

引用格式:侯志伟,诸云强,高楹,等.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,2018,20(1):17-27. [Hou Z W, Zhu Y Q, Gao Y, et al. Geologic time scale ontology and its applications in semantic retrieval[J]. Journal of Geo-information Science, 2018,20(1):17-27.] DOI:10.12082/dqxxkx.2018.170328

地质年代本体及其在语义检索中的应用

侯志伟^{1,2}, 诸云强^{1,4,5}, 高 楹³, 宋 佳^{1*}, 秦承志^{1,2,4}

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. 中国科学院大学, 北京 100049; 3. 首都师范大学燕都学院, 北京 100048; 4. 江苏省地理信息协同创新中心, 南京 210023; 5. 白洋淀流域生态保护与京津冀可持续发展协同创新中心, 保定 071002

Geologic Time Scale Ontology and Its Applications in Semantic Retrieval

HOU Zhiwei^{1,2}, ZHU Yunqiang^{1,4,5}, GAO Ying³, SONG Jia^{1*}, QIN Chengzhi^{1,2,4}

1. State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; 2. University of Chinese Academy of Sciences, Beijing 100049, China; 3. Honours College of Capital Normal University, Beijing 100048, China; 4. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China; 5. Baiyangdian Lake Ecological Protection and Sustainable Development of Jing-jin-ji Collaborative Innovation Center, Baoding 071002, China

Abstract: Different countries or organizations publish different versions or language editions of geologic time scales. Problems of ambiguity, such as synonymy or near-synonym and polysemy or other kinds of semantic heterogeneity arise in those geochronological concept systems, along with the lack of semantic linkage among different concepts and data, have hindered people from understanding and using those concepts accurately. Also, this caused insufficient results of integration and retrieval of geo-data for users' requirements. The ontology of geologic time proved to be an effective solution to these problems. However, current studies focus on international geologic time scales, and their content and formalizations are not fully applicable in China. In this paper, we present a new Geologic Time Scale Ontology which mainly focuses on Chinese geochronological and stratigraphic concept systems. It describes the formalizations of attributes, especially the temporal features and relations of those concepts. It adopts a modular method to build this ontology. In addition, we propose a design of a pilot system to study the utility of this ontology as the basis of a geoscience knowledge graph in knowledge retrieval. Furthermore, we implemented an approach of semantic geo-data retrieval in the pilot system which uses a hybrid strategy of fulltext and ontology-based search. Applications of knowledge graph and geo-data retrieval based on the abilities such as temporal reasoning of Geologic Time Scale Ontology, proved that the

收稿日期 2017-07-17;修回日期:2017-10-18.

基金项目:国家自然科学基金项目(41371381);国家自然科学基金重点项目(41631177);科技基础性工作专项(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);贵州省国土资源可持续发展战略研究(黔国土资资源函[2016]269号)。[**Foundation items:** Natural Science Foundation of China, No.41371381, 41631177; National Special Program on Basic Works for Science and Technology of China, No.2013FY110900; the Public and Basic Geological Project of Guizhou Province, China, No.[2014]23; GuiZhou Welfare and Basic Geological Research Program of China, No.201423, 2016269.]

作者简介: 侯志伟(1989-),男,湖南永兴人,博士生,研究方向为地学智能建模环境和地学数据共享。E-mail: houzw@lreis.ac.cn

***通讯作者:** 宋 佳(1980-),男,山西太原人,博士,助理研究员,研究方向为空间数据组织,地学数据共享和信息化。

E-mail: songj@lreis.ac.cn

proposed researches are effective in resolving semantic heterogeneity issues in geoscience knowledge and data. They can not only facilitate discovery of geoscience knowledge but also achieve the function of semantic data retrieval more intelligently, comprehensively and accurately for the discovery of specified and relevant data.

Key words: geological time; stratum; ontology; knowledge graph; semantic retrieval

***Corresponding author:** SONG Jia, E-mail: songji@lreis.ac.cn

摘要: 不同国家、语言和版本的地质年代概念体系中普遍存在的语义异构问题,以及概念之间、概念与数据之间缺少语义关联等问题,阻碍了概念的准确理解和应用,难以满足数据集成、检索的应用需求。地质年代本体是解决这些问题的有效方式。但现有研究主要针对国际地质年代概念体系,其内容和表达方式不完全适用于国内的概念体系。本文提出以中国地质年代和地层概念体系为主的地质年代本体,设计了本体中各概念属性信息,尤其是时间信息的语义表达方法。在此基础上采用模块化方法相应构建了本体,并设计实现了一个本体应用原型系统,提供了以地质年代本体知识库为基础的概念查询检索功能,以及融合全文检索技术和地质年代本体的数据语义检索功能。应用案例的初步结果表明,本研究构建的地质年代本体能为用户提供形式多样的知识查询服务,能较好地解决地学概念与数据中存在的语义异构、关联缺失问题,更为智能、完整、准确地完成地学数据的语义检索。

关键词: 地质年代;地层;本体;知识图谱;语义检索

1 引言

地质年代是表明地质事件发生时代及其相互关系的时间系统,是地球科学研究中最基本的要素之一^[1]。地质年代与时空上分布的地质事件、地层等都存在密切关系,对地学各领域、全球变化研究及地质信息化建设等都具有十分重要的理论和实际意义^[2-5]。然而,不同国家或组织所提出的地质年代等概念在定义、更新、应用等过程中都存在一词多义、异词同义等语义异质性问题^[6]。概念间复杂的、显性或隐性存在的时空、语义关联关系也缺少形式化表达等。这些问题严重阻碍了地学知识与数据的准确应用和相互关联与智能推荐、高效检索,是地学数据共享、地学信息化建设必须解决的重要问题之一^[7-9]。

