

引文格式: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. DOI:10.11947/j. AGCS.2018.20180113.  
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. DOI: 10.11947/j. AGCS. 2018. 20180113.

# 多源异构数据的大规模地理知识图谱构建

蒋秉川<sup>1,2</sup>, 万 刚<sup>2</sup>, 许 剑<sup>1,3,4</sup>, 李 锋<sup>2</sup>, 温荟琦<sup>2</sup>

1. 地理信息工程国家重点实验室, 陕西 西安 710054; 2. 信息工程大学地理空间信息学院, 河南 郑州 450001; 3. 清华大学建筑学院, 北京 100084; 4. 西安测绘研究所, 陕西 西安 710054

## Geographic Knowledge Graph Building Extracted from Multi-sourced Heterogeneous Data

JIANG Bingchuan<sup>1,2</sup>, WAN Gang<sup>2</sup>, XU Jian<sup>1,3,4</sup>, LI Feng<sup>2</sup>, WEN Huiqi<sup>2</sup>

1. State Key Laboratory of Geo-information Engineering, Xi'an 710054, China; 2. Institute of Geospatial Information, Information Engineering University, Zhengzhou 450001, China; 3. School of Architecture, Tsinghua University, Beijing 100084, China; 4. Xi'an Research Institute of Surveying and Mapping, Xi'an 710054, China

**Abstract:** As a new generation of geographic language, virtual geographic environments(VGE) needs to construct the virtual geographic knowledge engineering through the analysis and excavation of multi-sourced heterogeneous spatio-temporal data. Through the transformation about “data-information-knowledge-wisdom”, it can promote the rapid conversion and integration of geographic knowledge for intelligent VGE system, which may provide a theoretical and technical support for intelligent processing of geographic information and intelligent service of geographic knowledge. Knowledge graph plays the role of a bridge between artificial intelligence and knowledge engineering for VGE system. This paper firstly reviewed the research status in knowledge graph and geographic knowledge graph from the domestic and abroad perspectives. Then, the construction process of geographic knowledge graph is proposed. Furthermore, the key technologies of geographic knowledge graph are discussed in depth. Finally, the application direction of geographic knowledge graph is discussed and stated. The study of geographic knowledge graph in this paper may help to realize the knowledgization of geographic information and promote the intelligent service level for VGE system.

**Key words:** virtual geographic environments; geographic knowledge; knowledge graph; geographic knowledge graph

**Foundation support:** The National Defense Science and Technology Foundation of China (Nos. 3601023; 3601020); The Science and Technology Program of Henan (No. 182102210147); The Open Fund of State Key Laboratory of Geo-information Engineering (No. SKLGIE2017-M-4-3); The Foundation for New Research Direction of PLA Information Engineering University (No. 210502020)

**摘 要:** 虚拟地理环境作为新一代的地理语言, 需要通过对多源异构时空数据的分析与挖掘, 构建虚拟地理环境知识工程, 实现“数据—信息—知识—智慧”的转化, 促进智能虚拟地理环境系统中地理知识的快速转换和融合, 从而为虚拟地理环境中地理信息的智能化处理、地理知识的智慧化服务提供支撑。知识图谱是人工智能与虚拟地理环境知识工程相结合的桥梁。本文系统评述了知识图谱、地理知识图谱的研究现状, 提出了地理知识图谱的构建流程, 重点研究了地理知识图谱构建的关键技术, 讨论和阐述了地理知识图谱的应用方向。本文的研究将有助于实现地理信息的知识化, 提高虚拟地理环境系统的智能化服务水平。

**关键词:** 虚拟地理环境; 地理知识; 知识图谱; 地理知识图谱

中图分类号:P208文献标识码:A文章编号:1001-1595(2018)08-1051-11

基金项目:国防科技基金(3601023;3601020);河南省科技攻关计划(182102210147);地理信息工程国家重点实验室开放研究基金(SKLGIE2017-M-4-3);信息工程大学校新兴科研方向培育基金(210502020)

当前地理科学数据丰富但知识匮乏,而智能化虚拟地理环境需要构建 VGE 知识工程<sup>[1]</sup>,智慧城市的建设也需要构建地理知识基础设施<sup>[2]</sup>(geographic knowledge infrastructure),地理信息服务必须实现从“数据—信息—知识—智慧”的智能化转化,构建面向大众的地理空间知识服务体系<sup>[3-4]</sup>。

虚拟地理环境作为新一代的地理语言,改变了传统的地理科学知识的表达与获取方式,加速了地理数据、信息到知识的转换,是协助人类认识地理世界和解决地理问题的新型地理分析工具<sup>[5-6]</sup>。VGE 知识工程是基于知识的智能虚拟地理环境系统的支撑理论、方法和技术体系,其关键问题是地理知识的表达与建模、知识库的构建和管理、地理知识的智慧化服务<sup>[1]</sup>。将人工智能与地理空间信息领域相结合,是解决虚拟地理环境智能化的重要手段。

地理空间知识的存储、共享和分类是虚拟地理环境系统通向智能服务的重要知识库支撑。在大数据、人工智能等技术的推动下,需要利用高性能计算、云计算、大数据可视分析、知识图谱等技术体系,实现对多源异构时空数据的分析与挖掘,从地理数据到地理知识的转变,为虚拟地理环境的智能化提供技术支撑。在人工智能领域,知识图谱、知识表示、知识链接等技术逐渐成为机器人认知的重要手段,在智能搜索、机器翻译、机器理解、自然语言问答等领域得到了广泛应用。知识图谱核心是构建相关领域的语义知识网络,对 VGE 知识的表达与建模、VGE 知识库的构建和管理、VGE 知识的可视化方法具有广泛的借鉴意义。

本文将知识图谱技术与地理知识工程相结合,提出基于多源异构数据的大规模地理知识图谱的构建,目的是建立大规模的地理知识库,用于融合语义关系和空间关系的地理智能搜索和地理知识自然语言智能问答,最终为智能虚拟地理环境系统的构建提供技术支撑。论文首先评述了知识图谱及地理知识图谱的研究现状;然后,提出了地理知识图谱构建的技术体系和流程,探讨了地理知识图谱与其他领域知识图谱的区别和联系;最后,讨论和阐述了地理知识图谱的应用方向。

1 国内外研究综述

1.1 知识图谱

知识图谱是人工智能领域的分支,是大数据时代知识表示最重要的一种方式。本质上是由具有属性的实体通过关系链接而成的网状知识库,即具有有向图结构的一个知识库,其中图的节点代表实体(entity)或者概念(concept),而图的边代表实体/概念之间的各种语义关系<sup>[7]</sup>。

