SciEngine，WDS，GEOSS，PubScholar，CKRSC

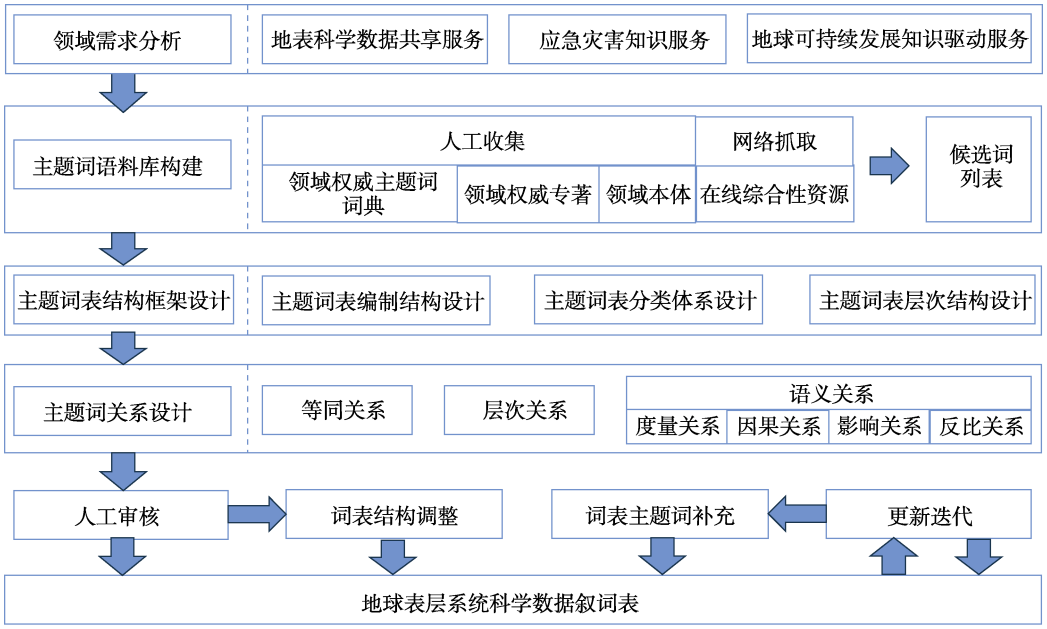

图 1　《地球表层系统科学数据叙词表》构建技术路线图

人工汇集或网络抓取等方式从领域权威主题词词典和综合性在线资源等多种数据来源中收

集并分析地表系统科学数据领域的术语和概念，建立地表系统科学数据主题词语料库，确定词表的候选词列表。然后，利用自顶向下的方式，从宏观角度建立词表的整体结构框架，确定词表的整体编制结构、分类体系和层次结构。接着，结合词表的结构框架，设计候选词间的基本语义关系，包括等同关系、层次关系以及语义关系等。最后，采用自底向上的方式从候选词中确定主题词，通过人工审核的方式对词表的结构进行调整，细分主题词的类别和层级，实现词表的更新迭代与词汇补充。

3.2 地球表层系统的定义与范围

构建系统完善的《地表科学数据叙词表》是地表系统数据的分析、应用、共享与知识服务的基础，其分类和内容的构建离不开对地表系统概念、内涵及范围的准确界定。

近代科学家对地表系统概念与研究范围的不同见解在《地表科学数据叙词表》的涵盖范围界定中具有深远的意义。德国地理学家李希霍芬于 1883 年提出“地球表层”的概念，苏联地理学家布罗乌诺夫于 1910 年进一步将其定义为同心圈层[17, 18]。随着大陆漂移[19]、海底扩张[20]、板块构造[21]和盖亚假说[22]等重要理论的提出，研究者们对地表系统的内涵与范围有了更深入的理解，如表 1 所示。地理学角度认为地表系统是围绕人类活动的地球各圈层的耦合整体[18, 23]。地球科学宏观角度将地表系统视为地球内外部能量和物质交换的复杂系统[24, 25]。生态学角度将地表系统定义为供给人类和生态系统的地理空间载体[26]。自然资源角度则认为地表系统是地球系统给人类生产生活提供基础性生存条件的核心空间[25]。

