文章编号:1000－2367(2017)06－0001－08 DOI:10.16366/j.cnki.1000－2367.2017.06.001

地学数据本体研究与发展思考

诸云强[1,2,3],潘鹏[4,5],宋佳[1,2,3],侯志伟[6,1],
王东旭[7],孙凯[6,1],李威蓉[8],杨杰[6,1],王筱莹[1]

(1.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;
2.江苏省地理信息资源开发与利用协同创新中心,南京 210023;3.白洋淀流域生态保护与京津冀
可持续发展协同创新中心,河北 保定 071002;4.环境保护部环境工程评估中心,北京 100012;
5. 国家环境保护环境影响评价数值模拟重点实验室,北京 100012;6.中国科学院大学 资源与环境学院,
北京 100049;7.北京吉威时代软件股份有限公司,北京 100043;8.山东理工大学 建筑工程学院,山东 淄博 255000)

摘 要:分散、多源、异构的地学数据资源的整合集成、交换共享、挖掘利用必须以本体为支撑解决语义异构问题.针对当前缺乏以数据为核心能够支撑数据资源全生命周期操作处理本体库的问题,提出并开展了地学数据本体的研究与构建实践,并进一步分析了地学数据本体的应用领域和重点发展方向.地学数据本体的实质就是地学数据资源各类特征的本体,通过对地学数据资源内容、时间、空间、形态、来源等方面的概念、实例及其关系的形式化描述,实现无歧义的语义表达与推理,可有效支撑地学数据资源的分类编码、集成建库、语义搜索、关联数据和挖掘分析等应用.未来地学数据本体应进一步加强自动构建更新方法的研究,加快完善地学数据本体库,大力推动地学数据本体的应用.

关键词:地球科学;数据本体;分类集成;共享服务;关联挖掘

中图分类号:P208;P628 **文献标志码:**A

大数据时代下,数据资源的作用与地位愈发突出和重要[1],特别是第四科研范式下的现代地球科学(以下简称"地学")研究,不仅依赖于学科领域数据资源,而且需要大量跨学科、跨区域、综合数据资源的支撑,涉及频繁的地学数据资源分类集成、存储管理、整合处理、交换共享、挖掘利用等.由于地学数据资源具有分散、多源、异构、海量等特点,为了能够保障上述地学数据资源全生命周期操作的一致性和准确性,迫切需要统一语义的支持.

本体作为领域共识概念及其关系的形式化表达,已经成为语义网的基础,自 20 世纪 80 年代引入到信息领域后得到了长足的进步和发展[2－7].

(1)本体认知和建模方法不断完善.本体是语义网的基础,是概念化的、明确的、共享的和形式化的领域内共同理解的知识,并对概念及其相互关系进行明确的规范化定义[8－10].本体可分为顶层本体、领域本体、任务本体和应用本体[11],通常采用三元组、五元组、六元组和七元组等进行建模[12－14].

(2)本体描述查询语言和构建工具不断发展.从资源描述框架(Resource Description Framework,RDF),RDF-S,DAML(DARPA Agent Markup Language),OIL(Ontology Inference Layer)到网络本体描

收稿日期:2017-09-10;**修回日期:**2017-09-25.

基金项目:国家自然科学基金(41771430;41631177);科技基础性工作专项重点项目(2013FY110900);贵州省公益性基础性地质工作项目(黔国土资地环函[2014]23 号);2016 年贵州省公益性基础性地质工作项目(黔国土资源函[2016]269 号).

作者简介:诸云强(1977－),男,江西广丰人,中国科学院地理科学与资源研究所研究员,博士,博士生导师,主要研究方向为地学数据共享关键技术,资源环境信息系统,E-mail:zhuyq@igsnrr.ac.cn.

通信作者:潘鹏(1985－),男,湖北武汉人,环境保护部环境工程评估中心助理研究员,博士,主要研究方向为资源环境信息集成与共享,E-mail:panpeng@acee.org.cn.