知识图谱的概念起源于 20 世纪 50 年代末 60 年代初提出的语义网络(semantic net)<sup>[8]</sup>,知识图谱的发展在相关文献中都有了深入的阐述<sup>[7-9]</sup>。从 20 世纪 70 年代出现的“专家系统(expert system)”,70 年代中后期利用哲学领域的本体论创建计算机模型,到万维网之父 Berners-Lee 提出的“语义网(semantic web)”<sup>[10]</sup>和“链接数据(linked data)”<sup>[11]</sup>,都是“知识图谱”的前身。当前,开放域知识图谱比较有代表性的有: DBpedia<sup>[12]</sup>、YAGO<sup>[13]</sup>、Probase<sup>[14]</sup>、BableNet<sup>[15]</sup>等,国内的有 Zhishi.me<sup>[16]</sup>、CN-DBpedia<sup>[17]</sup>,见表 1。知识图谱具有规模大、语义丰富、质量高和结构友好等特点。

表 1 主要开放知识图谱的实体及关系数量  
Tab.1 The number of entities and relations of open KG

知识图谱	实体/概念数量	关系数量
YAGO	10 Million	120 Million
DBpedia	28 Million	9.5 Billion
Probase	2.7 Million	70 Billion
BableNet	14 Million	5 Billion

1.2 地理知识图谱

对地理知识的认识,不同的学者有不同的看法,大致可以分为 3 类:

(1) 地理知识的专业性认识。地理知识包括大量的地理概念、地理现象的描述,是关于地理时空问题的认知、理解与规律表达<sup>[18]</sup>,是高层次的地理信息<sup>[19]</sup>。

(2) 地理知识的泛在化认识。由于泛在地理信息获取渠道的不断拓展,认为地理知识是描述地球系统中自然和人文环境的信息<sup>[20]</sup>,互联网中

隐含了大量的地理知识<sup>[21]</sup>,是广义 GIS 的知识支撑<sup>[22]</sup>。

(3) 地理知识的综合性认识。从地理知识共享角度,按照不同的分类方法,将地理知识进行了系统性的分类。代表性的有:文献[3]将虚拟地理环境地理知识划分为事实型知识、规则及控制型知识和决策型知识 3 个层次。文献[1]认为地理知识基础 (geographic knowledge base, GKB) 包括地理实体 (geographic objects)、地理结构 (geographic structures)、地理关系 (geographic relations)、地理规则 (geographic Rules)、地理本体 (geographic ontology)、地名词典 (gazetteer)、物理数学模型 (physico-mathematical models) 和外部知识 (external knowledge)。

与地理知识图谱字面较为相近的概念是陈述彭和廖克等提出的地学信息图谱。文献[23]认为地学信息图谱是应用地学分析的系列多维图解来描述现状,并通过建立时空模型来重建过去和虚拟未来。文献[24]认为其是一种空间图形谱系,经过空间模型与地学认知的深入分析,可进行推理、反演与预测,形成对事物和现象更深层次的认识。由此可见,地学信息图谱是一种时空分析方法,是谱系 (spectrum) 的概念。地理知识图谱 (geographic knowledge graph) 是地理知识工程的一种知识表示,是网络图 (graph) 的概念。二者有着本质的不同。地理知识图谱的研究,主要是作为开放语义网的一个子集出现的,代表性的成果有: GeoNames Ontology, LinkedGeoData, GeoWorldNet 等<sup>[25]</sup>。针对地理知识图谱构建的相关技术研究,主要有地理实体抽取<sup>[26-27]</sup>、拓扑和方位关系的抽取<sup>[28-29]</sup>和地理知识图谱存储<sup>[30-32]</sup>等。在地理知识图谱应用方面,典型的有地理知识语义共享网络系统<sup>[33]</sup> (Geo-Wiki) 和基于地理知识的地名词典<sup>[34]</sup> (KIDGS)。

如图 1 所示,地理知识图谱的核心是建立地理知识库,在其基础上形成地理知识语义网,然后通过语义模型,实现地理知识的语义搜索、地理知识推荐、关联分析等功能,从而具备对地理、空间上分散的人、环境、事件等进行大规模实时关联和因果分析的能力。

## 2 地理知识图谱构建

### 2.1 地理知识图谱的内涵

地理知识图谱是知识图谱在地理学的拓

展,是结构化的地理语义知识库,通过形式化地描述地理学领域的概念、实体、属性及其相互关系,使得概念、实体间相互联结,构成网状知识结构。

地理知识可采用资源描述框架 (resource description framework, RDF) 表示为三元组形式,即“<s(主语), p(谓语), o(宾语)>”,形成由“点—边”组成的大规模有向图,如图 2 所示。其中点表示地理概念、地理实体及属性值,边表示概念与概念之间的关系、概念与实体的关系、实体与实体之间的关系、实体与属性的关系、属性与属性值的关系。

举例说明,如图 3 所示,实体与概念的关系: <中国,属于,国家>; 实体与实体之间的关系: <中国,首都,北京>; 实体与属性的关系: <北京,人口, 2 069.3 万>。

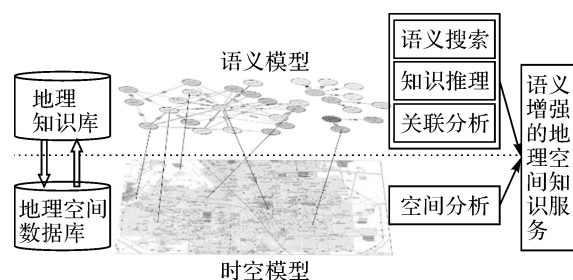


图 1 地理知识图谱概念图

Fig.1 Concept map of geographic knowledge graph

### 2.2 地理知识图谱的特点

地理知识图谱不单单是一个“增强型”的开放域知识图谱,而是需要针对地理知识自身的特点,对知识的概念、实体和关系进行拓展,具有以下特点:

(1) 地理知识图谱的构建是一项地理知识工程。地理知识图谱的构建是对地理知识的形式化表达的基础上,实现对地理知识抽取、融合、众包、表示、推理、链接、问答、语义搜索和可视化等,是一个系列的地理知识工程。

(2) 地理知识图谱描述的关系包括语义关系、空间关系和时间关系。开放域知识图谱描述的关系主要是实体之间的语义关系,地理知识图谱除了描述语义关系外,还需要考虑空间关系和时间关系的描述。如何建立地理时空关系和自然语言空间关系的映射,是地理知识图谱构建的核心关键问题。