本体是共享概念模型明确的形式化规范说

明^[10],用于确定领域内共享的词汇,并明确定义这些词汇和词汇间的关联关系^[11]。本体还具备语义推理能力,能根据规则自动建立词汇间的关联。本体的这些特性使其被广泛应用于地质年代相关领域知识、时空数据的形式化表达和语义关联应用,以满足用户对分布式多源异构、多语言知识和数据服务的查询检索、共享与互操作需求^[12-14]。知识图谱泛指各种大规模的、结构化的语义知识库,是一种揭示实体之间关系的语义网络,在智能搜索、个性化推荐等领域有重要的应用价值^[15]。本体对公理、规则和约束条件的支持能力被用于规范知识图谱中的实体及其联系^[16]。包含大量概念实体及各类关系的地质年代本体,可用于研究地质年代知识图谱的基本模式与应用。

国际上已对地质年代本体开展了较多的研究,代表性研究如表1所示。结合 Ma 等^[12]及 Cox 等^[17]

表1 部分代表性地质年代本体

Tab. 1 Some representative ontologies of geologic time scale

本体名称	文献	主要内容	年代关系表达	概念时间值表达	应用
SWEET	Raskin 等 ^[19]	国际地质年代	时间区间代数	“{百万年值} ^[8] MYA”对象及其数值属性	建立概念关联,促进地学数据的发现与使用
Geological Time & Dating Ontology	Perrin 等 ^[18]	国际地质年代、地层、地质事件	OWL-Time 本体	GeochronologicBoundary 与 GeochronologicInstant 对象及其数值属性	地质年代及地层层序的形式化描述与关联、定年
GTS ontology	Ma 等 ^[14]	国际地质年代、GSSP、多语言注释	gts:lowerThan 等关系谓词	gts:upperBoundaryTime、gts:lowerBoundaryTime 属性及其值	在线地质图地质年代信息标注、可视化
GTS ontology (isc {版本年})	Cox 等 ^[17]	多版本国际地质年代与地层、GSSP、多语言注释	THORS 本体	“{isc:Base 年代}Time”对象及其数值属性	地学专家的地质年代参考系统,地质年代调查、发展参考资料

的梳理可发现,这些本体在所表达的概念内容、时间关系和属性(起始和终止时间、地质年龄)表达方式方面都存在一定的共性。在内容方面,主要以多语言地质年代概念,全球标准层型剖面 and 点位(Global Standard Stratotype- Section and Points, GSSP)、地层等内容为主^[14, 17-18]。而在概念时间属性表达方面,为同时表达年代值与误差、单位等内容,多数本体使用了面向对象的表达方式,即将时间属性表达为一个概念(或称实例)。如 SWEET (Semantic Web for Earth and Environmental Terminology) 本体中的“251.0MYA”概念^[19]及 GTS 本体中的“isc:BaseMiddleJurassicTime”^[17]实例等。在地质年代时间关系表达方面,主要趋势是采用通用的时间拓扑模型。SWEET 本体采用了时间区间代数^[20], Perrin 等^[21]采用了 OWL-Time 本体,而 Cox 等^[4, 17]则采用实现了 ISO 19108:2002 时间拓扑模型和时间顺序参考系统模型 TORS (Temporal Ordinal Reference System Model) 的 THORS (Temporal Hierarchical Ordinal Reference System) 本体,通过 thors:nextEra 等关系谓词表示年代概念间的时间拓扑关系。在本体应用方面,主要关注点在于建立地质年代与地层知识库,克服概念语义歧义,建立地质年代相关概念与数据间的关联,促进知识发现、数据集成与查询检索、共享^[8, 12, 17-18, 22]。

国内目前对地质年代本体的研究较少。董少春等^[6]研究了地质时间本体的构建,探讨了地质时间本体在异构数据语义检索中的应用;吴立新等^[23]研究了沉积地层本体及其在数据集成与地层关系判断中的应用。二者在本体内容、多语言概念、时间关系的表达上都不完善,难以解决语义异质性问题和实现时间语义推理、地质年代与地层等的关联。中国古生物学与地层学数据库^[5]、基于地层时空拓扑关系的地质图数据自动检查系统^[24]等信息系统,尚不具备语义查询、推理等功能,无法充分发挥地质年代和地层信息的作用。综上所述,国际上以国际年代地层表为主要依据的地质年代本体研究目前已取得不少成果。但中国地质年代在概念定义和命名、时间段划分等方面与国际地质年代划分不完全一致。如《中国地层表》^[25]中的震旦纪在《国际年代地层表》^[26]中称为埃迪卡拉纪,“泥河湾期”被划分为 Calabrian (“卡拉布里雅期”)与 Gelasian (“杰拉期”或“格拉斯期”)。而国内相关研究较少,不能解决中国地质年代本身及其与国际地质年代概念之间的语义异构、地质年代与地层等概念的

语义关联缺失等问题。为此,通过分析地质年代概念及其特征,本文构建了以中国地质年代和地层划分等为主要内容的地质年代本体,对概念和时间信息的语义表达、本体构建方法等进行了设计。最后,以应用案例形式展示了该本体在地质年代知识查询和地学数据语义检索中的应用。

2 地质年代概念及其特征

2.1 地质年代与年代地层表

地质年代分为相对年代和绝对年龄(同位素年龄)。相对地质年代根据生物演化及岩石、地质构造等特征,把整个地质历史划分为一系列不同级别的历史阶段,即地质年代单位,借以展示岩石、地质事件等在时间上的顺序和包含关系。同位素年龄是指利用放射性元素的蜕变规律来测定的地质事件发生时刻距今(公元 1950 年)的单位时间值^[22, 27-28]。在表述地质年龄时,必须同时包含样品和测年方法等内容。因为对同一样品,采用不同的元素和方法测定的年龄及其误差是存在差异的。地质年代表(年代地层表)是地质年代的主要内容载体和表现形式。不同国家或组织的不同版本的地层表在年代单位命名和时间划分方面不尽相同。例如,2014 版《中国地层表》^[25]对部分界线的划分方案做了调整,同时对一些阶进行了重命名^[29]。