述语言(Web Ontology Language,OWL)和 SPARQL 查询语言.出现了南加利福尼亚大学的 Ontosaurus,英国开放大学研发的 WebOnto,德国卡尔鲁厄大学的 OntoEdit 以及斯坦福大学的 Protege 等一批本体构建工具软件[15].提出了基于关系数据库、形式概念分析、概念格构造以及叙词表、Wiki 等网络语料抽取等一系列自动和半自动本体构建方法[16-18].

(3)本体知识库不断丰富.在自然语言理解和翻译领域,已经建立了叙词网(Wordnet)、框架网(Framenet)、通用上层模型(GUM)、常识知识库(Cyc)、中文知网(HowNet)等一系列通用的本体库及检索系统,以及地球与环境术语语义网(SWEET)、全球地理数据库(GeoNames)、时间本体(DAML)等地学领域知名的本体库.

(4)本体应用不断广泛和深化.通用本体知识库已经广泛应用于信息检索与智能推理、多语言翻译、知识管理与数据挖掘等领域中.地学本体也在多源、异构地学数据资源的分类集成[19-20]、组织管理[21-22]、检索发现[23-28]以及地理空间模型数据匹配、数据关联[29-34]等方面得到了大量的应用.

尽管已有的本体研究及其应用已经取得了丰硕的成果,但对比地学数据资源全生命周期操作对语义的需求,存在以下两个主要问题:(1)已有的本体库主要是通用、常用的词典库、地名库或学科术语库,到目前为止还缺乏以数据为核心,支撑数据资源全生命周期操作的本体库;(2)由于缺乏完善本体库的支撑,大部分地学本体应用,主要停留在理论方法探索、试验或小规模应用层面,未得到大范围的应用.为此,本文围绕地学数据资源分类集成、存储管理、整合处理、交换共享、挖掘利用等全生命周期操作对语义本体的需求,提出并研究地学数据本体,探讨地学数据本体的应用与发展.在数据本体的支撑下,有效解决分散、多源、异构地学数据资源整合集成与共享利用问题,同时推动和促进地学本体及其应用的发展.

1 地学数据本体模型与构建

1.1 地学数据本体模型

地学数据资源全生命周期操作处理的基础和依据主要是地学数据资源的各类特征,这些特征包括:数据内容、时间范围、空间范围、主题类别、数据类型、数据格式、数学基准、数据精度、属性语义、数据来源等内容[29-30].例如:数据分类主要依据数据内容、主题分类、数据类型等特征,数据整合主要依据数据类型、格式、属性语义等特征,数据发现主要依据数据内容、时间和空间范围等特征.因此,地学数据本体的实质就是建立地学数据资源各类特征的本体,即明确各类数据资源特征概念、属性及其实例以及相互间关系的形式化表达.

从描述、发现和使用数据资源过程中发挥的作用来看,地学数据资源特征可以分为本质特征、形态特征和来源特征三大类.地学数据资源本质特征是指标识数据资源唯一性的特征,通常包括:数据内容、时间和空间三要素;形态特征是指描述地学数据资源内在结构和外在形状的特征,通常包括:数据类型、格式、数学基准、数据精度、语言字符、属性语义等特征;来源特征是指表达数据源、数据采集、处理工具、模型方法等的特征.

从本体构建的角度,地学数据本体可以进一步分为基础本体、领域本体和应用本体.基础本体是指与领域无关的共性本体,包括:时间、空间、数学基准、语言字符等本体;领域本体是指地学领域内共识的本体,包括:主题内容、类型格式、数值语义、模型方法等本体;应用本体是指与具体应用相关的本体.

从本体功能的角度,地学数据本体由概念及其概念的属性、实例,概念之间、概念与实例、实例之间的关系,约束规则构成,即通常所说的本体五元组.

因此,地学数据本体模型可以表达为图 1 所示的三维模型.