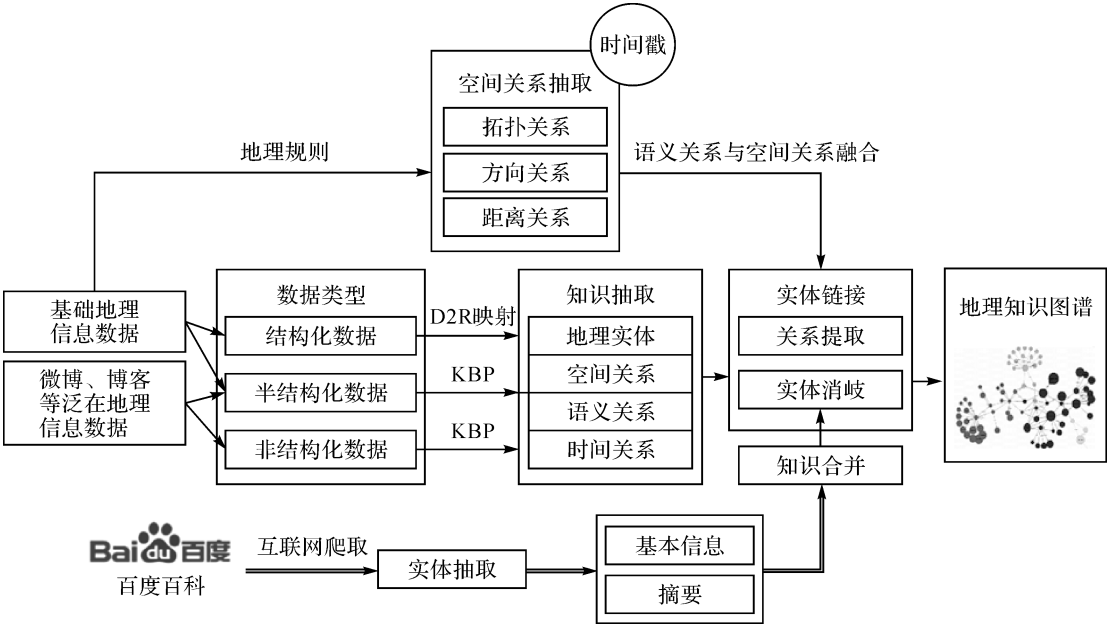


图 2 地理知识图谱数据层构建流程

Fig.2 Flowchart of knowledge graph data layer building

(3) 地理知识图谱需要描述事实型知识和过程型知识。事实型知识主要指已经得到地理学家或人们认可的地理术语、地理名词、地理分布、地理数据等反映地理事物外部特征和联系的知识,属于“轻量型”知识。过程型知识是指描述地理演变规律、地理预测规律等地理时空变换的地理模型,属于专业性较强的知识。知识图谱技术能够较好地描述、存储和表示事实型知识,并建立知识之间的联系。对地理规则、地理模型等知识如何

用图谱技术建模与表示,也是需要研究的。

2.3 地理知识图谱构建技术体系

地理知识图谱的构建采用“自顶向下”的方法构建,主要分为模式层和数据层构建。模式层包括地理本体和概念,主要通过本体构建方法实现地理本体概念分类体系的构建。数据层包括地理实体和相关属性,基于多源异构地理信息数据实现地理知识抽取、知识融合和知识更新,构建流程见图 3。

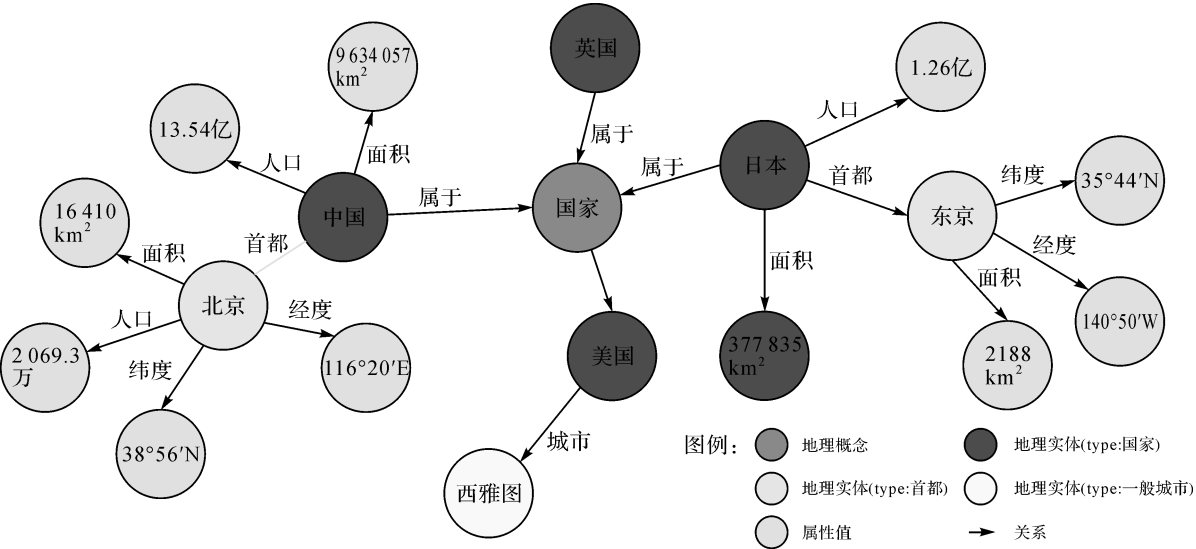


图 3 地理知识图谱示例

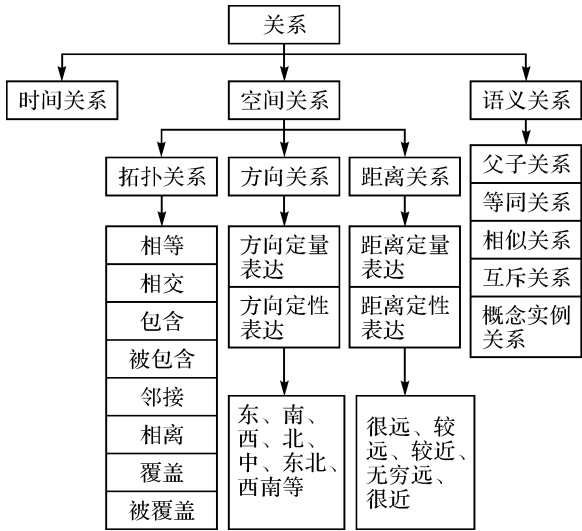
Fig.3 Example of geographic knowledge graph



2.3.1 多源异构数据的地理知识抽取

多源异构地理实体抽取和关系建立包括地理知识抽取、关系链接、关系推理等过程。地理知识的来源包括地理本体库、基础地理信息数据、百科数据、微博客、微信、新闻网站、专题网站等,可从中提取地理概念、地理实体、地理属性、语义关系、空间关系和时间关系。