2.2 地质年代时间属性及其特征

根据侯志伟等^[30]的论述,时间实体可分为时间点与时间段两种类型,并具有时间坐标系、时间位置与距离、时间关系、时间方向等基本属性。这些时间属性对于地质年代时间同样适用。例如,万晓樵等^[1]指出,地质年代时间包含 3 个基本问题:① 地质事件是何时发生或结束的? ② 各事件是否同时发生? ③ 事件的持续时间和发生的速率如何? 问题①的实质是确定时间位置,问题②的实质是确定事件间的时间拓扑关系。问题③与时间距离相关,涉及时间的量测。时间的量测首先需要确定时间坐标系,即时间起算点、尺度基准(时间单位系统)和时间方向;然后确定时间点的坐标值;最后进行持续时间及发生速率的计算。在时间特征方面,地质年代时间也符合侯志伟等^[30]总结的地学数据时间的多尺度性、相对性、不确定性等特征。如不同级别地质年代单位的划分体现了其多尺度性和相

对性,地质年龄误差、地质年代划分及底界年龄的不一致等则体现了不确定性。

2.3 地质年代与其他领域概念关系

地质年代是地层划分和对比、古生物、构造地质、地球化学等研究的最基本要素,在科学研究及工程实践中都有重要应用^[1,6]。地层分为岩石地层、生物地层、事件地层、年代地层等不同的类型。其中,生物地层和事件地层研究的基础,在于确定古生物化石、地质事件的相对年代或绝对年龄。年代地层单位与地质年代单位一一对应,其他地层类型则存在穿时现象^[31]。年代地层还是划分和对比地层的最重要手段之一,任何其他相关地质学科如沉积环境分析、古地理重建等都首先要建立在时间对比准确的地层体基础之上^[5]。为了建立这样一套精确划分的地层(时间)框架,国际地层委员会推行 GSSP 来定义年代地层单位^[32]。地质年代信息也是地质图编绘、地质数据库建设中的主要内容之一。地质年代代码是地质图数据库中最重要代码,在计算机生成图件时,它决定着地质体的着色及地质符号的生成^[3]。

3 本体结构与语义表达

3.1 地质年代本体结构设计

为充分利用现有时间本体和地质年代研究成果,并解决中国与国际地质年代系统、地层学、古生物学等领域中与地质年代相关的语义异构和数据关联问题,为科学研究与工程应用提供丰富的知识服务,基于上述地质年代概念及特征的分析,设计了由基础层、核心层与应用层3个层次内容组成的地质年代本体,如图1所示。

基础层由领域无关的本体内容组成,主要内容

包括侯志伟等^[30]构建的定义了时间单位、时间坐标系、时间拓扑关系等基本时间概念、关系和公理的基础时间本体,用于知识共享和关联的通用数据模型 SKOS^① (Simple Knowledge Organization System) 及网络资源描述元数据标准都柏林核心元数据集术语 DCMI (Dublin Core Metadata Initiative) Metadata Terms^②。基础层的目标是为地质年代本体提供“建立与人类社会接轨的时间单位和地质学时间坐标”^[1,33],同时便于与其他类型的本体整合集成,扩展其应用范围。

核心层是地质年代本体的核心内容,是连接本体内部不同概念、实现概念相互关联的纽带和进行时态推理的基础。包含多语言的中国和国际地质年代单位概念及其属性和关系,还包括定义了精确地层时间框架,为地质年代提供了大量的点位地层和时空间信息的 GSSP。

应用层由多个地层模块组成,以年代地层等地质信息及其与地质年代的关联关系等内容为主。另外,由于地层、地质过程和地质事件、古生物等将地质年代概念与其他地学学科概念相关联,是地质年代本体面向更广泛的地学科研、应用服务的关键内容,因此也是本体的重要组成部分。

地质年代本体中各主要模块基本内容、代表性实例和主要属性或关系的简要说明如表2所示。

3.2 概念语义表达

本体的核心是领域内共享概念的形式化表达。将概念用一个词和一组结构化的属性组表示,是实现概念形式化表达的有效途径^[34]。属性是关于类成员的一般事实的断言以及个体的具体事实的断言,一个属性是一个二元关系^[35]。概念属性包括性质 (Property) 与关系 (Relation) 2 类。因而,本



图1 地质年代本体结构

Fig. 1 Architecture of the Geologic Time Scale Ontology (GTSO)

① <https://www.w3.org/TR/skos-reference/#L895>。

② <http://dublincore.org/documents/dcmi-terms/#H3>。

表2 地质年代本体模块定义

Tab. 2 Definitions of ontology modules of geologic time scale

模块名称	主要内容	主要概念类	主要属性/关系	概念实例示例
基础时间本体	时间类型、单位、方向;时间属性与关系、时间信息描述规则	时间单位,时间坐标系,百万年等	timeAfter、intAfter、intMeets、intContains	B.P.、逆时、百万年、145Ma
中国地质年代	中国地层表中确定的地质年代	宙、代、纪、世、期等	对应年代地层/对应国际地质年代	阳新世、周口店期
国际地质年代	国际地层委员会确定的地质年代	Eon、Era、Period、Epoch、Age等	对应年代地层、对应中国地质年代	Cambrian Period、Calabrian Age
全球界线层型剖面与点位	国际地层委员会确定的作为两个年代地层单位之间界线的定义和识别标准的专有标志点	按国际年代地层单位分类	地质年龄、经度、纬度、地理位置、综合图链接	Tortonian、Messinian
中国年代地层	在特定地质时间间隔内形成的岩石体。其顶底界线都是以等时面为界的	宇、界、系、统、阶等	标准RGB颜色、地质年龄、符号、岩性特征	第四系、更新统、周口店阶
国际年代地层	国际地层委员会确定的年代地层划分	Eonothem、Erathem、Series、System、Stage等	标准RGB颜色、地质年龄、符号、岩性特征	Guadalupian、Calabrian
中国岩石地层	根据地壳中岩石的特征和相互关系组织成的地层单位	群、组、段、层等	特点、正层型、分布	泰山岩群、上湖组、绒布寺冰碛层
中国事件地层	利用能在地层中留下某种印记并可被识别的较大范围分布的等时地质事件划分的对比地层	事件层、事件带、事件面及沉积事件、气候事件等	事件发生年代、事件发生区域、事件标志物	加里东运动、三叶虫首现、南极冰盖重大扩张
中国生物地层	将岩层根据地层中所含化石的特性编制成的若干地层单元	富集带、谱系带等;脊索动物门等	标准化石、包含生物种类	Claraia aurita富集带、三叶虫