表 2 不同研究者对地表系统范围的差异性描述

序号	包含圈层	地表系统范围描述	学者
1	大气圈、水圈、生物圈、岩石圈	从大气对流层的顶层延伸至岩石圈表面及海洋深处	李希霍芬(1883)[17]；布罗乌诺夫(1910)[18]；黄秉维(1996)[23]；Jin *et al.* (2023)[27]
2	大气圈、水圈、土壤圈、岩石圈	由地面以下凝固相区间延伸至人造地球卫星运行轨道高度以内的流动相区间所构成的地球表面附近空间	窦学成(1998)[28]
3	大气圈、水圈、生物圈、人类圈、智慧圈、土壤圈、岩石圈	由大气圈、水圈、生物圈、人类圈、智慧圈、土壤岩石圈等形成层状分离、相互渗透的同心圈层	浦汉昕(1983)[29]
4	大气圈、水圈、生物圈、人类圈、土壤圈、岩石圈	由大气圈、生物圈、人类圈、水圈、土壤圈、岩石圈等圈层相互作用所构成的复杂开放巨系统	陆大道(1999)[30]；Phillips (2002)[31]；张猛刚等(2005)[32]；王成善等(2009)[24]；李晓亮等(2022)[33]；Chen, et al. (2023)[34]
5	大气圈、水圈、生物圈、岩石圈、土壤圈	包括地球表面上下的岩石圈、水圈、大气圈、生物圈和近地物理场，下界是软流层，上界为大气圈最外层	周俊(2004)[17]
6	大气圈、水圈、生物圈、人类圈、岩石圈、地核、天体	包括大气对流层、水体层、陆地结构以及与这些层系交互的生物层和人文层。气、水、壳三大层系的界面耦合以及地外和地内动力作用是研究的核心	马宗晋等(2006)[35]
7	大气圈、冰冻圈、水圈、人类圈、陆地、岩石圈	大气圈、冰冻圈、陆地、海洋和岩石圈相互作用，涵盖物理、化学和生物过程，人类活动是系统功能的一部分	Steffen (2020)[25]

学术界对于地表系统的分歧主要体现在圈层边界的划定方面，具体包括地表系统下边界岩石圈和上边界大气圈的准确界定。尽管不同视角对地表系统的理解存在侧重的差异，但总体存在一定共识。首先，在地表系统的基本定位上，都认同地表系统是由多圈层共同构成的有机复杂系统，这些圈层相互耦合，变化态势显著，物质能量信息流通频繁。其次，都认为地表系统的核心圈层自下而上涵盖了岩石圈（部分）、土壤圈、生物圈、人类圈、水圈、大气圈（部分）。最后，都认为地表系统是地球各圈层交互作用和人类活动最为活跃的区域，为人类提供持续供给。

综上所述，本文将地表系统定义为以大气圈、水圈、生物圈、人类圈、土壤圈和岩石圈为研究对象的开放复杂巨系统，涉及大气、水域、生态系统、人类社会和地质构造等要素，这些要素之间相互联系，形成一个相互依存、不断变化的整体系统。

3.3　地球表层系统科学数据叙词表数据源

考虑到地表系统研究的科学性与严谨性，《地表科学数据叙词表》的词汇来源主要包含四类：领域权威主题词词典、领域权威专著、领域本体以及在线综合性资源（详见表 3）。领域权威主题词词典由地球科学领域的权威机构或组织维护，包含专业领域审核和认可的术语或概念，有助于确保词表的专业性和准确性。领域专家编写的权威专著涵盖丰富的科学知识和术语，为词表提供可靠的背景信息和专业词汇。地表系统领域本体是对领域概念和关系形式化的知识表示，有助于更好地理解领域内的知识结构，进而帮助建立词汇的分类和层次结构。互联网上存在大量关于地表系统的综合性资源，提供广泛的背景信息和分类索引，可用于与其他数据源进行相互验证、补充和丰富词表的概念内容，从而提高词表的质量和覆盖范围。通过综合利用这些数据源，可以确保词表的全面性、准确性和适应性，为地表系统科学数据的组织和标准化提供有力支持。

表 3　《地球表层系统科学数据叙词表》数据源

数据源类型	数据源	数据源覆盖的研究领域
领域权威主题词词典	全球变化主目录（Global Change Master Directory, GCMD）[36]	大气圈、生物圈、人文因素、陆地表层、陆地水圈、固体地球
领域权威专著	《地理辞典》[37]	自然地理、人文地理、资源地理
	《地球系统研究与科学数据》[38]	大气、陆地表层、海洋、岩石圈、外层空间
	《地球系统科学数据资源体系研究》[39]	大气圈、人地关系、固体地球、陆地表层、海洋
	《地球系统科学数据集成共享研究：标准视角》[40]	大气圈、陆地表层、生物圈、冰冻圈、自然资源、人文因素、海洋极地、固体地球
领域本体	地球与环境术语语义网(Semantic Web for Earth and Environmental Terminology, SWEET)[41, 42]	地质特征、人类活动、自然现象
在线综合资源	维基百科[1]	自然科学、人文社科
	百度百科[2]	自然科学、人文社科

3.4　叙词表结构框架设计

3.4.1　叙词表编制结构设计

叙词表由主表和辅表组成[43]。主表是叙词表的核心组成部分，按照一定的顺序进行组

[1] https://zh.wikipedia.org.
[2] https://baike.baidu.com.