地理概念主要源于地理本体的建立,如国家、城市、河流、草原等都是概念范畴,地理实体是概念的实例,含有具体的地理属性,如中国、郑州、黄河、呼伦贝尔大草原等都属于地理实体。地理关系主要有语义关系和空间关系(图 4)。语义关系包括“is-a 关系”、等同关系、相似关系、互斥关系等<sup>[35]</sup>。is-a 关系,既包括概念之间的父子关系(如“水系”与“河流”),也包括概念与实体(如“河流”与“黄河”)的实例关系;等同关系用来描述同级地理概念或实例之间的等价关系,如“黄河”与“母亲河”;互斥关系用来描述同级地理概念之间及相同性质实例之间的互斥关系,如“平原”与“山地”;相似关系用来描述意思上相近的同级地理概念或实例,如“第三世界”与“发展中国家”。



时间关系主要用于描述时变特征较为明显的各类地理现象。对时间关系的形式化描述,大致可分为两类:①地理事件、过程等地理实体之间的时变关系,如“人物→国家,人物→地点”等之间的关系<sup>[36]</sup>。非结构化的地理事件主要通过新闻网站等网络中获取,如“2018 年 4 月 3 日,普京访问土耳其”,可对“普京—访问—土耳其”三元组加入

时间标签拓展为四元组,用于描述时间关系。②空间信息的时间属性值变化。可用于图谱实体和关系的更新,如城市地名、空间位置、可随时间变化而变化,如“人行走在路上”,地点随着时间不断变化,则对人的位置属性值(如经纬度坐标)描述需要加入时间戳。

空间关系抽取的过程是 GIS 计算模型到自然语言空间关系的语义转换<sup>[37]</sup>,通过建立地理空间关系和自然语言描述对照词典实现关联<sup>[38]</sup>,如表 2 所示,拓扑相离关系和拓扑包含关系对应方位词,拓扑相交关系对应空间动词。不同的地图要素需要建立不同的对照表,如桥和河流之间有“横跨”动作,道路和公园之间有“穿过”动作等<sup>[39]</sup>。由计算机图形学计算方法(如点与线、面的关系)判断得到的空间关系类别,对应的自然语言描述存在多个词汇的情况,可根据地理实体的不同类别确定其常用的自然语言描述词汇。举例:线、面的相交关系对应“相交、交叉、流经、途径……”等多种描述,根据水系、交通、居民地等不同要素类型,将河流与城市的相交关系定义为“流经”,道路与城市的相交关系定义为“途径”。如图 5 是通过地图抽取出的城市、交通和河流之间的语义关系示例(部分)。

表 2 空间拓扑关系与自然语言描述对应<sup>[37]</sup>  
Tab. 2 Spatial topological relations corresponding to natural language descriptions<sup>[37]</sup>

空间拓扑关系	自然语言描述
相等	相等、相当、相同、相近、接近……
相交	相交、交叉、顺着、沿着、流经、途径……
穿越	穿越、横过、横穿、划分、横越、横断、横贯、穿过、经过……
内部	内部、里、内、里面……
包含	包含、围绕、环绕……
相离	旁路、近旁、旁边、远处、附近……
重叠	重叠、覆盖、交迭……
相接	邻接、相邻、连接、周围……

多源异构地理信息数据按存储类型可分为结构化、半结构化和非结构化数据。针对结构化数据(如地图、地名库等),通过建立数据库中概念与知识图谱中本体的映射关系以及基于规则的推理,实现从数据库中自动抽取地理实体、属性及其关系。例如基于电子地图和百科知识抽取地理实体,首先按照不同图层设计不同的抽取规则,抽取实体类、属性和属性值,然后借助开放知识库(如百度百科等)通过实体消歧、去重和属性填充

等操作丰富地理实体的属性信息。

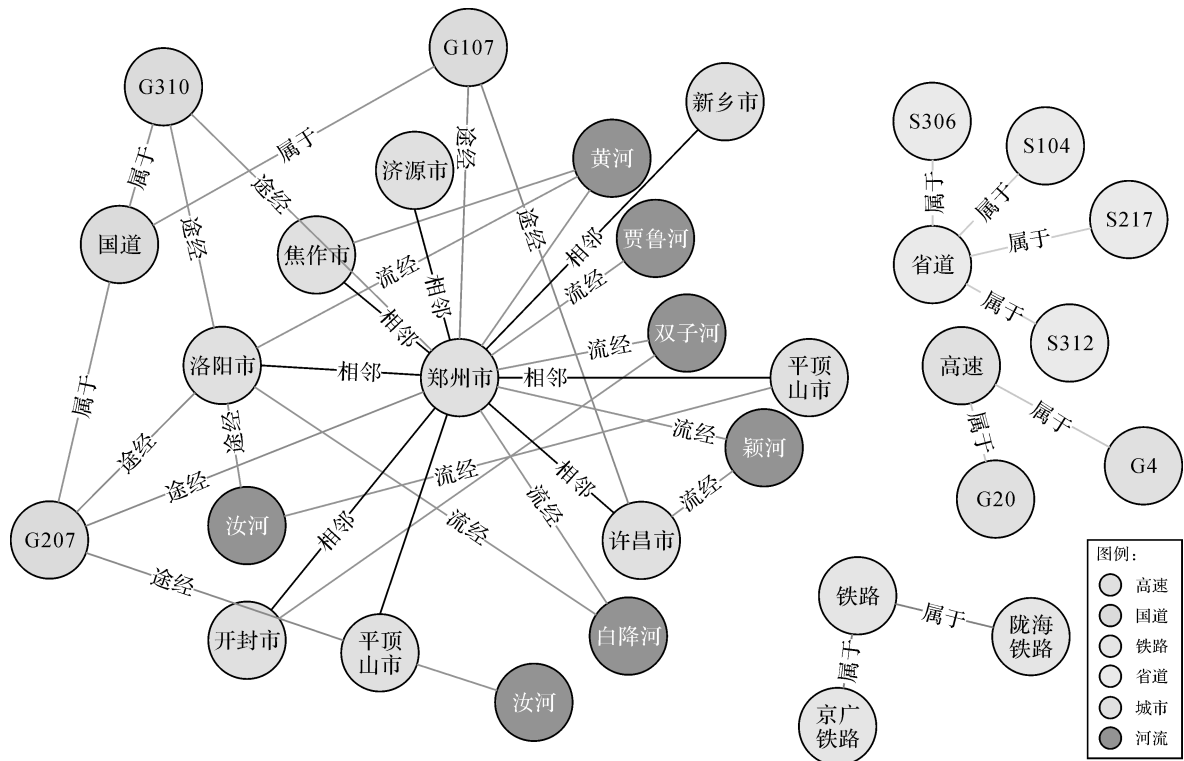


图 5 1 : 25 万郑州市地图提取的地理实体语义关系(部分)

Fig.5 Semantic relationships of geographic entities extracted from 1 : 250 000 map of Zhengzhou