体中概念C的语义定义可表达为:

$$\exists (term, P, R)[\forall C \equiv term \wedge R \wedge P] \quad (1)$$

式中: C 表示概念; $term$ 指概念术语; P 表示概念的性质; R 表示概念的关系。

概念的性质既可以是文本或数值,也可以是其他概念。前者在本体中使用数据属性(data property)或注释属性(annotation property)表达,如地层单位概念的层形剖面属性等;后者表达了概念之间的某种关联,使用对象属性(object property)表达,如地质年代单位对应的年代地层单位,年代地层单位对应的GSSP单位等。

依据上述定义,本文根据地层表等文献资料抽取了地质年代本体模型各主要概念术语及概念的属性和概念间关系,以实现地质年代本体各组成模块概念及其相互关系的形式化表达。各主要概念分类及其主要属性与关系如图2所示。

各地质年代单位相互之间通过来自基础时间本体的时间拓扑关系“intMeets”(时间段相接)等建立时态关联,并通过“cor.CCU”(“对应的中国年代地层单位”)等关系谓词建立与各地层单位的关联。同时,地层单位又通过“hasGSSP”与GSSP概念相关联。基于相互之间的语义关联,建立起各概念之间的语义关联网络。

版本信息、概念多语言注释(标注)也是地质年

代概念的重要属性。不同版本概念其定义、属性既有区别又有联系。本文采用DCMI中的“terms:hasVersion”来表示一个概念的版本信息,并定义关系“hasNewVersion”与“hasOldVersion”来表达概念的新旧版本,规定没有变化的概念不需要对URI中的版本信息做出更新。相对于GTS本体^[15]的版本表达方式,这种方式不仅便于对概念进行更新,同样便于对概念进行版本追溯和版本信息提取。对于概念多语言标注,本文借鉴Ma等^[8]的方法,使用SKOS中的“altLabel”与“prefLabel”来表示概念多语言标注,并使用“xml:lang”来表示语言类型。

3.3 时间信息语义表达

时间是地质年代本体概念的本质属性。地质年代本体中的时间语义表达包括两个方面,即时间属性的表达(如地质年龄)和时间拓扑关系的表达。

3.3.1 时间属性表达

地质年代单位起止时间、生物始现或绝灭时间等可认为是时间点;事件持续时间、旋回周期等可认为是时间段。因而时间属性的语义表达主要指地质年龄及其时间误差等点信息,和持续时间、间隔时间与重叠时间等段信息的属性值及相关的定年方法、地层剖面点位等信息的表达。使用简单的文字或数字难以表达如此丰富的时间属性内容。