1.2 地学数据本体总体架构

基于地学数据本体三维模型[35],地学数据本体顶层概念包括:本质本体、形态本体、来源本体.本质本体包括数据内容、时间范围、空间范围等;形态本体包括数据类型格式、数学基准、数据精度、属性语义、语言字符等;来源本体包括数据源、数据采集处理的模型方法与工具软件、责任者以及遵循的标准规范等.本质本体是地学数据本体的核心,形态本体依附本质本体,保障地学数据资源集成处理过程数据结构、基准、语义理解

的一致性，而来源本体同样依附本质本体，通过对地学数据资源原始的资料来源及其处理方法的描述，辅助数据资源集成处理过程中质量的控制.时间本体又包含时间基准（时间坐标系、时区、时间单位等）、地质时间、历史时间、现代时间等本体；空间本体又包含空间基准（平面坐标系、高程系、投影方式）、自然空间要素、人文空间要素等本体.限于篇幅，其他本体的下级子本体不再一一介绍.采用统一建模语言（Unified Modeling Language，UML）表达的地学数据本体总体架构如图 2 所示.

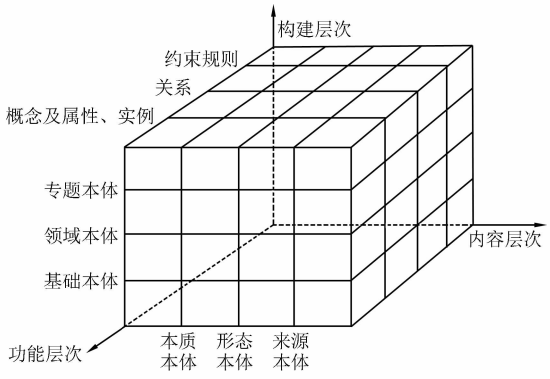

图 1　地学数据本体三维模型

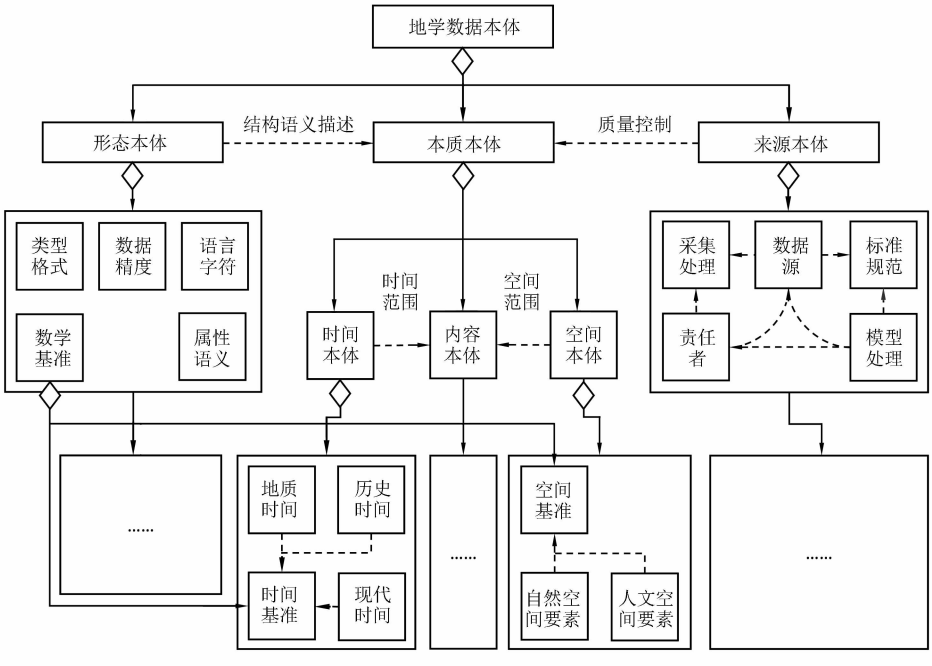

图 2　地学数据本体总体架构 UML 图

1.3　地学数据本体构建

为了便于本体的重用和集成，本文主要采用模块化设计方法，将每一个独立的本体构建成单独的 OWL（Web Ontology Language）文件，通过本体文件引用的方式实现相互之间的共享重用.目前已经建立形成了包含时间、空间、形态和来源的 200 个地学数据本体文件（如图 3 所示），共 14 279 个地学数据本体概念、实例及其它们相互间的关系（如图 4 所示）.具体实现时，利用 Protege 工具进行地学数据本体的构建.同时，为了提高地学数据本体存储管理、查询推理与应用的效率，采用 Jean TDB（Triple Database）作为本体持久化数据库管理工具，提供基于 HTTP 的 SPARQL（SPARQL Protocol and RDF Query）查询服务等.

