

Research on the Personalized Discovery of Geospatial Resources Based on CSW

Wu Du¹, Hao Tian²

¹State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan Hubei

²School of Information Engineering, Hubei University of Economics, Wuhan Hubei

Email: haotian@whu.edu.cn

Received: Sep. 28th, 2017; accepted: Oct. 10th, 2017; published: Oct. 17th, 2017

Abstract

In order to solve the main problems existing in current space geographic information retrieval, and enhance the capacity of CSW on geographical resources retrieval, this paper combined ontology query technology system and the preference theory into CSW, explored the geographic resources metadata model that adapts to preference queries, proposed a personalized geospatial resource discovery system, designed the geographic resource discovery method based on semantic technology, then set up a prototype system. The examples prove that the proposed method can effectively retrieve the high quality geospatial service resources.

Keywords

Geospatial Resources, Discovery, Personalization, Engine, CSW

基于CSW的个性化地理空间资源发现研究

杜武¹, 田浩²

¹武汉大学测绘遥感信息工程国家重点实验室, 湖北 武汉

²湖北经济学院信息工程学院, 湖北 武汉

Email: haotian@whu.edu.cn

收稿日期: 2017年9月28日; 录用日期: 2017年10月10日; 发布日期: 2017年10月17日

摘要

为解决目前空间地理信息检索中存在的主要问题, 增强CSW对地理资源的检索能力, 将本体技术与偏好

查询理论引入CSW, 采用适应偏好查询的地理资源元数据模型, 提出了一个个性化地理空间资源发现系统, 设计基于语义的地理资源发现方法, 并搭建了原型系统。实例证明本文提出的方法能够有效检索出优质的地理空间服务资源。

关键词

空间资源, 发现, 个性化, 引擎, 网络目录服务

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在空间数据基础框架(Spatial Data Infrastructure, SDI)的分布式环境中, 地理空间资源的有效发现和访问是重要任务之一。网络目录服务(Catalog Service for the Web, CSW)是管理对地观测数据的标准协议框架, 支持 ebRIM、ISO 等应用纲要(Application Profile), 提供了数据、服务和其他资源的元数据的发现、浏览和查询的接口[1]。随着地球空间信息产业的不断发展, 越来越多的地理信息资源以元数据的形式被发布到 CSW 中, 针对网络目录服务的检索需求也日益增多。目前在 SDI 内, 地理数据和服务的检索都主要依赖于 CSW。但由于分布式网络环境的复杂性和异构性, 以及跨学科领域的语义差别, CSW 的应用受到了较大限制。

为解决以上问题, 本文将本体技术与偏好查询理论引入 CSW, 采用适应偏好查询的地理资源元数据模型, 提出基于语义的地理资源发现方法, 支持多源 CSW 的统一检索, 增强 CSW 对地理资源的检索能力, 为分布式环境下地理空间信息的有效发现提供可行的理论和技术支持。

2. 背景分析

2.1. 相关研究分析

为提高 CSW 的应用性能, Yue P.等利用地理空间语义和服务, 提出了一种基于工作流的复杂地理空间特征的发现方法。通过特征类型的基本特征和主题语义之间的空间语义, 构建基于工作流的服务链来定位图像中与语义相关的复杂特征。该方法也扩展了 CSW 的 ebRIM 信息模型, 以增强目录服务的地理信息处理和知识发现能力, 更好地支持服务发现和编排[2]。为实现地表覆盖信息空间服务的有效检索, 徐生磊等在分析地表覆盖信息的特征的基础扩展了 CSW 目录的地理信息元数据标准, 以此设计了一个全球地表覆盖信息元数据模型, 并提出了一个相应的全球地表覆盖信息元数据目录系统软件架构[3]。文献[4]中也进行了类似的工作, 文中也对 CSW 进行了扩展, 设计了一个统一空间资源发现服务框架。Stock K.等则使用一个语义丰富的 ebRIM 注册表, 包含一个特征类型目录(FTC)来表示地理特征类型的语义, 包括操作、属性和特征类型之间的关系, 以提升 SDI 的适用性[5]。之后其又提出了一种专门用于语义数据存储的 CSW 应用纲要, 定义了信息模型到 OWL、RDF 模型的映射, 避免了语义信息在映射过程中的丢失问题[6]。为利用地理数据空间本体来实现地理数据的智能关联, 文献[7]将该本体表示为一个五元组, 并通过具体检索应用流程在特定平台中进行对其进行了性能验证。Lutz M.等人则通过扩展本体推理机和级联目录服务这两个组件, 提出了一个专门针对 SDI 的地理信息检索框架。在此框架中, 所有 WFS 的要素类型都可以通过注册映射的方式被关联到对应的本体概念上[8]。文献[9]专门设计了一个空间信息的本