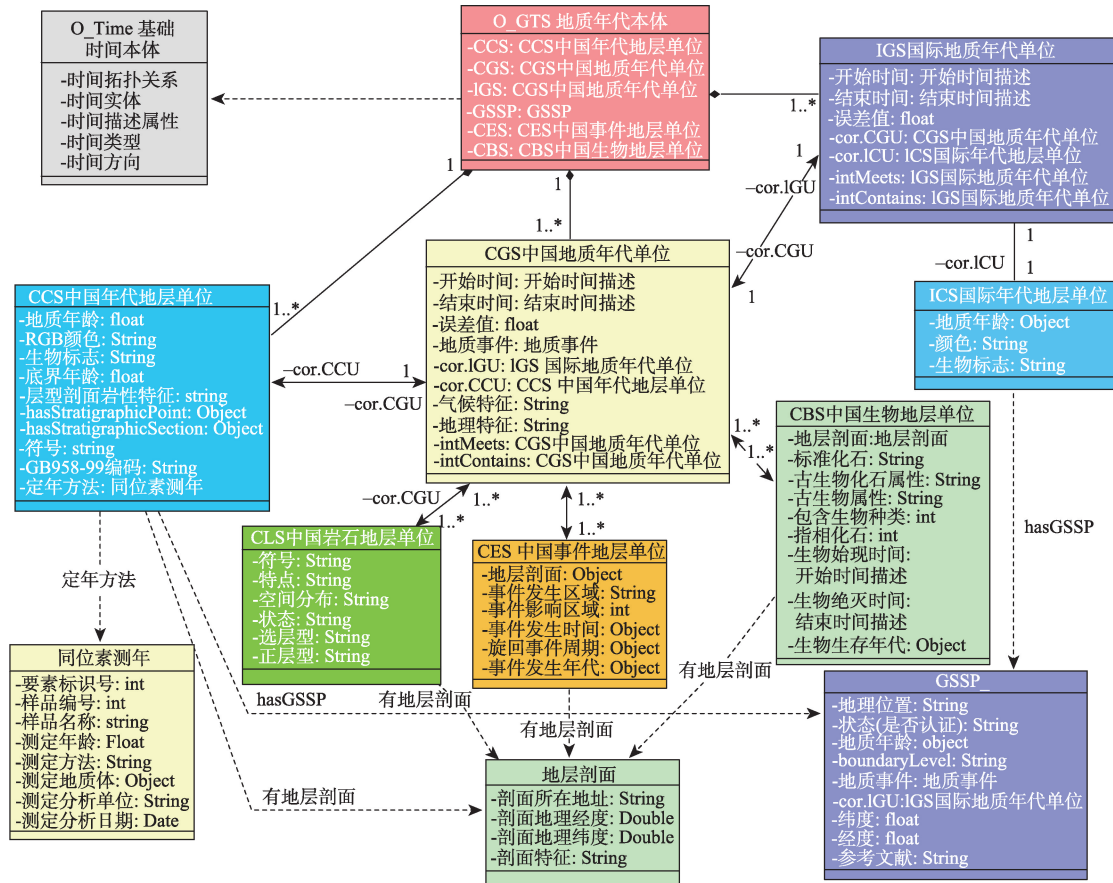


图2 地质年代本体模型概念语义表达

Fig. 2 Semantic representations of Geologic Time Scale Ontology modules

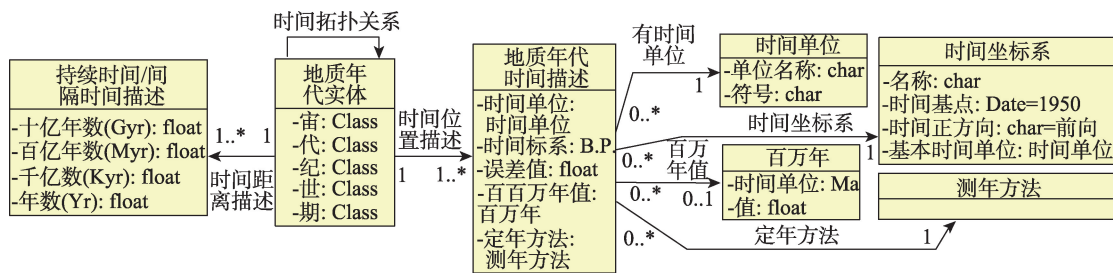


图3 地学数据对象时间属性语义表达

Fig. 3 Semantic representations of temporal properties of a geo-data object

为此,本文采用以时间描述对象为中心的RDF(Resource Description Framework)链实现时间属性的表达(图3)。1个地质年代概念实体可有多多个时间描述对象(如“开始时间”、“生物始现时间”),而每一时间描述对象可有时间坐标系类型(如B.P.(Before Present)等)和单位、测年方法等对象属性以及时间误差等数值属性。时间值对象通过“hasValue”等关系谓词来表达具体的时间值和时间单位属性。根据时间计算精度及应用需要,时间值对象及

持续时间等时间距离可以用一个时间粒度进行描述(如“1.806 Ma”),或采用不同粒度的组合进行描述(如“1 Ma, 806 Ka”)。

相比使用数据属性以字符串形式表达时间信息的方式,用时间概念对象表达时间属性这种多数地质年代本体采用的方法,虽然相对复杂,但更便于扩展和提取时间信息,可表达更丰富的时间语义。

3.3.2 时间拓扑关系表达

时间拓扑关系是实现时态推理的基础,其描述

地质现象或过程在时间上的相关关系,是时间语义关系的子集。地质年代单位之间的先后、包含关系,古生物种类与其始现与灭亡时间、生存年代的关系,地质事件与其发生年代的关系等都属于时间拓扑关系。侯志伟等^[30]基于时间区间代数扩展了一系列时间段、时间点之间的时间拓扑关系描述,主要包括相接、相离、相交、相等和包含5类关系及早于、晚于等衍生关系。本文采用其方法进行时间拓扑关系描述,具体内容包含于基础时间本体之中。

4 本体的构建与应用

4.1 地质年代本体构建

本体构建是本体由概念模型向工程应用过渡的关键步骤。本体的构建将产生一系列相关的 OWL 或 RDF 文件,便于分发和应用程序解析、应用。在借鉴吸收七步法^[36]、模块化本体构建方法^[37]及本体论开发步骤^[38]的基础上,本文采用了一种本体模

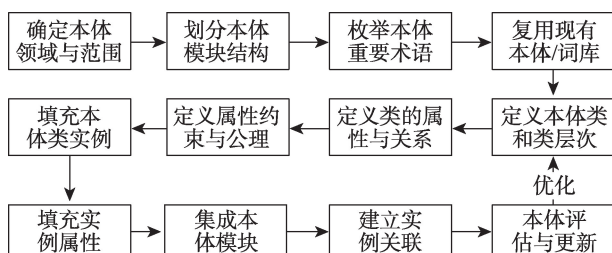


图4 本体构建流程

Fig. 4 Procedures of ontology construction

块化构建流程。多模块的划分使本体结构清晰、关系明确,便于复用、维护和团队分工协作,保证了本体构建的高效性和科学性。具体流程如图4所示。

本文以《中国地层表》和《国际年代地层表》的中、英、法、日等多个语言版本及 GSSP 表^③作为地质年代本体的重要术语来源,同时参照地层表说明等资料及表2、图2、图3中的本体设计,将本体划分为多个模块,并完成类的定义与类的层次结构划分。最终使用 Protégé 软件完成类及类属性、关系和实例、约束等内容的定义、生成和关联操作,初步构建完成了地质年代本体。目前共完成宙、代、宇、界等 120 个类,显生宙、新生代等 1988 个实例及几十种各类型对象属性与数据属性的构建,并为绝大多数属性定义了取值类型、个数与范围约束。本体中主要的类及其关系使用 Protégé 的 OntoGraf 插件展示如图 5 所示。

图5中,GTS为地质年代本体根类,CGS等为对应各个子模块的子类,具体含义及属性如图2所示。ICE、IGE、地质年代实体等类为概念实体类,其概念实例的时间描述信息由GTD与地质年代时间描述等辅助类中实例提供。