织，如英文字母顺序、汉语拼音顺序等，包括全部主题词词汇及其相关语义关系。辅表则重新组织了主表的结构，以满足用户多角度的检索需求，通常包括分类表、索引表、附表等形式。为满足地表系统科学数据的研究需求，《地表科学数据叙词表》以主表和分类表两种形式呈现。分类表根据词表分类体系构建，便于用户分析主题词间的层级关系。

3.4.2 叙词表分类体系设计

如图 2 所示，为实现地表系统科学数据的有效挖掘、管理与共享，本文基于地表系统结构特征，结合 GCMD 分类思想和科学数据共享分类特征，将地表系统科学数据分为近地空间层、地表覆被层和地表机制层三大类[44, 45]。其中，近地空间层涵盖大气科学和气象学等领域，关注大气圈的各种特性和过程，用于理解气象、气候和大气环境的变化。地表覆被层包括水体、土壤、人类与其他生物活动区域等，涵盖海洋活动、生态系统相互作用、土地利用和覆盖类型等，有助于理解生态系统、资源管理等问题。地表机制层包括岩石圈和地壳内部的地质和地球物理过程，涵盖地质构造、火山活动、岩石矿物资源等，有助于理解固体地球学和矿物资源管理等问题。

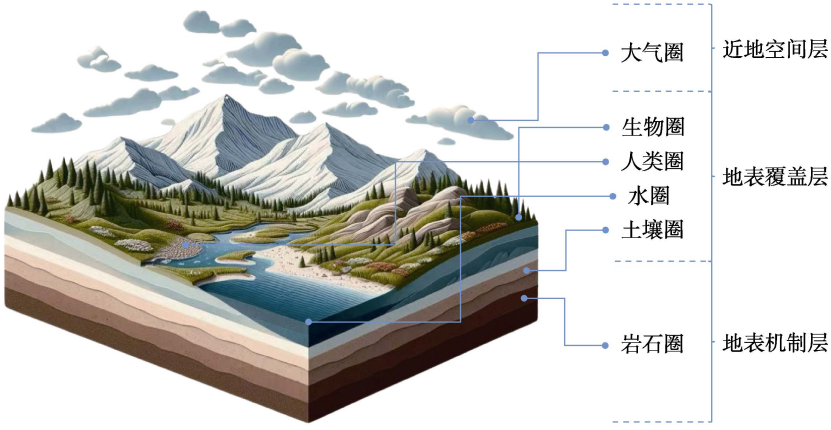

图 2 地球表层系统科学数据圈层分类示意图

在此基础上，结合地球圈层结构，将三大主题类型进一步划分为 6 个圈层类型。各个圈层结合自身的特点进行细分，以更好地反映地表系统科学数据的复杂性和多样性。《地表科学数据叙词表》三级分类体系如表 4 所示，包含 6 个二级大类，35 个三级类目，三级类目下直接列出主题词。

3.4.3 叙词表层次结构设计

层次结构的设计旨在凸显主题词之间的层级关系。在《地表科学数据叙词表》的层次结构设计中，主要参考了 GCMD 的树形层级结构思想和地学相关学科的分类标准。GCMD 关键词按照“类别>主题>术语>变量>详细变量”的多级树形结构对概念进行分类和关联。因此，在设计《地表科学数据叙词表》的层次结构时，遵循科学性、系统性、精确性等原则，将其按照“基础空间>圈层>系统>子系统>对象>要素>属性”的层次结构进行组织，如图 3 所示。其中，基础空间代表地表系统科学数据所涵盖的地理和空间范围，是词表的顶级层次。圈层即构成地表系统的 6 种基本圈层结构。系统表示每个圈层内部的主要领域。

子系统进一步细化圈层的子领域，以更好地表示各个研究领域的差异。对象表示子系统内更为具体的实体或概念。要素表示对象的基本构成部分，更详细地描述对象的组成和特性。属性则详细描述要素的特征和内容。

表 4　《地球表层系统科学数据叙词表》三级分类

一级	二级	三级
近地空间层	大气圈	大气物理
		大气化学
		气象气候
		天气
		大气环境
地表覆被层	水圈	海洋
		极地
		冰川冻土
		地表水
		地下水
		水化学
	人类圈	自然地理
		古地理
		人文地理
		资源环境
	生物圈	生态系统
		植物
		动物
		原生生物
		细菌
		真菌
		病毒
	土壤圈	土壤物理
		土壤化学
		土壤生物
		土壤地理
		土壤资源与环境
地表机制层	岩石圈	大地测量
		岩石/矿物
		地磁
		地震
		地质构造
		地质灾害
		地热
		火山