针对半结构化数据(如百科数据等互联网数据),由于互联网网站主要通过模板方式构建,可建立相应的模板抽取器实现知识抽取。如图 6 所示,从百度百科、互动百科和中文维基百科抽取的实体,主要抽取的是实体名、标签、基本描述、信息卡、扩展链接、分类等信息,重点需要解决不同源数据的实体融合。

针对非结构化数据(如文本、微博微信等数据),主要利用已有的知识图谱知识,通过远程监督的方式来构建训练集,并利用深度学习的方法<sup>[40]</sup>学习出抽取器,利用抽取器来进文本中的知识进行抽取。

2.3.2 地理知识融合

由于地理知识的来源渠道众多,必须进行地理知识融合。地理知识融合是不同数据中不同标识实体的语义理解,关联到同一实体上,实现对同名、多名和缩写等多种实体语义的消歧和共指消解(例如:“中国”、“中华人民共和国”等都是指同一个地理实体)。

地理知识的融合包括模式层(即概念层)和实体层的融合。概念层的融合主要是基于地理本体

库的地理本体知识扩展。实体层的融合主要用到实体链接技术,主要包括 3 个步骤<sup>[41]</sup>:①生成候选地理实体。通过地理实体指称项从地理知识库中选取候选地理实体,主要有基于名称词典、基于搜索引擎的方法。②候选实体排序。主要方法可分为两类:一类是有监督的排序方法,通过标注训练数据集训练候选实体排序模型,包括二进制分类方法,概率方法和基于图的方法;另一类是无监督的排序方法,基于无标签的语料库进行模型训练。包括向量空间模型(vector space model, VSM)<sup>[42]</sup>和信息检索方法<sup>[43]</sup>。③无链接指称项预测。当地理知识库中没有相关的候选实体选项时,需要给出近似的实体。

实体链接技术按照不同的实体消歧方法可分为四种<sup>[7]</sup>。①基于概率生成模型方法:通过构建候选实体与实体指称项之间的概率模型<sup>[44]</sup>,提高实体链接的效率。②基于主题模型的方法:训练数据集训练 LDA 主题模型,通过语义相似度实现实体消歧<sup>[45]</sup>。③基于图的方法:建立基于图的模型,实现目标实体的选择<sup>[46]</sup>。④基于神经网络的方法:主要是利用深度神经网络模型以监督或

半监督的方式训练实体表示模型,依据语义相似度进行排序<sup>[47]</sup>。

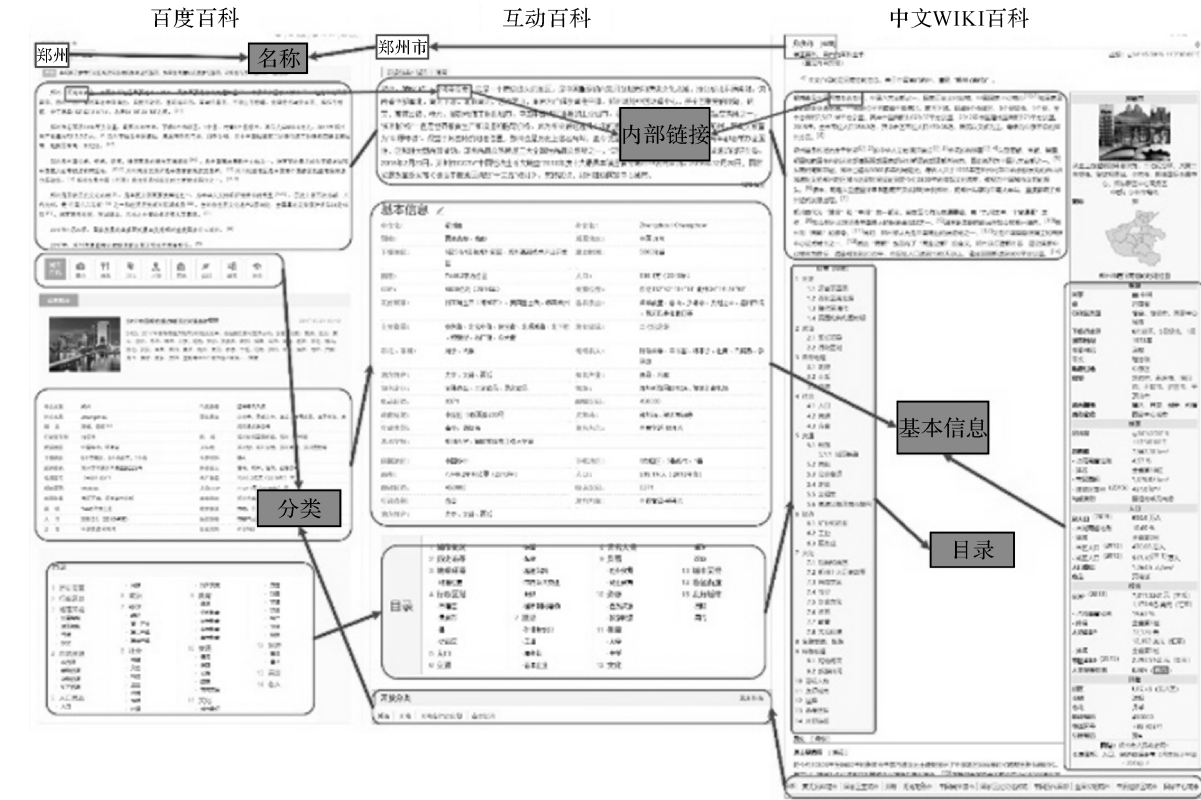


图 6 多源非结构化数据抽取  
Fig.6 The example of geographic knowledge card

2.3.3 地理知识推理

地理知识推理是指从地理知识库中的地理实体关系数据出发,经过计算机推理,建立地理实体间的新关联,从而拓展和丰富地理知识网络<sup>[9]</sup>。地理知识推理包括概念(即本体)的推理、地理实体的推理和实体属性值的推理等。地理本体推理,例如已知(河南,属于,中国)和(郑州,属于,河南),可以推出(郑州,属于,中国)。地理实体推理,例如已知(乾隆,父亲,雍正)和(雍正,父亲,康熙),可以得到(乾隆,祖父,康熙)或(康熙,孙子,乾隆)。

知识推理可以粗略地分为基于符号的推理和基于统计的推理<sup>[7]</sup>。基于符号的推理主要是利用相关规则,从已有实体关系推理出新的实体关系,并对知识图谱进行逻辑冲突检测。基于统计的推理是利用机器学习方法,通过统计规律从知识图谱中学习新的实体间关系,主要包括实体关系学习方法、类型推理方法和模式归纳方法<sup>[7]</sup>。也可按解决方法分为:基于描述逻辑的推理<sup>[48]</sup>、基于规则挖掘的推理<sup>[49]</sup>、基于概率逻辑的推理和基