4.2 地质年代本体应用案例

4.2.1 地质年代概念语义查询

知识图谱的一个主要应用是概念语义查询。这对于增强用户对地质年代知识的理解,解决异词同义等语义异构问题,促进知识的有效利用有重要

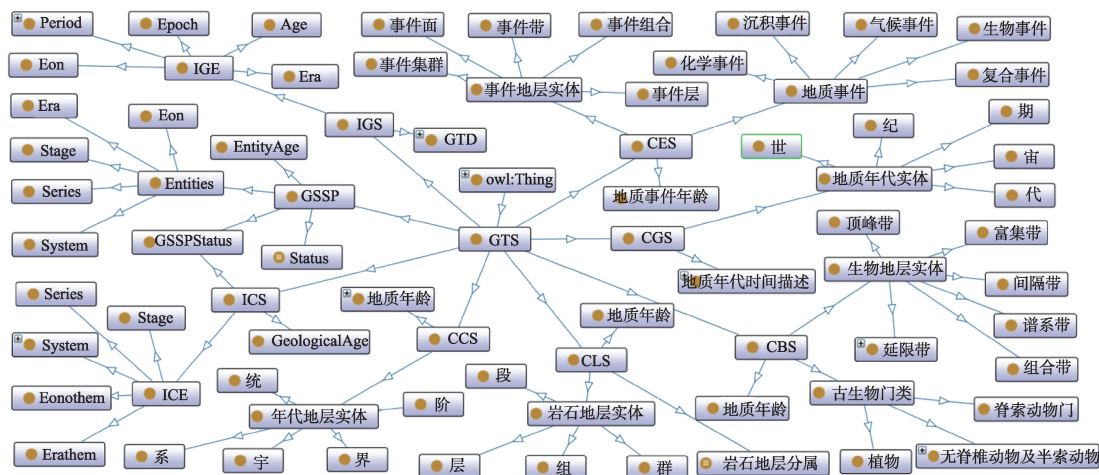


图5 地质年代本体类关系图

Fig. 5 Classes of GTSO built with Protégé

③ <https://engineering.purdue.edu/Stratigraphy/gssp/index.php?parentid=all>.

意义。以“新近纪”概念查询为例,系统先对术语“新近纪”进行语义解析,查询获得本体中对对应概念与定义、多语言标注等属性,并通过“owl:sameAs”和“owl:differentFrom”或“owl:AllDifferent”声明获取等价和非等价概念。然后,根据如式(2)、(3)所示规则(采用 Apache Jena^④规则语法)进行语义推理,获得起止时间属性和不同层级的地质年代、地层、地质事件、GSSP 等概念和概念属性。

$[(?a \text{ TimeRela:} \text{intContains } ?b)(?b \text{ TimeRela:} \text{intContains } ?c) \rightarrow (?a \text{ TimeRela:} \text{intContains } ?c)]$ (2)

$[(?a \text{ TimeRela:} \text{intContains } ?b)(?b \text{ CGS:有地质事件 } ?c) \rightarrow (?a \text{ CGS:有地质事件 } ?c)]$ (3)

查询结果经过转换最终以知识卡片的形式展现给用户。概念属性等结构化知识可以表格形式展现,

地质年代概念之间的层级关系则以动态树状图的方式进行展示,而地质年代关联的地层、地质事件和古生物、GSSP 等概念中包含的丰富空间信息,可结合 WebGIS 技术实现地理可视化。

4.2.2 地学数据语义检索

自然语言存在的语义歧义和数据间缺少关联的现实,使基于关键词匹配的数据检索功能难以满足日益复杂的应用需求^[39]。本文采用一种结合本体和全文检索技术的语义检索方法,基于本体对一词多义及概念间显性和隐性关联进行形式化处理和语义推理,经过语义转换、本体解析、推理、查询等过程对输入的检索条件进行响应,能获得比关键词检索更好的效果^[11]。本文从百度学术、地球系统科学数据共享平台、东亚古环境科学数据库等整理了200余条

查询类型: ☐ 年代值 ☐ 概念术语 ☐ 概念属性 ☒ 数据语义检索 ☐ SQL检索 ☐ 全文检索

600 Ma

结果列表 结果可视化

序号	标题
1	天津蓟县中新元古代沉积层序的初步研究——前寒武纪(1800~600Ma)一级层序划分及其与显生宙的一致性
2	中国东部及邻区晚前寒武纪(c.1850~600 Ma)古地理的再造
3	华北地台震旦纪—早古生代地震节律
4	天山古生代洋盆开启、闭合时限的岩石学约束——来自震旦纪、石炭纪火山岩的证据
5	震旦系地质年代学新研究
6	关于中国元古宙地质年代划分几个问题的讨论
7	新元古代岩浆活动与全球变化
8	新元古代超大陆构型中华南的位置

图6 地学数据语义检索结果

Fig. 6 Results of semantic geodata retrieval

地质年代相关中英文文献和数据集元数据进行试验。部分数据检索结果如图6所示。

由图6可发现,直接使用地质年代时间值进行查询时,本文实现的数据语义检索方法,不仅获得了字面上完全匹配的数据,还通过时态语义推理,获得了600 Ma所对应的地质年代单位概念震旦纪、新元古代、元古宙的相关数据。而传统的全文检索结果无法获得任何与时间值对应的地质年代概念相关的数据。显然地,语义检索所得结果更符合用户需求,说明本文的方法能够较好的理解用户搜索意图,具有一定的关联数据发现和数据推荐能力。由图7可发现,使用地质年代概念作为检索关键词时,在概念多语言注释的帮助下,可获得中文、日语、英

语等不同语言的数据结果。此外,在本体的支持下,还可同时检索获得“大湾阶”及等价的“大坪阶”概念的数据。术语变更已不再是人们获取数据的阻碍。由此可见,本文的方法解决了由于多语言和概念更新等造成的异词同义等语义异构问题,弥补了关键词检索在发现等价语义概念数据方面的不足。

5 结论

解决地质年代概念和地学数据中普遍存在的语义异构问题,建立各概念之间及概念与数据之间的语义关联,对于提升地学概念和数据的查询检索效果,促进数据集成、共享等都有重要作用。国内

④ <http://jena.apache.org/>

查询类型: <input type="radio"/> 年代值 <input type="radio"/> 概念术语 <input type="radio"/> 概念属性 <input checked="" type="radio"/> 数据语义检索 <input type="radio"/> SQL检索 <input type="radio"/> 全文检索	
寒武纪	
结果列表	结果可视化
序号	标题
1	闽北前寒武纪基底的地质年代学研究
2	华北克拉通前寒武纪超大陆旋回的基本制约
3	太行山-五台山区前寒武纪变质岩系同位素地质年代学研究
4	中国东部及邻区晚前寒武纪(c.1850-600 Ma)古地理的再造
5	天津蓟县中元古代沉积序列的初步研究——前寒武纪(1800~600Ma)一级层序划分及其与显生宙的一致性
6	カンブリア紀のジルコンU-Pb年代を示す南嶺長麓岩類について
7	A Cambrian peak in morphological variation within trilobite species.