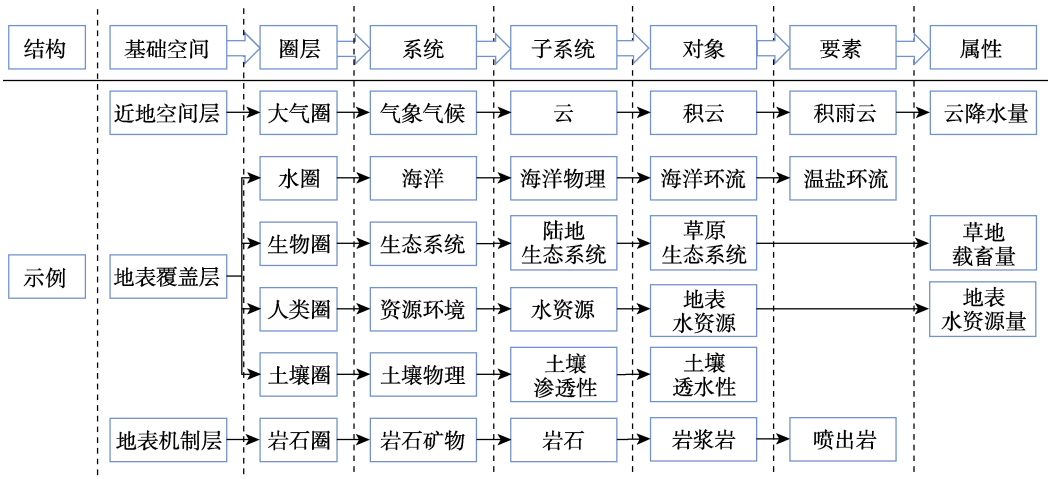

图 3 《地球表层系统科学数据叙词表》层次结构示例

3.4.4 叙词表主题词关系设计

ISO 25964[46]标准明确了叙词表中的三种基本语义关系，即等同关系（Equivalence Relation）、层级关系（Hierarchy Relation）和相关关系（Association Relation）[8]。

等同关系即等价关系，表示两个或多个可以相互替换的语义相同或相似的主题词，包括同义关系、缩写关系、名称演变关系等。同义关系表示不同的主题词具有相同或近似的含义，如“地壳运动”和“地质构造”是一组用来描述地质作用导致的地壳形变过程的同义词。缩写关系即主题词的简写或缩略形式与其完整形式之间的关系，如“CO_2”与“二氧化碳”。名称演变关系表示主题词的名称可能会随着认识的发展而发生改变，例如不同历史时期地名的更替。通过对主题词概念的分析，可以识别其中存在的同义关系。

层级关系即属分关系，表示主题词之间的上下位的级别关系，主要包含属种关系、整体与局部关系以及实体与实例关系[47]。属种关系指两个主题词之间呈父类和子类关系，如“根瘤菌”是“细菌”的一个子类。整体与局部关系指一个主题词是另一个主题词的一部分，如“北极”是“极地”的一部分。实体与实例关系指一个主题词表示某一实体，另一个主题词是该实体的一个实例，如“青藏高原”是“高原”的一个实例。通过构建主题词之间的层级关系，可以确保词表的清晰性和多层次性。

相关关系表示主题词之间存在不涉及等同关系或层级关系的某种关联，其种类复杂，主要关系如表 5 所示。

表 5 《地球表层系统科学数据叙词表》主要相关关系

关系	关系中文名	关系含义
hasImpactOn	影响关系	表示某主题词影响另一主题词
influencedBy		表示某主题词受另一主题词的影响
hasPossibleCause	因果关系	表示某主题词可能导致另一主题词
causedBy		表示某主题词受另一主题词而产生
measures	度量关系	表示某主题词度量另一主题词
measuredBy		表示某主题词被另一主题词度量
inverseOf	反比关系	表示某主题词与另一主题词成反比关系

4　数据结果与验证

4.1　数据集组成

《地表科学数据叙词表》包括三个部分：（1）《地表科学数据叙词表主表》（.xlsx）包括中英文主题词名称、同义词、关系、定义以及数据源等信息；（2）《地表科学数据叙词表分类表（中文版）》（.xlsx）包括中文主题词分类信息与数据源信息；（3）《地表科学数据叙词表分类表（英文版）》（.xlsx）包括英文主题词分类信息与数据源信息。其中，各字段及其描述如表 6 所示。

表 6　《地球表层系统科学数据叙词表》字段表

条目	描述
Keyword	英文主题词名称
ChineseName	中文主题词名称
AltLabel	英文主题词同义词
ChineseAltLabel	中文主题词同义词
SubClassOf	主题词所属父类
OnProperty	主题词关系
SomeValuesFrom	主题词关系作用对象
Comment	英文主题词定义
ChineseComment	中文主题词定义
Source	主题词来源

4.2　数据结果

《地表科学数据叙词表》将主题词划分为 7 级树状层次结构，共计 3,463 个主题词。在 3 大核心主题下，涵盖 6 个二级类和 35 个三级类，分别对应词表层次结构中的基础空间、圈层和系统。在三级类下，一个主题词可以隶属于多个子类，共计包含 166 个子系统主题词，589 个对象主题词，2,480 个要素主题词及 532 个属性主题词。各圈层的主题词数量分布情况如图 4 所示，水圈、人类圈、大气圈占主导地位，而土壤圈的主题词数量相对较少。图 5 展示了各数据源中所参考的主题词数量分布情况。