于表示学习与神经网络的推理。由于知识库足够大,知识网络足够丰富,可以从地理知识库中推理出隐含的关系和知识。

2.3.4 地理知识的动态感知与更新

地理知识并非一成不变的,多是随着时间的变化而变化,在地理实体、关系和属性值都会发生变化,需要建立地理知识库的动态感知与更新机制。地理知识库的知识来源需建立不同的知识更新方法,可分为 3 种:

- (1) 基于结构化的地理信息数据文件更新。传统的地理信息数据已经有了一套较为完整的更新机制,地理知识库需保持与地理信息数据的更新周期一致,完成知识的局部更新。
- (2) 周期性更新。由于一些非结构化的地理知识数据,是从互联网得到的,新闻类、专题类网站,可建立周期性更新机制,但这样的更新方法花费很大,效率较低。
- (3) 基于新闻热搜词的更新方法。利用互联网爬虫系统实时监控互联网媒体,包括各类新闻网站、搜索关键词、微博等。从中识别出每日热



词,根据热词进行地理知识的更新。

3 地理知识图谱应用

按照应用服务对象,地理知识图谱的应用可分为面向人的和面向智能平台的应用。面向人的应用可使地理信息服务轻量化、大众化,通过智能语义搜索、地理知识智能问答、知识挖掘与决策分析等更好地为人认知地理世界服务;面向智能平台(如机器人)的地理知识应用,主要使机器人具备地理知识,从而在感知世界的基础上进一步实现认知和理解现实世界的目标。具体应用可分为以下几类:

3.1 基于地理知识库的智能知识服务

基于地理知识图谱的搜索结果是地理知识的形式化表达,多为地理知识卡片的形式(如图 7 所示),可将地理环境、地理实体、人文要素等资源进行融合和关联,以地理实体的形式对地理知识实现语义搜索和查询,在统一视图里进行管控。

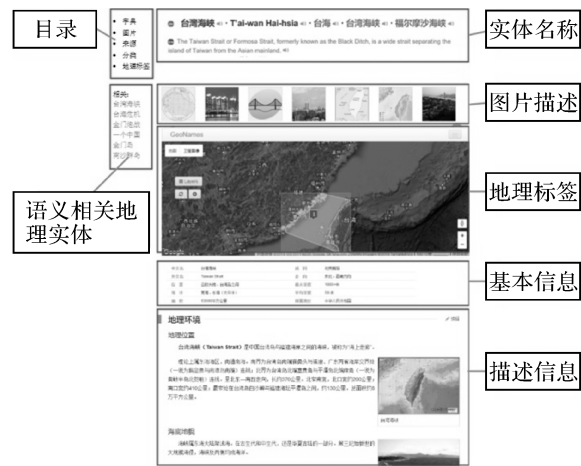


图 7 地理知识卡片示例

Fig.7 The example of geographic knowledge card

3.2 地理知识智能问答

自然语言在交互形式上更接近人类的交流习惯,基于地理知识库的智能问答系统可以实现高度智能化,能够适应地理信息快速、准确、启发式获取信息的需求。由于地理知识图谱具有结构化、关联化的特征,地理知识图谱相比纯文本资料、结构化数据库等,具有更丰富的语义表达、更精确的数据内容和更高效的检索方式等优势,基于地理知识库的智能问答是地理知识图谱的一个重要应用方向。

3.3 基于地理知识图谱的决策支持分析

基于地理知识图谱,可通过相关的大数据挖

掘算法,包括分类、聚类方法等,以及图计算方法包括图遍历、最短路径、路径探寻、权威节点分析、族群分析、相似节点发现等,实现地理实体关联分析、地理事件聚类分析等功能,实现时间空间上离散分布的人、环境、事件等的关联分析与挖掘,进而为相关决策服务。

3.4 为机器人等智能平台提供地理空间知识库

当前,机器学习、深度学习等技术的应用,使机器人等智能平台具备了对周围环境的动态感知能力,但若要使机器人能够具备环境的理解能力,则必须要有相关知识库的支撑。地理知识图谱可为机器人等智能平台提供地理空间相关知识,为机器人导航地图提供丰富的语义网,使其具备地理空间理解能力。

4 总结与展望

虚拟地理环境的知识工程是智能虚拟地理环境的基础工程。地理信息智能化服务是虚拟地理环境智能化发展的新趋势。对多源异构隐性的泛在地理信息处理,是地理信息向大众化、普适化方向转变的基础。业界已经对地理知识表示、地学信息图谱等方面进行了相关的研究,但其主要仍是对专业的地理学领域专家知识的刻画与表示,难以适应大规模地理知识的共享、交互与应用。地理知识图谱适合对事实型地理知识的大规模建模与表示,可为虚拟地理环境的知识工程建设提供基础支撑。

本文重点剖析了地理知识图谱的概念、内涵及研究现状,提出了基于多源异构地理信息数据的地理知识构建技术体系和流程。地理环境知识图谱的构建是地理信息服务通向智能化的桥梁,其有着广阔的应用前景,存在诸多挑战,需重点解决地理知识的多源数据抽取、地理实体空间关系、语义关系抽取、地理知识融合、地理知识链接、地理知识质量评价及地理知识的更新等关键科学问题,为实现面向智能虚拟地理环境的自动化、智能化的地理知识服务奠定理论与方法基础。

参考文献:

[1] 林琤, 游兰. 虚拟地理环境知识工程初探[J]. 地球信息科学学报, 2015, 17(12): 1423-1430.  
LIN Hui, YOU Lan. A Tentative Study on Knowledge Engineering for Virtual Geographic Environments [J]. Journal of Geo-information Science, 2015, 17(12): 1423-1430.  
[2] LAURINI R. Towards Smart Urban Planning through