2　地学数据本体应用

2.1　数据分类编码

数据分类编码是指根据数据资源特征，将数据资源进行区分和归类，并进行排序编码的过程.数据分类编码是数据资源组织、存储管理、查询检索、共享交换与互操作的基础.由于分类的目的和应用场景的不同，

数据资源往往采用不同的分类编码,而不同分类编码之间由于语义内涵的不一致,导致不同分类编码的数据资源很难进行统一的存储管理与共享交换.

图 3 地学数据本体文件

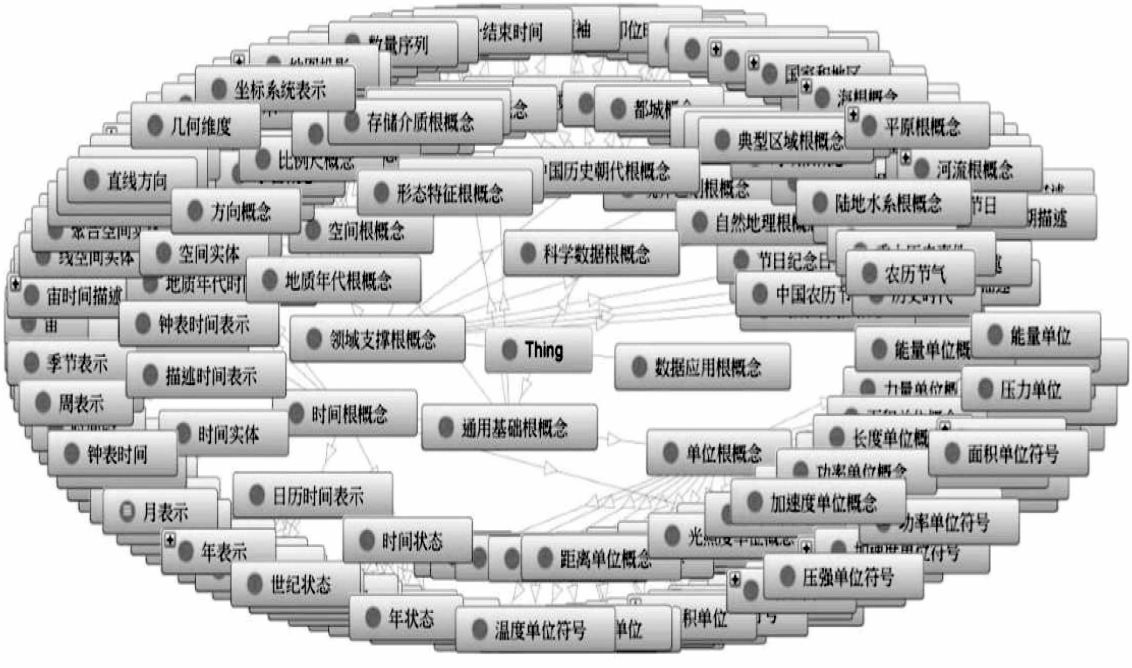

图 4 地学数据本体概念网络图

地学数据本体的实质就是建立地学数据资源各类特征的本体,通过规范化的形式对数据资源各类特征的概念、属性及其实例以及相互间关系进行明确的表达.为了解决传统数据资源分类编码存在的问题,可以依托地学数据本体,在数据分类时,选择数据资源对应的本体概念,如:数据内容分类时,将每个数据资源与

地学数据内容本体中规范化的概念关联起来即可；空间范围分类时，将每个数据资源与地学数据空间本体中规范化的概念关联起来即可.基于地学数据本体的数据分类编码可以很好地解决以下 3 个问题：一是利用本体概念之间的关系，可以自动、实时地实现数据资源的分类；二是由于本体概念语义的规范性与一致性，不同分类体系可以很好地进行映射和转换；三是数据资源与本体概念关联后，数据分类体系会随着本体库的更新而自动更新，不会影响数据的编码，从而保障数据资源在存储管理和共享利用中编码的唯一性.