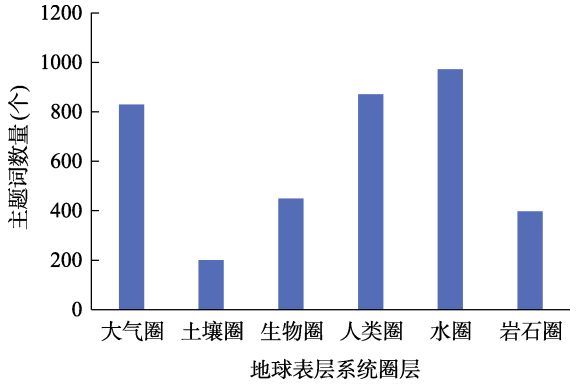

图 4　《地球表层系统科学数据叙词表》主题词数量圈层分布图

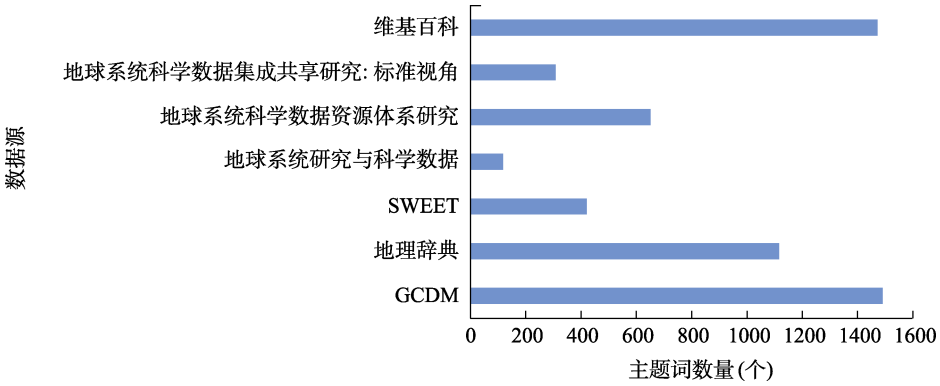

图 5 《地球表层系统科学数据叙词表》数据源分布情况

（注： SWEET 表示地球与环境术语语义网，即 Semantic Web for Earth and Environmental Terminology 的缩写；GCMD 表示全球变化主目录，即 Global Change Master Directory 的缩写）

本体是对实体、属性及其关系的一种形式化的知识表示方式，能够更精确地描述叙词表中的术语和概念。通过构建《地表科学数据叙词表》的本体模型，可以更直观地表达主题词间的关系。本体中所有的类别按照《地表科学数据叙词表》中的概念对象构建，根据词表的层次结构进行组织，并对构建的类添加各自的属性及与其他类之间的关系。本体的属性包括共有属性和数据属性。共有属性是多个类别之间共享的属性，可以被多个不同的类别所拥有或继承，以表示它们具有相似特征或共同属性。数据属性用于描述概念或实体的基本特征或属性，主要包括名称、定义、唯一标识编码、数据源等。本体的关系由词表中主题词间的语义关系建立而成。最终得到的地表科学数据叙词表本体模型可视化结果（前三级）如图 6 所示。

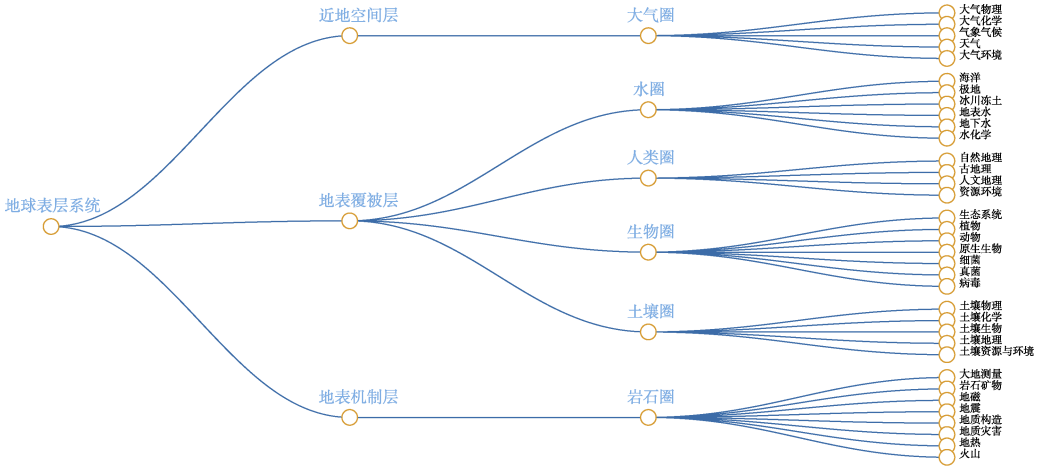

图 6 地球表层系统科学数据叙词表本体模型可视化结果（前三级）

4.3 数据结果验证

为验证和分析《地表科学数据叙词表》的词表规模与词表性能，本文参考相关研究的方法和经验，将其与《地质学汉语叙词表（Chinese Thesaurus of Geology）》[48]、《全球

变化主目录地球科学数据关键字（Global Change Master Directory Earth Science Data keywords）》[49]和《文物数字化保护主题词表（Cultural Relics Digital Protection Thesaurus）》[50]进行比较，以更好地支持地球科学的研究工作。

4.3.1　词表规模分析

词表规模是指词表所涵盖的词汇量，是评估词表对领域知识的覆盖程度的关键指标。《地表科学数据叙词表》《地质学汉语叙词表》《全球变化主目录地球科学数据关键字》和《文物数字化保护主题词表》的词表规模如表 7 所示。其中，正式主题词指经过筛选得到的代表核心主题的词汇；非正式主题词是与正式主题词的语义相同或相似的词汇；分项代表下位主题词，即在某个更广泛的主题下，更具体或更细分的概念；属项代表上位主题词，即包含或概括了多个下位主题词的更广泛或更一般的概念；参项指具备相关关系的主题词，它们可能在不同的上下文中与主题词有联系，或者在概念上与主题词有交叉。