- Knowledge Infrastructure[C]//Proceedings of the 9th International Conference on Advanced Geographic Information Systems, Applications, and Services. Nice, France: IARIA, 2017: 75-80.
- [3] 林琿, 张春晓, 陈旻, 等. 论虚拟地理环境对地理知识的表达与共享[J]. 遥感学报, 2016, 20(5): 1290-1298. DOI: 10.11834/lrs.20166185.
- LIN Hui, ZHANG Chunxiao, CHEN Min, et al. On Virtual Geographic Environments for Geographic Knowledge Representation and Sharing[J]. Journal of Remote Sensing, 2016, 20(5): 1290-1298. DOI: 10.11834/lrs.20166185.
- [4] ROWLEY J. The Wisdom Hierarchy: Representations of the DIKW Hierarchy[J]. Journal of Information Science, 2007, 33(2): 163-180.
- [5] LIN Hui, CHEN Min, LÜ Guonian, et al. Virtual Geographic Environments (VGE): A New Generation of Geographic Analysis Tool[J]. Earth-Science Reviews, 2013, 126: 74-84.
- [6] CHEN Min, LIN Hui, LÜ Guonian. Virtual Geographic Environments[C]//International Encyclopedia of Geography. [S.l.]: American Association of Geographers (AAG), 2017. DOI: 10.1002/9781118786352.wbieg0448.
- [7] 漆桂林, 高桓, 吴天星. 知识图谱研究进展[J]. 情报工程, 2017, 3(1): 4-25. DOI: 10.3772/j.issn.2095-915x.2017.01.002.
- QI Guilin, GAO Huan, WU Tianxing. The Research Advances of Knowledge Graph[J]. Technology Intelligence Engineering, 2017, 3(1): 4-25. DOI: 10.3772/j.issn.2095-915x.2017.01.002.
- [8] SOWA J F. Principles of Semantic Networks: Exploration in the Representation of Knowledge[M]. San Mateo, Calif: Morgan Kaufmann, 1991: 135-157.
- [9] 李涓子, 侯磊. 知识图谱研究综述[J]. 山西大学学报(自然科学版), 2017, 40(3): 454-459.
- LI Juanzi, HOU Lei. Reviews on Knowledge Graph Research[J]. Journal of Shanxi University (Natural Science Edition), 2017, 40(3): 454-459.
- [10] BERNERS-LEE T, HENDLER J, LASSILA O. The Semantic Web: A New Form of Web content that is Meaningful to Computers Will Unleash A Revolution of New Possibilities[J]. Scientific American, 2001, 284(5): 34-43.
- [11] BIZER C, HEATH T, BERNERS-LEE T, et al. Linked Data: The Story So Far[J]. International Journal on Semantic Web and Information Systems, 2009, 5(3): 1-22.
- [12] AUER S, BIZER C, KOBILAROV G, et al. Dbpedia: A Nucleus for A Web of Open Data[M]// Aberer K, Choi K S, Noy N, et al. The Semantic Web. Berlin, Heidelberg: Springer, 2007: 722-735.
- [13] SUCHANEK F M, KASNECI G, WEIKUM G. Yago: A Core of Semantic Knowledge[C]//Proceedings of the 16th International Conference on World Wide Web. Banff, Alberta, Canada: ACM, 2007: 697-706.
- [14] WU Wentao, LI Hongsong, WANG Haixun, et al. Probase: A Probabilistic Taxonomy for Text Understanding[C]//Proceedings of 2012 ACM SIGMOD International Conference on Management of Data. Scottsdale, Arizona: ACM, 2012: 481-492.
- [15] NAVIGLI R, PONZETTO S P. BabelNet: Building A Very Large Multilingual Semantic Network[C]//Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics. Uppsala, Sweden: Association for Computational Linguistics, 2010: 216-225.
- [16] NIU Xing, SUN Xinruo, WANG Haofen, et al. Zhishi. Me-Weaving Chinese Linking Open Data[M]//AROYO L, WELTY C, ALANI H, et al. The Semantic Web-ISWC 2011. Berlin, Heidelberg: Springer, 2011: 205-220.
- [17] XU Bo, XU Yong, LIANG Jiaqing, et al. CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System[C]//Proceedings of the 30th International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems. Arras, France: Springer, 2017: 428-438.
- [18] 王伟星, 龚建华. 地学知识可视化概念特征与研究进展[J]. 地理与地理信息科学, 2009, 25(4): 1-7.
- WANG Weixing, GONG Jianhua. Concept, Characteristics and Advances of Geographic Knowledge Visualization[J]. Geography and Geo-Information Science, 2009, 25(4): 1-7.
- [19] 龚建华, 李亚斌, 王道军, 等. 地理知识可视化中知识图特征与应用——以小流域淤地坝系规划为例[J]. 遥感学报, 2008, 12(2): 355-361.
- GONG Jianhua, LI Yabin, WANG Daojun, et al. Characteristics and Applications of Knowledge Maps in Geographic Knowledge Visualization: A Case Study of Spatial Planning of Dam Systems in Watersheds[J]. Journal of Remote Sensing, 2008, 12(2): 355-361.
- [20] DANGERMOND J. Geographic Knowledge: Our New Infrastructure[M]. Redlands, CA: ArcNews, 2010.
- [21] 陆锋, 余丽, 仇培元. 论地理知识图谱[J]. 地球信息科学学报, 2017, 19(6): 723-734. DOI: 10.3724/SP.J.1047.2017.00723.
- LU Feng, YU Li, QIU Peiyuan. On Geographic Knowledge Graph[J]. Journal of Geo-Information Science, 2017, 19(6): 723-734. DOI: 10.3724/SP.J.1047.2017.00723.
- [22] 陆锋, 张恒才. 大数据与广义 GIS[J]. 武汉大学学报(信息科学版), 2014, 39(6): 645-654. DOI: 10.132023/j.whugis20140148.
- LU Feng, ZHANG Hengcai. Big Data and Generalized GIS[J]. Geomatics and Information Science of Wuhan University, 2014, 39(6): 645-654. DOI: 10.132023/j.whugis20140148.
- [23] 陈述彭, 岳天祥, 励惠国. 地学信息图谱研究及其应用[J]. 地理研究, 2000, 19(4): 337-343.