图7 寒武纪语义查询结果

Fig. 7 Results of Cambrian Period semantic retrieve

外学者对地质年代本体的构建与应用研究,缺少对中国地质年代和地层划分的本体表达及其与国际地质年代和地层划分的关系的分析。为此,本文建立了由基础层、核心层和应用层组成的,以中国和国际地质年代与地层系统、GSSP等为主要内容的多模块地质年代本体。对地质年代概念和时间语义的表达方法进行了详细的设计。探索了地质年代本体在地学知识图谱和数据语义检索中的应用方法,设计开发了具有本体解析、存取、查询等功能的应用系统。实验结果表明,本研究建立的地质年代本体,建立了不同概念间的语义关联,较完整地表达了地质年代概念的语义信息,为快速而准确地发现地学研究与应用所需的地学知识与数据,提供了一种科学有效的方法。所提出的语义检索方法,不仅提高了检索效果,同时保证了检索工具的实用性和高效性,尤其对于领域专题数据库,具有一定的理论和应用价值。

本体的应用效果如何,很大程度上取决于本体的规模。因此,在后续研究中,需进一步完善地质年代本体结构和数据,补充地层剖面、古生物化石、地质图、古地理重建图等各类知识和数据,形成较完善的大规模的地质年代知识图谱。此外,还需对时间语义相似度、相关度的计算和数据语义标注方法等进行研究,提高数据检索等应用的准确度;而在应用方面,开展与各类地学数据共享平台的系统集成方法研究,向外开放知识查询服务接口,为社区提供知识服务。增加对概念属性信息的多语言标注,并建立与国际上现有地质年代本体的概念与关系映射,实现成果的国际化应用,也是后续需要研究的问题。

参考文献(References):

- [1] 万晓樵,王成善,吴怀春,等.从地层到地时[J].地学前缘, 2014,21(2):1-7. [Wan X Q, Wang C S, Wu H C, et al. From stratigraphy to Earthtime[J]. Earth Science Frontiers, 2014,21(2):1-7.]
- [2] 陈宣华,董树文,史静.地质年代学发展历史的简要回顾及前景[J].世界地质,2009,28(3):384-96. [Chen X H, Dong S W, Shi J. Brief review and prospect of geochronological development[J]. Global Geology, 2009,28(3):384-96.]
- [3] 郭宽廉,左群超,郭可筠.建立地质图数据库的地质年代代码问题[J].地质通报,2008,27(7):1103-6. [Wu K L, Zuo Q C, Wu K Y. Problems in the codes of geological ages in the process of establishment of the geological map database[J]. Geological Bulletin of China, 2008,27 (7):1103-1106.]
- [4] Cox S J, Richard S M. A formal model for the geologic time scale and global stratotype section and point, compatible with geospatial information transfer standards[J]. Geosphere, 2005,1(3):119-37.
- [5] 樊隽轩,张华,侯旭东,等.古生物学和地层学研究的量化趋势——GBDB 数字化科研平台的建设及其意义[J].古生物学报,2011,50(2):141-53. [Fan J X, Zhang H, Hou X D, et al. Quantitative research trends in Palaeobiology and Stratigraph - Construction of the GeoBiodiversity Database (GBDB) using e-science technology[J]. Acta Palaeontologica Sinica, 2011,50(2):141-153]
- [6] 董少春,尹宏伟,许刚.地质时间本体在异构数据检索中的应用[J].地球信息科学学报,2010,12(2):194-199. [Dong S C, Yin H W, Xu G. Application of geologic time ontology in heterogeneous data searching[J]. Journal of Geo-Information Science, 2010,12(2):194-199.]
- [7] 李学军.地学论文中常见的表达问题及解决方法和建议[J].中国科技期刊研究,2011,22(5):779-83. [Li X J. Solutions and suggestions for common expression problems