2.2　数据集成建库

为了能够高效利用分散、多源、异构的地学数据资源，通常需要集成并建立地学数据资源库.传统的数据集成建库方法，首先利用 ER(实体关系)模型，设计统一的关系型数据库，然后利用 ETL(抽取、转换、装载)以及互操作等方法，将数据集成导入到数据库中.数据集成的难点在于不同来源数据的语义差异，而数据建库 ER 模型的核心则是实体的准确识别、抽取，实体属性的科学设计，以及实体关系的全面梳理.

地学数据本体可以很好地支持地学数据集成与建库工作.数据集成时，首先利用地学数据本体对异构数据资源进行语义标注，然后根据本体概念的语义关系，通过语义的映射和转换，在保持语义一致性的前提下，实现异构数据的统一集成.数据建库时，依据地学数据本体可以方便地抽取出科学、系统的 ER 模型：本体概念对应数据表实体、概念属性对应数据表字段、概念实例对应字典表记录或属性值域、本体关系可以直接转换为表间的关系，从而为数据建库奠定基础.

2.3　数据语义搜索

语义搜索是解决传统关键词匹配搜索“查不全、查不准”等问题的有效办法，已经被广泛应用于各类信息搜索中.语义搜索的基础是本体及其语义推理，其基本过程是首先利用本体对待查询的数据库进行语义标注，然后对用户搜索关键词进行语义扩展，再利用扩展后的搜索关键词与规范化语义标注后的数据库进行匹配查询，最终将查询结果按语义相关度进行优化排序.

因此，地学数据本体可以很好地支持和应用到地学数据的语义搜索中.如：利用地学数据空间本体，可以进行空间拓扑推理(如：雄安新区在空间上包含河北省的雄县、容城和安新)；利用地学数据形态本体，可以进行数据类型与格式实例的匹配判断(如：矢量类型包括 GML，ArcGIS，SuperMap，MapGIS 等格式)，从而实现地学数据的语义搜索.

2.4　关联数据

关联数据(Linked Data)被认为是语义网的一种实现，它通过明确的语义表达，使得不同领域、来源和结构的数据可以相互链接，从而促进数据的查找、集成与利用，为构建一个富含语义、人机都可理解的、互连互通的全球数据网络奠定基础[36−39].谷歌 2012 年启动的知识图谱就是典型的关联数据.它采用资源描述框架(Resource Description Framework，RDF)模型来表示数据，将内部信息资源都唯一地关联起来.如果查询词匹配到了谷歌知识图谱中的某个实体，谷歌就会以知识卡片的形式返回这个实体的属性以及与其他实体的关系[40].

关联数据的基础就是数据资源的特征语义，其价值则在于数据之间的关联关系.这恰恰是数据本体的核心所在.基于数据本体，可以对数据资源进行不同维度的语义标注，在此基础上，利用本体本身具有的概念与概念、概念与实例、实例与实例间的关系，就可以自动产生不同数据之间的关联关系[29].

2.5　数据挖掘分析

数据挖掘分析是指利用聚类、回归、机器学习等方法对数据资源的变化特征及隐藏的规律进行认识，对数据资源的发展趋势和异常特征进行预测和预警.地学数据挖掘分析通常需要用到不同来源、不同类型、不同部门和不同时间周期的数据资源，往往需要对这些数据资源进行预处理、融合或同化，包括：数据空间基准的统一、数据类型格式的转换、数据时空尺度的变换、数量单位、分类体系的统一、语义的一致性处理等.

只有在地学数据本体的支持下，才能准确理解数据资源各类特征的语义信息，为不同来源、类型、部门和时间的数据资源提供统一的语义基准，才能正确进行数据的预处理、融合、同化和挖掘分析.