- CHEN Shupeng, YUE Tianxiang, LI Huiguo. Studies on Geo-Informatic Tupu and Its Application[J]. Geographical Research, 2000, 19(4): 337-343.
- [24] 廖克. 地学信息图谱的探讨与展望[J]. 地球信息科学, 2002(1): 14-20.
- LIAO Ke. The Discussion and Prospect for Geo-informatic Tupu[J]. Geo-information Science, 2002(1): 14-20.
- [25] BALLATORE A, WILSON D C, BERTOLOTTO M. A Survey of Volunteered Open Geo-knowledge Bases in the Semantic Web[M]//PASI G, BORDOGNA G, JAIN L C. Quality Issues in the Management of Web Information. Berlin, Heidelberg: Springer, 2013: 93-120.
- [26] MARRERO M, URBANO J, SÁNCHEZ-CUADRADO S, et al. Named Entity Recognition: Fallacies, Challenges and Opportunities[J]. Computer Standards and Interfaces, 2013, 35(5): 482-489.
- [27] DAVIES C. Reading Geography Between the Lines: Extracting Local Place Knowledge from Text[M]//TENBRINK T, STELL J, GALTON A, et al. Spatial Information Theory. Cham: Springer, 2013, 8116: 320-337.
- [28] 余丽, 陆锋, 刘希亮. 开放式地理实体关系抽取的 Bootstrapping 方法[J]. 测绘学报, 2016, 45(5): 616-622. DOI: 10.11947/j.AGCS.2106.20150181.
- YU Li, LU Feng, LIU Xiliang. A Bootstrapping Based Approach for Open Geo-entity Relation Extraction[J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(5): 616-622. DOI: 10.11947/j.AGCS.2106.20150181.
- [29] LOGLISCI C, IENCO D, ROCHE M, et al. Toward Geographic Information Harvesting: Extraction of Spatial Relational Facts from Web Documents[C]//Proceedings of the 12th International Conference on Data Mining Workshops. Brussels: IEEE, 2012: 789-796.
- [30] KYZIRAKOS K, KARPATHTOTAKIS M, KOUBARAKIS M. Strabon: A Semantic Geospatial DBMS[C]//Proceedings of the 11th International Semantic Web Conference (ISWC 2012). Boston, MA, USA: Springer, 2012: 295-311.
- [31] 段红伟, 孟令奎, 黄长青, 等. 面向 SPARQL 查询的地理语义空间索引构建方法[J]. 测绘学报, 2014, 43(2): 193-199. DOI: 10.13485/j.cnki.11-2089.2014.0028.
- DUAN Hongwei, MENG Lingkui, HUANG Changqing, et al. A Method for Geo Semantic Spatial Index on SPARQL Query[J]. Acta Geodaetica et Cartographica Sinica, 2014, 43(2): 193-199. DOI: 10.13485/j.cnki.11-2089.2014.0028.
- [32] GÜR N, PEDERSEN T B, ZIMÁNYI E, et al. A Foundation for Spatial Data Warehouses on the Semantic Web [J]. Semantic Web, 2017: 1-31. DOI: 10.3233/SW-170281.
- [33] GAO Yong, GAO Song, LI Runqiang, et al. A Semantic Geographical Knowledge Wiki System Mashed up with Google Maps[J]. Science China Technological Sciences, 2010, 53(S1): 52-60.
- [34] LIU Yu, LI Runqiang, CHEN Kaichen, et al. KIDGS: A Geographical Knowledge-informed Digital Gazetteer Service [C] // Proceedings of the 17th International Conference on Geoinformatics. Fairfax, VA: IEEE, 2009: 1-6.
- [35] 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.
- WANG Dongxu, ZHU Yunqiang, PAN Peng, et al. Construction of Geodata Spatial Ontology and Its Application in Data Retrieval[J]. Journal of Geo-information Science, 2016, 18(4): 443-452.
- [36] TRIVEDI R, DAI Hanjun, WANG Yichen, et al. Know-evolve: Deep Temporal Reasoning for Dynamic Knowledge Graphs[C]//Proceedings of the 34th International Conference on Machine Learning. Sydney, Australia: International Convention Centre, 2017, 70: 3462-3471.
- [37] 张雪英, 闻国年. 自然语言空间关系及其在 GIS 中的应用研究[J]. 地球信息科学学报, 2007, 9(6): 77-81.
- ZHANG Xueying, LÜ Guonian. Natural-language Spatial Relations and Their Applications in GIS[J]. Geo-information Science, 2007, 9(6): 77-81.
- [38] 乐小虬, 杨崇俊, 于文洋. 基于空间语义角色的自然语言空间概念提取[J]. 武汉大学学报(信息科学版), 2005, 30(12): 1100-1103.
- LE Xiaoqi, YANG Chongjun, YU Wenyang. Spatial Concept Extraction Based on Spatial Semantic Role in Natural Language[J]. Geomatics and Information Science of Wuhan University, 2005, 30(12): 1100-1103.
- [39] 杜清运, 任福. 空间信息的自然语言表达模型[J]. 武汉大学学报(信息科学版), 2014, 39(6): 682-688.
- DU Qingyun, REN Fu. Representation Model of Spatial Information in Natural Language[J]. Geomatics and Information Science of Wuhan University, 2014, 39(6): 682-688.
- [40] 张春菊. 面向中文文本的事件时空与属性信息解析方法研究[J]. 测绘学报, 2015, 44(5): 590. DOI: 10.11947/j.AGCS.2015.20140657.
- ZHANG Chunju. Interpretation of Event Spatio-temporal and Attribute Information in Chinese Text[J]. Acta Geodaetica et Cartographica Sinica, 2015, 44(5): 590. DOI: 10.11947/j.AGCS.2015.20140657.
- [41] SHEN Wei, WANG Jianyong, HAN Jiawei. Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(2): 443-460.
- [42] SOCHER R, CHEN Danqi, MANNING C D, et al. Reasoning with Neural Tensor Networks for Knowledge Base Completion[C]//Proceedings of Neural Information Processing Systems. Nevada, USA: MIPS, 2013: 926-934.

[43]

ARTILES J, LI Qi, CASSIDY T, et al. CUNY BLENDER TAC-KBP2011 Temporal Slot Filling System Description [C]//Proceedings of Text Analysis Conference (TAC). [S.l.]: TAC, 2011.

[44]

BLANCO R, OTTAVIANO G, MEIJ E. Fast and Space-efficient Entity Linking for Queries[C]// Proceedings of the Eighth ACM International Conference on Web Search and Data Mining. Shanghai, China: ACM, 2015: 179-188.

[45]

ZHANG Wei, SIM Yanchuan, SU Jian, et al. Entity Linking with Effective Acronym Expansion, Instance Selection, and Topic Modeling [C] // Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence. Barcelona, Catalonia, Spain: [s.n.],2011: 1909-1914.

[46]

ALHELBAWY A, GAIZAUSKAS R. Graph Ranking for Collective Named Entity Disambiguation[C]//Proceedings of the 52nd Annual Meeting of the association for computational linguistics. Baltimore, Maryland: ACL, 2014: 75-80.

[47]

HUANG Hongzhao, HECK L, JI Heng. Leveraging Deep Neural Networks and Knowledge Graphs for Entity Disambiguation[J]. arXiv preprint arXiv: 1504. 07678, 2015: 1275-1284.

[48]

LEE T W, LEWICKI M S, GIROLAMI M, et al. Blind Source Separation of More Sources Than Mixtures Using Overcomplete Representations[J]. IEEE Signal Processing Letters, 1999, 6(4): 87-90.

[49]

QUINLAN J R. Learning Logical Definitions from Relations [J]. Machine Learning, 1990, 5(3): 239-266.

(责任编辑:宋启凡)

收稿日期: 2018-03-16

修回日期: 2018-05-07

第一作者简介: 蒋秉川(1984—),男,博士,讲师,研究方向为时空大数据分析 with 地理知识图谱。

First author: JIANG Bingchuan(1984—),male,PhD,lecturer,majors in spatio-temporal big data and geographic knowledge graph.

E-mail: jbc021@163.com