- in geoscience papers[J]. Chinese Journal of Scientific and Technical Periodicals, 2011,22(5):779-83.]
- [8] Ma X, Carranza E J M, Wu C, et al. A SKOS-based multilingual thesaurus of geological time scale for interoperability of online geological maps[J]. Computers & Geosciences, 2011,37(10):1602-1615.
- [9] 诸云强,孙九林.面向 e-GeoScience 的地学数据共享研究进展[J].地球科学进展,2006,21(3):286-90. [Zhu Y Q, Sun J L. The research progress in Geo-data sharing based on e-GeoScience[J]. Advances in Earth Sciences, 2006,21(3):286-290.]
- [10] Studer R, Benjamins V R, Fensel D. Knowledge engineering: Principles and methods[J]. Data & Knowledge Engineering, 1998,25(s1-2):161-197.
- [11] 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. [Deng Z H, Tang S W, Zhang M, et al. Overview of ontology[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2002,38(5):730-738.]
- [12] Ma X, Fox P. Recent progress on geologic time ontologies and considerations for future works[J]. Earth Science Informatics, 2013,6(1):31-46.
- [13] Richard S M, Boisvert E, Brodaric B, et al. GeoSciML - A GML Application for Geoscience Information Interchange[J]. Philadelphia Annual Meeting, 2007:47-59.
- [14] Ma X, Carranza E J M, Wu C, et al. Ontology-aided annotation, visualization, and generalization of geological time-scale information from online geological map services[J]. Computers & geosciences, 2012,40:107-119.
- [15] 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. [Xu Z L, Sheng Y P, He L R, et al. Review on techniques of knowledge graph [J]. Journal of University of Electronic Science and Technology of China, 2016,45(4):589-606.]
- [16] 刘峤,李杨,段宏,等.知识图谱构建技术综述[J].计算机研究与发展,2016,53(3):582-600. [Liu Q, Li Y, Duan H, et al. Knowledge graph construction techniques[J]. Journal of Computer Research and Development, 2016,53(3):582-600.]
- [17] Cox S J D, Richard S M. A geologic timescale ontology and service [J]. Earth Science Informatics, 2015,8(1):5-19.
- [18] Perrin M, Mastella L S, Morel O, et al. Geological time formalization: an improved formal model for describing time successions and their correlation[J]. Earth Science Informatics, 2011,4(2):81-96.
- [19] Raskin R G, Pan M J. Knowledge representation in the semantic web for Earth and environmental terminology (SWEET) [J]. Computers & Geosciences, 2005,31(9):1119-1125.
- [20] Allen J F. Maintaining knowledge about temporal intervals [J]. Communications of the ACM, 1983,26(11):832-843.
- [21] Hobbs J R, Pan F. Time Ontology in OWL. 2006, <http://www.w3.org/TR/2006/WD-owl-time-20060927/>.
- [22] Laxton J, Serrano J-J, Tellez-Arenas A. Geological applications using geospatial standards: An example from One-Geology-Europe and GeoSciML[J]. International Journal of Digital Earth, 2010,3(1):31-49.
- [23] 吴立新,徐磊,车德福.地层本体及其在大区域钻孔数据集成中的应用[J].武汉大学学报·信息科学版,2008,33(2):144-148. [Wu L X, Xu L, Che D F. Stratum-Ontology and its application in bore-hole data integration[J]. Geomatics and Information Science of Wuhan University, 2008,33(2):144-148.]
- [24] 熊义辉,汪新庆,李国庆.地层时空拓扑关系的计算机自动检查设计与实现[J].国土资源遥感,2016,28(1):197-202. [Xiong Y H, Wang X Q, Li G Q. Design and realization of computer automatic checking of the stratigraphic spatio-temporal topological relationships[J]. Remote Sensing for Land & Resources, 2016,28(1):197-202.]
- [25] 全国地层委员会.中国地层表[J].地球学报,2014,35(3):附图 1. [National Commission on Stratigraphy of China, The Stratigraphic Chart of China (Illustration 1)[J]. Acta Geoscientica Sinica, 2014,35(3): Attached map 1.]
- [26] Cohen K M, Finney S C, Gibbard P L, et al. The ICS International Chronostratigraphic Chart[J]. Episodes, 2013, 36(3):199-204.
- [27] 刘德良.地球科学中时间段的定年问题——一种不同于地球化学的地球物理定年方法[J].中国科学技术大学学报,2011,41(1):1-8. [Liu D L. Duration measurement in geological sciences : A geophysical approach[J]. Journal of University of Science and Technology of China, 2011,41(1):1-8.]
- [28] 柳成志,赵荣,赵利华.地球科学概论[M].北京:石油工业出版社,2006. [Liu C Z, Zhao R, Zhao L H. Introduction to earth science[M]. Beijing: Petroleum Industry Press, 2006.]
- [29] 王泽九,黄枝高,姚建新,等.中国地层表及说明书的特点与主要进展[J].地球学报,2014,35(3):271-276. [Wang Z J, Huang Z G, Yao J X, et al. Characteristics and main progress of the stratigraphic chart of China and directions [J]. Acta Geoscientica Sinica, 2014,35(3):271-276.]
- [30] 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. [Hou Z W, Zhu Y Q, Gao X, et al. Time-ontology and its application in geodata retrieval[J]. Journal of Geo-information Science, 2015,17(4):379-390.]
- [31] 龚一鸣,张克信.地层学基础与前沿[M].武汉:中国地质大学出版社,2007. [Gong Y M, Zhang K X. Basics and

- frontiers in stratigraphy[M]. Wuhan: China University of Geosciences Press, 2007.]
- [32] 彭善池. 全球标准层型剖面 and 点 (“金钉子”) 和中国的 “金钉子” 研究[J]. 地学前缘, 2014, 21(2): 8-26. [Peng S C. Global standard stratotype- section and points (GSSP, "Golden Spike") and the GSSP research in China[J]. Earth Science Frontiers, 2014, 21(2): 8-26.]
- [33] 龚一鸣, 史晓颖. 关于时间、沉积与地层学的几点思考[J]. 地学前缘, 2014, 21(2): 27-35. [Gong Y M, Shi X Y. Thinking on time, sedimentation and stratigraphy[J]. Earth Science Frontiers, 2014, 21(2): 27-35.]
- [34] 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——以陆地水系要素为例[J]. 测绘学报, 2008, 37(2): 230-235. [Li L, Zhu H H, Wang H, et al. Semantic analyses of the fundamental geographic information based on formal ontology: Exemplifying hydrological category[J]. Acta Geodaetica et Cartographica Sinica, 2008, 37(2): 230-235.]
- [35] 钟秀琴, 刘忠, 丁盘苹. 基于混合推理的知识库的构建及其应用研究[J]. 计算机学报, 2012, 35(4): 761-766. [Zhong X Q, Liu Z, Ding P P. Construction of knowledge based on hybrid reasoning and its application[J]. Chinese Journal of Computers, 2012, 35(4): 761-766.]
- [36] Noy N F, McGuinness D L. Ontology development 101: A guide to creating your first ontology[M]. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880. 2001.
- [37] 冯兰萍, 朱礼军, 蒋亚东. 一种模块化本体构建方法研究[J]. 现代图书情报技术, 2010, 26(6): 53-59. [Feng L P, Zhu L J, Jiang Y D. Study on a method of building modular onotlogy[J]. New Technology of Library and Information Service, 2010, 26(6): 53-59.]
- [38] 冯志勇, 李文杰, 李晓红. 本体论工程及其应用[M]. 北京: 清华大学出版社, 2007. [Feng Z Y, Li W J, Li X H. Ontology engineering and its applications[M]. Beijing: Tsinghua University Press, 2007.]
- [39] Lutz M, Klien E. Ontology-based retrieval of geographic information[J]. International Journal of Geographical Information Science, 2006, 20(3): 233-260.