3 地学数据本体发展思考

3.1 发展自动构建更新方法

本体构建方法可以分为三类:人工构建、半自动构建和自动构建.人工构建方法依靠领域专家的知识和群体智慧,通过手工的方式,分领域逐一确定领域内共识的本体概念、概念属性、实例及其相互间的关系.人工构建具有结果准确、权威等特点,但同时也存在费力、费时、构建周期漫长、更新困难等问题.最初的基础性字典、领域叙词表等大多都是通过人工构建方法完成的.半自动构建方法,主要基于已有的字典和知识库,通过抽取、转换等方式,先生成领域或应用本体,然后再通过领域专家对自动生成的本体进行检查和修改.半自动构建方法既降低了本体构建的复杂性,又可以较好地保证本体的质量,但它受限于所依赖的知识库大小和完善程度,仍然需要专家的参与,限制了本体的快速构建与更新.因此,在当今大数据时代下,如何充分利用网络上开放的语料库(如:维基百科、百度百科等),利用自然语言处理、深度学习、人工智能等技术,发展本体自动构建与更新方法,已经成为地学数据本体构建的迫切需求及研究热点与前沿.

3.2 完善地学数据本体库

本体自 20 世纪 80 年代被引入到信息领域,借助于完善的字典库,在信息检索、多语言翻译等领域取得了巨大的成功,但在地学领域一直没有得到广泛的、大规模的应用,其根本原因在于缺乏系统完善的、可实际应用的本体库.地学数据本体由本文作者提出后,尽管本文作者已经分别构建了时间、空间、形态等方面的数据本体[33,41-43],但作为地学本体中的新成员,地学数据本体的构建才刚刚开始.因此,必须充分整合和集成已有的地学本体库(如:SWEET,Geonames 等)以及网络上各种相关的开放信息资源,利用众包、众筹、志愿数据采集等模式,利用自动化构建方法,不断完善和丰富地学数据本体库.

3.3 推动地学数据本体应用

研究和构建地学数据本体的目的就是通过规范化的本体概念、实例及其相互间的关系,解决语义异构问题,支撑地学数据的分类集成、存储管理、智能检索、交换共享、数据关联和挖掘分析等应用.同时,通过本体的广泛应用,提升和改进地学数据本体构建的理论方法与技术体系,推进地学数据本体的发展.因此,应结合国家重大工程和计划,如:国家科技基础条件平台跨平台、跨类型科技资源的关联与共享,国务院要求全国各省/自治区开展的政务信息资源共享管理、政务信息系统整合共享等,大力推动地学数据本体的应用,不断丰富和完善地学数据本体.

4 结 语

大数据时代,地学数据资源作为一种战略资源已经引起各国政府、科技界和产业界的高度重视.分散、多源、异构的地学数据资源的分类集成、存储管理、整合处理、交换共享和挖掘利用,必须借助本体解决地学数据资源语义异构的问题.本文提出并开展了地学数据本体模型及其构建实践,在此基础上对地学数据本体的主要应用领域和发展方向进行了探索.

1)地学数据本体是地学领域本体重要的组成部分,其实质是地学数据资源各类特征的本体.地学数据本体模型,从内容上分本质、形态和来源三类本体,从功能上分概念、属性、实例、关系和约束规则五元组,从构建上分基础、领域和专题三层本体.

2)地学数据本体能够对地学数据资源全生命周期的概念、实例及其关系等进行明确的形式化描述,实现无歧义的语义表达与推理,从而可有效支撑地学数据资源的分类编码、集成建库、语义搜索、关联数据和挖掘分析等应用.

3)地学数据本体的研究才刚刚开始,未来还需要重点开展地学数据本体自动化构建与更新方法研究,加快构建形成全面系统、可实际应用的地学数据本体库,大力推进地学数据本体的应用,通过应用不断完善和丰富地学数据本体理论方法和知识库.

参 考 文 献

[1] 诸云强，朱琦，冯卓，等.科学大数据开放共享机制研究及其对环境信息共享的启示[J].中国环境管理，2015，7(6)：38-45.

[2] 孙敏，陈秀万，张飞舟.地理信息本体论[J].地理与地理信息科学，2004，20(3)：6-11.