表 7　《地球表层系统科学数据叙词表》《地质学汉语叙词表》《全球变化主目录地球科学数据关键字》和《文物数字化保护主题词表》词表规模比较表

主题词表	收词量	正式主题词数	正式主题词占比/%	非正式主题词数	分项数	属项数	参项数
地质学汉语叙词表	10,510	8,572	81.56	1,938	\	\	\
全球变化主目录地球科学数据关键字	1,556	1,556	100	0	1,541	254	0
文物数字化保护主题词表	2,605	2,468	94.74	137	407	1,648	182
地球表层系统科学数据叙词表	3,463	3,130	90.38	333	3,460	979	354

对比可知，在词表规模方面，《地表科学数据叙词表》处于中等规模，表明其具有良好的词汇覆盖质量。此外，其分项数、属项数和参项数也呈现出此特征。

4.3.2　主题词性能分析

《情报检索词汇控制》中指出词表的性能指标主要包括等同率、关联比、参照度及先祖度等[51]。等同率指非正式主题词数与正式主题词数的比例，较高的等同率有助于提高词表的检索效果，而较低的等同率表明词表更侧重于核心概念的精确表达。关联比和参照度则用于衡量词汇之间的关联程度。具体而言，关联比指具有语义关系的主题词数与正式主题词数之间的比例。参照度分为属分参照度、参项参照度以及总参照度。属分参照度指具有属分关系的词占正式主题词的比例，反映了词表在分类和层级结构上的清晰度。参项参照度指具有相关关系的主题词占正式主题词的比例，反映了词汇之间横向联系的丰富性。总参照度是属分参照度与参项参照度之和，它提供了一个综合的度量，反映了词表中词汇关系的丰富性和复杂性。《地表科学数据叙词表》《地质学汉语叙词表》《全球变化主目录地球科学数据关键字》和《文物数字化保护主题词表》的上述性能指标如表 8 所示。

由表 8 可知，四个词表的等同率普遍偏低，这可能表明这些词表在提供同义词或近似词方面不够丰富，从而在一定程度上限制了检索的广度和深度。特别是《全球变化主目录地球科学数据关键字》，其等同率为 0，这可能意味着该词表没有包含非正式主题词，或者其检索系统不区分正式和非正式主题词，这可能会对用户检索的灵活性和准确性造成影响。

表 8 《地球表层系统科学数据叙词表》《地质学汉语叙词表》《全球变化主目录地球科学数据关键字》和《文物数字化保护主题词表》主题词性能比较

主题词表	等同率	关联比	属分参照度	参项参照度	总参照度
地质学汉语叙词表	0.226	0.813	0.850	1.530	2.380
全球变化主目录地球科学数据关键字	0	1.000	1.154	0	1.154
文物数字化保护主题词表	0.053	0.746	0.789	0.070	0.859
地球表层系统科学数据叙词表	0.106	1.000	1.418	0.110	1.528

尽管等同率普遍偏低，但《全球变化主目录地球科学数据关键字》和《地表科学数据叙词表》的关联比达到了 1.000，显示出这两个词表在词汇之间的关联度较高，能够为每个主题词提供多个相关联的词汇，这有助于提高检索的深度和准确性。相比之下,《地质学汉语叙词表》和《文物数字化保护主题词表》的关联比虽然低于 1，但仍然显示出了一定的词汇关联性，表明它们在词汇关联方面也有一定的优势。属分参照度和参项参照度提供了词表内部结构的视角。《地表科学数据叙词表》在属分参照度上表现突出，说明该词表在词汇的层级结构和分类上具有较高的清晰度和组织性，有助于用户更好地理解主题词之间的关系。《地质学汉语叙词表》在参项参照度上表现突出，显示出该词表在提供词汇间的横向联系和多样性方面做得很好，这有助于丰富用户的检索视角和增加检索的覆盖面。总参照度综合了属分参照度和参项参照度,反映了词表中词汇关系的全面性。《地质学汉语叙词表》和《地表科学数据叙词表》的总参照度较高，这表明它们在词汇关系的构建上相对较优，有助于提供更全面的检索结果。

5 讨论和总结

随着对地球表层的深入认识，地表系统科学数据已成为不可或缺的科技资源。本文在明确地表系统内涵与范围的基础上，综合自顶向下和自底向上的方式构建《地表科学数据叙词表》。词表涵盖了大气圈、水圈、生物圈、人类圈、土壤圈和岩石圈在内的各种要素，整合了领域权威词典和综合性在线资源等数据源，将词汇划分为 3 个一级类、6 个二级类和 35 个三级类，共计 3,463 个主题词，为地球科学领域的数据管理和知识共享提供了有力的基础数据支持。未来研究将围绕《地表科学数据叙词表》这一核心成果，进行常态化更新与系列化的应用分析，主要有以下几方面：

（1）词汇扩展与自动更新：进一步拓展词汇的广度和深度，建立自动更新机制，定期整合与地球表层系统相关的新兴领域和跨学科领域的最新科学研究成果和领域知识，确保主题词表的时效性和新颖性，针对性地进行主题词的挖掘补全工作，以完善主题词类别分布。