[3] 李红梅，翟亮，朱煜.基于本体的地理空间实体类型语义相似度计算模型的研究[J].测绘科学，2009，34(2)：12-14.

[4] 李霖，朱海红，王红，等.基于形式本体的基础地理信息语义分析—以陆地水系要素类为例[J].测绘学报，2008，37(2)：230-235.

[5] 崔巍.用本体实现地理信息系统语义集成和互操作[D].武汉：武汉大学，2004.

[6] 黄茂军.地理本体的关键问题和应用研究[M].合肥：中国科学技术大学出版社，2006.

[7] 王磊，周宽久，仇鹏.领域本体自动构建研究[J].情报学报，2010，29(1)：45-52.

[8] Studer R，Benjamins V R，Fensel D.Knowledge engineering：principles and methods[J].Data & Knowledge Engineering，1998，25(1/2)：161-197.

[9] 邓志鸿，唐世渭，张铭，等.Ontology 研究综述[J].北京大学学报(自然科学版)，2002，38(5)：730-738.

[10] 陈建军，周成虎，王敬贵.地理本体的研究进展与分析[J].地学前缘，2006，13(3)：81-90.

[11] Guarino N.Understanding，Building and Using Ontologies[J].International Journal of IIuman Computer Studies，1997(46)：293-310.

[12] Perez A G，Benjamins V R.Overview of knowledge sharing and reuse components：Ontologies and problem-solving methods[EB/OL].[2017-08-23].https：//www.researchgate.net/publication/2505095.

[13] 安杨.基于本体的网络地理服务中的关键问题研究[D].武汉：武汉大学，2005.

[14] 黄茂军.地理本体的形式化表达机制及其在地图服务中的应用研究[D].武汉：武汉大学，2005.

[15] 徐国虎，许芳.本体构建工具的分析与比较[J].图书情报工作，2006，50(1)：44-48.

[16] 费静婷，顾君忠，杨静，等.基于 WordNet 和聚焦爬虫的半自动领域本体构建[J].计算机应用，2008，28：67-70.

[17] 刘丽斌，任瑞娟，米佳，等.基于叙词表构建本体的中文叙词词间关系细化研究[J].山东图书馆学刊，2010(1)：73-76.

[18] 陈琨，张蕾.基于知识图的领域本体构建方法.计算机应用，2011，31(6)：1664-1670.

[19] 何建邦，李新通，毕建涛，等.资源环境信息分类编码及其与地理本体关联的思考[J].地理信息世界，2003，1(5)：6-11.

[20] 王敬贵，杜云艳，苏奋振，等.基于地理本体的空间数据集成方法及其实现[J].地理研究，2009，28(3)：696-704.

[21] 李德仁，崔巍.地理本体与空间信息多级网格[J].测绘学报.2006，35(2)：143-148.

[22] 刘耀林，李红梅，杨淳惠.基于本体的土地利用数据综合研究[J].武汉大学学报(信息科学版)，2010，35(8)：883-886.

[23] 郭庆胜，杜晓初，闫卫阳.地理空间推理[M].北京：科学出版社.2006.

[24] 王欢，曹菡.基于本体和 SWRL 的空间关系的表示与推理方法[J].微电子学与计算机，2007，24(7)：166-168.

[25] 孙海霞，钱庆，成颖.基于本体的语义相似度计算方法研究综述[J].现代图书情报技术，2010，26(1)：51-56.

[26] 王强，王家耀，姜艳媛.本体支持的智能化空间信息服务发现[J].信息工程大学学报，2010，11(2)：170-174.

[27] 宋佳，王卷乐，诸云强，等.基于地理空间本体的语义检索相关度研究[J].计算机工程与应用，2011，47(5)：114-117.

[28] 魏勇，胡丹露，李响，等.基于 geonames 和 solr 的地名数据全文检索[J].测绘工程，2016，25(2)：28-32.

[29] Zhu Y，Zhu A，Song J，et al.Multidimensional and quantitative interlinking approach for Linked Geospatial Data[J].International Journal of Digital Earth，2017.DOI：10.1080/17538947.2016.1266041.