（2）丰富语义关联：通过引入先进的深度学习和自然语言理解技术，增强主题词之间的相关性，实现更准确和丰富的语义关联，进一步提高主题词表的可用性和有效性。

（3）多样化应用：将词表扩展应用到更多的领域，包括教育、环境保护、灾害管理等，推动地表系统科学数据的广泛应用，为社会发展和跨学科合作提供更大支持。

综上所述,《地表科学数据叙词表》在词汇关联比和属分参照度方面表现较为出色，说

明其具有更加健壮的词汇层次关系以及更加稠密的词汇关联密度，能够较好反应地表系统间复杂的概念间关系。但是,《地表科学数据叙词表》在等同关系和相关关系方面的性能相对较低，使其等同率和参项参照度表现略显不足。因此，词表仍需要结合具体领域应用需求（如灾害应急数据共享等）进行有侧重点的扩展，以进一步提高词表的检索效果。

作者分工：诸云强和段福洲对数据集的开发做了总体设计；陈锦采集和处理了地球表层系统科学数据叙词表构建数据源数据；王曙设计了整体模型；陈锦、王曙和王斌做了数据验证；陈锦撰写了数据论文；王曙审核了数据论文。

利益冲突声明：本研究不存在研究者以及与公开研究成果有关的利益冲突。

参考文献

[1] 吴传钧. 论地理学的研究核心——人地关系地域系统[J]. 经济地理, 1991(3): 1–6.
[2] Phillips, J. D. Earth Surface Systems [M]. Oxford: Blackwell, 1999.
[3] 钱学森. 谈地理科学的内容及研究方法(在 1991 年 4 月 6 日中国地理学会“地理科学”讨论会上的发言)[J]. 地理学报, 1991(3): 257–265.
[4] 诸云强, 孙凯, 胡修棉等. 大规模地球科学知识图谱构建与共享应用框架研究与实践[J]. 地球信息科学学报, 2023, 25(6): 1215–1227.
[5] Li, X., Feng, M., Ran, Y., *et al*. Big Data in Earth system science and progress towards a digital twin [J]. *Nature Reviews Earth & Environment*, 2023, 4: 319–332.
[6] Knight, J., Harrison, S. The impacts of climate change on terrestrial Earth surface systems [J]. *Nature Climate Change*, 2013, 3(1): 24–29.
[7] Reichstein, M., Camps-Valls, G., Stevens B., *et al*. Deep learning and process understanding for data-driven Earth system science [J]. *Nature*, 2019, 566(7743): 195–204.
[8] Martínez-González, M. M., Alvite-Diez, M. L. Thesauri and semantic web: discussion of the evolution of thesauri toward their integration with the semantic web [J]. *IEEE Access*, 2019, 7: 153151–153170.
[9] 郭杨. 地理科学叙词表[M]. 北京: 科学出版社, 1995.
[10] 薛山顺, 周峰, 王春宁等. 基于主题词表的知识组织体系重构——以地学知识组织系统为例[J]. 国土资源信息化, 2020(3): 9–14.
[11] 史静. 地质学汉语叙词表[M]. 北京: 地质出版社, 2010.
[12] 《环境科学叙词表》编制组. 环境科学叙词表[M]. 北京: 中国环境科学出版社, 1989.
[13] 国家图书馆《中国图书馆分类法》编辑委员会. 中国分类主题词表[M]. 北京: 国家图书馆出版社, 2017.
[14] Timmer, R. C., Mark, M., Khoo, F. S., *et al*. NASA science mission directorate knowledge graph discovery [Z]. Companion Proceedings of the ACM Web Conference 2023. Austin, TX, USA; Association for Computing Machinery, 2023: 795–799. DOI: 10.1145/3543873.3587585.
[15] 陈锦, 王曙, 诸云强等. 地球表层系统科学数据叙词表[J/DB/OL]. 全球变化数据仓储电子杂志，2024. https://doi.org/10.3974/geodb.2024.07.10.V1. https://cstr.escience.org.cn/CSTR:20146.11.2024.07.10.V1.
[16] 全球变化科学研究数据出版系统. 全球变化科学研究数据共享政策[OL]. https://doi.org/10.3974/dp.policy.2014.05 (2017 年更新).
[17] 周俊. “地球表层”再讨论[J]. 自然灾害学报, 2004(6): 1–7.
[18] 谢家泽. 关于地球表层系统观的几个问题[J]. 地球科学进展, 1995(5): 432–435.
[19] Ramos, V. A. Hans Keidel and Alexander du Toit’s relationship and its impact on Wegener’s continental drift hypothesis [J]. *Geological Society, London, Special Publications*, 2023, 531(1): SP531-2022-2181.
[20] Conder, J. A. An active role for the ocean in seafloor spreading [Z]. American Geophysical Union Fall Meeting 2022, Chicago, American Geophysical Union, 2022: T26B-06
[21] Zheng, Y. F. Plate tectonics in the twenty-first century [J]. *Science China Earth Sciences*, 2023, 66(1): 1–40.
[22] Pausas, J. G., Bond, W. J. Feedbacks in ecology and evolution [J]. *Trends in Ecology & Evolution*, 2022,

37(8): 637–644.
[23] 黄秉维. 区域持续发展的理论基础——陆地系统科学[J]. 地理学报, 1996(5): 445–453.
[24] 王成善, 曹珂, 黄永建. 沉积记录与白垩纪地球表层系统变化[J]. 地学前缘, 2009, 16(5): 1–14.
[25] Steffen, W., Richardson, K., Rockström, J., *et al*. The emergence and evolution of Earth System Science [J]. *Nature Reviews Earth & Environment*, 2020, 1(1): 54–63.
[26] 杨顺华, 宋效东, 吴华勇等. 地球关键带研究评述: 现状与展望[J]. 土壤学报, 2023: 1–14.
[27] Jin, Z., Wang, X., Wang, H., *et al*. Organic carbon cycling and black shale deposition: an Earth System Science perspective [J]. *National Science Review*, 2023, 10: nwad243.
[28] 窦学成. 关于地球表层空间的本体模态构成[J]. 开发研究, 1998(1): 50–51.
[29] 浦汉昕. 地球表层的系统与进化[J]. 自然杂志, 1983(2): 126–128.
[30] 陆大道. 地球表层系统研究与地理学理论发展[Z]. 纪念中国地理学会成立九十周年学术会议, 北京. 中国地理学会. 1999: 8–13
[31] Phillips, J. D. Global and local factors in earth surface systems [J]. *Ecological Modelling*, 2002, 149(3): 257–272.
[32] 张猛刚, 雷祥义. 地球表层系统浅论[J]. 西北地质, 2005(2): 99–101.
[33] 李晓亮, 吴克宁, 冯喆等. 陆地表层系统分类研究进展——从土地类型到地球关键带类型[J]. 地理科学进展, 2022, 41(3): 531–542.
[34] Chen, M., Qian, Z., Boers, N., *et al*. Iterative integration of deep learning in hybrid Earth surface system modelling [J]. *Nature Reviews Earth & Environment*, 2023, 4(8): 568–581.
[35] 马宗晋, 高祥林, 杜品仁. 全球表层系统研究的思考[J]. 地学前缘, 2006(6): 96–101.
[36] Parsons, M. A., Duerr, R., Godøy, Ø. The evolution of a geoscience standard: an instructive tale of science keyword development and adoption [J]. *Geoscience Frontiers*, 2023, 14(5): 101400.
[37] 谭见安. 地理辞典[M]. 北京: 化学工业出版社, 2008.
[38] 孙久林. 地球系统研究与科学数据[M]. 北京: 科学出版社, 2009.
[39] 廖顺宝. 地球系统科学数据资源体系研究[M]. 北京: 科学出版社, 2010.
[40] 王卷乐. 地球系统科学数据集成共享研究: 标准视角 [M]. 北京: 气象出版社, 2015.
[41] Haribabu, S., Kumar, P. S. S., Padhy, S., *et al*. A novel approach for ontology focused inter-domain personalized search based on semantic set expansion [Z]. 2019 fifteenth international conference on information processing (ICINPRO). Bengaluru, India; *IEEE*, 2019: 1–5
[42] Whetzel, P. L., Noy, N. F., Shah, N. H., *et al*. BioPortal: enhanced functionality via new web services from the National Center for Biomedical Ontology to access and use ontologies in software applications [J]. *Nucleic Acids Research*, 2011, 39(suppl_2): W541–W545.
[43] 陈瑞, 曾建勋. 叙词表集成化体系及应用推进研究[J]. 情报学报, 2022, 41(4): 401–411.
[44] 王卷乐, 林海, 冉盈盈等. 面向数据共享的地球系统科学数据分类探讨[J]. 地球科学进展, 2014, 29(2): 265–267+273–274.
[45] 王卷乐, 王明明, 石蕾等. 科学数据管理态势及其对我国地球科学领域的启示[J]. 地球科学进展, 2019, 34(3): 306–315.
[46] ISO. ISO 25964-2:2013 Information and documentation-Thesauri and interoperability with other vocabularies Part 2:Interoperability with other vocabularies [EB/OL]. (2013-03-04) [2016-03-20]. http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=53658.
[47] 景雪芹. 基于维基百科分类体系的多语海洋叙词表构建[D]. 青岛: 中国海洋大学, 2016.
[48] 鲍秀林, 吴雯娜. 40 年来中文叙词表修订情况概览[J]. 图书情报工作, 2013, 57(2): 109–113.
[49] Global Change Master Directory (GCMD). GCMD Keywords, Version 17.3 [Z]. Greenbelt, MD: Earth Science Data and Information System, Earth Science Projects Division, Goddard Space Flight Center, NASA. 2023. URL (GCMD Keyword Forum Page). https://forum.earthdata.nasa.gov/app.php/tag/GCMD+Keywords.
[50] 罗威. 文物数字化保护主题词表的构建研究[D]. 北京: 北京化工大学, 2018.
[51] Hider, P. A survey of the coverage and methodologies of schemas and vocabularies used to describe information resources [J]. *Knowledge Organization*, 2015, 42: 154–163.