[30] Zhu Y，Zhu A X，Feng M，et al.A similarity-based automatic data recommendation approach for geographic models[J].International Journal of Geographical Information Science，2017，31(7)：1403-1424.DOI：10.1080/13658816.2017.1300805.

[31] 赵红伟，诸云强.地理空间元数据关联网络构建与应用[M].北京：电子工业出版社，2017.

[32] 赵红伟，诸云强，杨宏伟，等.地理空间数据本质特征语义相关度计算模型[J].地理研究，2016，35(1)：58-70.

[33] 赵红伟，诸云强，侯志伟，等.地理空间元数据关联网络的构建[J].地理科学，2016，36(8)：1180-1189.

[34] 罗侃，诸云强，程文芳，等.极地科学数据关联方法及应用研究[J].极地研究，2016，28(3)：361-369.

[35] 潘鹏.地理空间数据本体及其在数据发现中的应用研究[R].北京：中国科学院地理科学与资源研究所博士后报告，2015.

[36] Goodwin J，Dolbear II G.Geographical Linked Data：The Administrative Geography of Great Britain on the Semantic Web[J].Transactions in GIS，2008，12：19-30.

[37] 白海燕，朱礼军.关联数据的自动关联构建研究[J].现代图书情报技术，2010，26(2)：44-49.

[38] 郭少友.关联数据的动态链接维护研究[J].图书情报工作，2011，55(17)：112-116.

[39] 游毅，成全.基于关联数据的科研数据资源共享[J].情报杂志，2012，31(10)：146-151.

[40] 邹磊.知识图谱的数据应用和研究动态[J].中国计算机学会通讯，2017，13(8)：49-54.

[41] 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.

[42] 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.

[43] 孙凯，诸云强，潘鹏，等.形态本体及其在地理空间数据发现中的应用研究[J].地球信息科学学报，2016，18(8)：1011-1021.

Research and Development of Geoscience Data Ontology

Zhu Yunqiang[1,2,3], Pan Peng[4,5], Song Jia[1,2,3], Hou Zhiwei[6,1],
Wang Dongxu[7], Sun Kai[6,1], Li Weirong[8], Yang Jie[6,1], Wang Xiaoxuan[1]

(1. State Key Laboratory of Resources and Environment Information Systems; Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 2. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 3. Center for Collaborative Innovation in Baiyangdian River Basin Ecological Protection and Beijing-Tianjin-Hebei Sustainable Development, Baoding 071002, China; 4. Appraisal Center for Environmental and Engineering, Ministry of Environmental Protection, Beijing 100012, China; 5. State Environmental Protection Key Laboratory of Numerical Modeling for Environment Impact Assessment, Beijing 100012, China; 6. College of Resource and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; 7. Beijing GEOWAY Software Co., Ltd, Beijing 100043, China; 8. School of Civil and Architecture Engineering, Shandong University of Technology, Zibo 255000, China)

Abstract: The semantic heterogeneous problem must be solved based on ontology to integrate, exchange, share, and mining as well as the use of decentralized, multi-source and heterogeneous geoscience data resources. Aiming at the problem that existed ontology, which doesn't take geoscience data as the core and cannot support the data resource's whole lifecycle operation processing, this paper puts forward research and construction practice on geoscience data ontology, and further analyzes the application fields and key development directions of geoscience data ontology. The essence of geoscience data ontology is a normalized specification on geoscience data's various characteristics, including content, time, space, pattern, source and so on, which logically consists of class, instance and relationship between them. Based on geoscience data ontology, semantics of geoscience data can be expressed and reasoned unambiguously, which then can effectively support application of geoscience data, including classification and coding, data integration and database construction, accurate search , data linked, date mining and others. In future, Followingresearch ongeoscience data ontology should be carried out that include methods of automatic construction and updating, building a perfect geoscience ontology database and promoting the application of geoscience ontology.

Keywords: geoscience; data ontology; data integration; sharing services; data linking and mining

[责任编辑 陈留院]