体模型, 并使用它来完成网络地理信息的推理, 以此来实现空间信息的描述和以及空间定位。Fitzner D. 等提出了一种新的方法来发现地理服务(WPS)。该方法中, 发现请求和 Web 处理服务被注释为逻辑编程语言中的连接查询, 而发现过程则是基于这些描述之间的逻辑编程查询控制。该方法将输入和输出之间的关系形式化, 使之能够更精确地捕获 WPS 的功能并能使用逻辑编程查询在发现过程中进行有效的推理 [10]。Janowicz K. 等认为地理语义有利于支持跨异构信息源的复杂查询和检索以及服务编排、语义转换和即时集成。他们提出了将 OGC 和语义技术之间进行映射配置的方法并指出了如何将 SDI 与链接数据相结合 [11]。

这些研究都在不同程度上解决了 SDI 中的服务资源检索及应用问题。但是, 由于 CSW 自身的原理结构特点和用户专业背景的局限, 地理空间信息资源的检索中依然存在着可操作性差, 检索功能弱, 查询效率低和用户决策支持少等突出问题。而且现有研究较少考虑用户的偏好问题, 没能有效解决用户的个性化需求。

2.2. 问题分析

具体而言, 目前阻碍空间地理信息有效检索的主要问题有以下几点: 1) ebRIM、ISO 等应用纲要都是面向地理资源提供者的, 不是面向地理资源消费者, 不同数据中心使用不同的发布方式和元数据模型, 用户如果不了解目录服务使用的信息模型, 便无法准确描述自己的需求; 2) OGC filter 规范定义的结构只能描述一些相对单一的查询条件, 并不能准确描述用户的需求, 无法使用自然的逻辑语法来表达用户对地理资源的偏好; 3) 目录服务缺乏语义性质的描述, 用户查询主要依赖关键词, 当用户查询使用的关键词和描述地理资源的词汇表不一致时, 便无法检索出结果, 即使目录服务中已经有相关的地理资源; 4) 目前的检索结果表示方法基本上是基于纯文本的, 缺乏足够的信息和功能来辅助用户有效做出最终选择。

3. 系统设计

为解决以上问题, 在前期研究成果上, 本文设计如图 1 所示的个性化地理空间资源发现系统框架。

3.1. 用户个性化引擎(User Personalization Engine, UPE)

用户个性化引擎 UPE 接受和解析用户的请求, 将其分解为基本功能需求和偏好需求两个子项, 并根据基本查询规则和基本偏好项对这两个子项进行规范化。该引擎扩展偏好查询中的基本偏好关系, 建立面向地理资源的偏好查询关系模型。同时基于操作信息数据库中的用户操作信息建立其混合偏好模型。

3.2. 元数据规范引擎(Metadata Standardization Engine, MSE)

CSW 提供了核心查询属性来实现不同应用纲要的统一查询, 但 OGC 提供的抽象信息模型描述能力比较弱, 无法有效地描述数据和服务的信息。元数据规范引擎中针对地理资源建立一个面向用户的元数据模型, 明确、统一地描述地理资源的偏好属性。用户不需要理解应用纲要中复杂的信息模型, 通过地理资源元数据模型, 就可以提出相应的查询, 进而完成地理资源的检索。

3.3. 分布式发现引擎(Distributed Discovery Engine, DDE)

分布式发现引擎依据发现策略将规范后的基本功能需求子项解析为 CSW Filter 查询条件, 并将这些 Filter 查询条件将转发给底层的各个 CSW, 也负责整合底层各个 CSW 返回的结果, 根据地理资源的标注和元数据模型生成统一的简单地理资源对象; 然后基于规范后的基本偏好子项, 在地理资源监控模块的协助下, 根据算法构建偏好网络, 对返回的候选项进一步筛选。

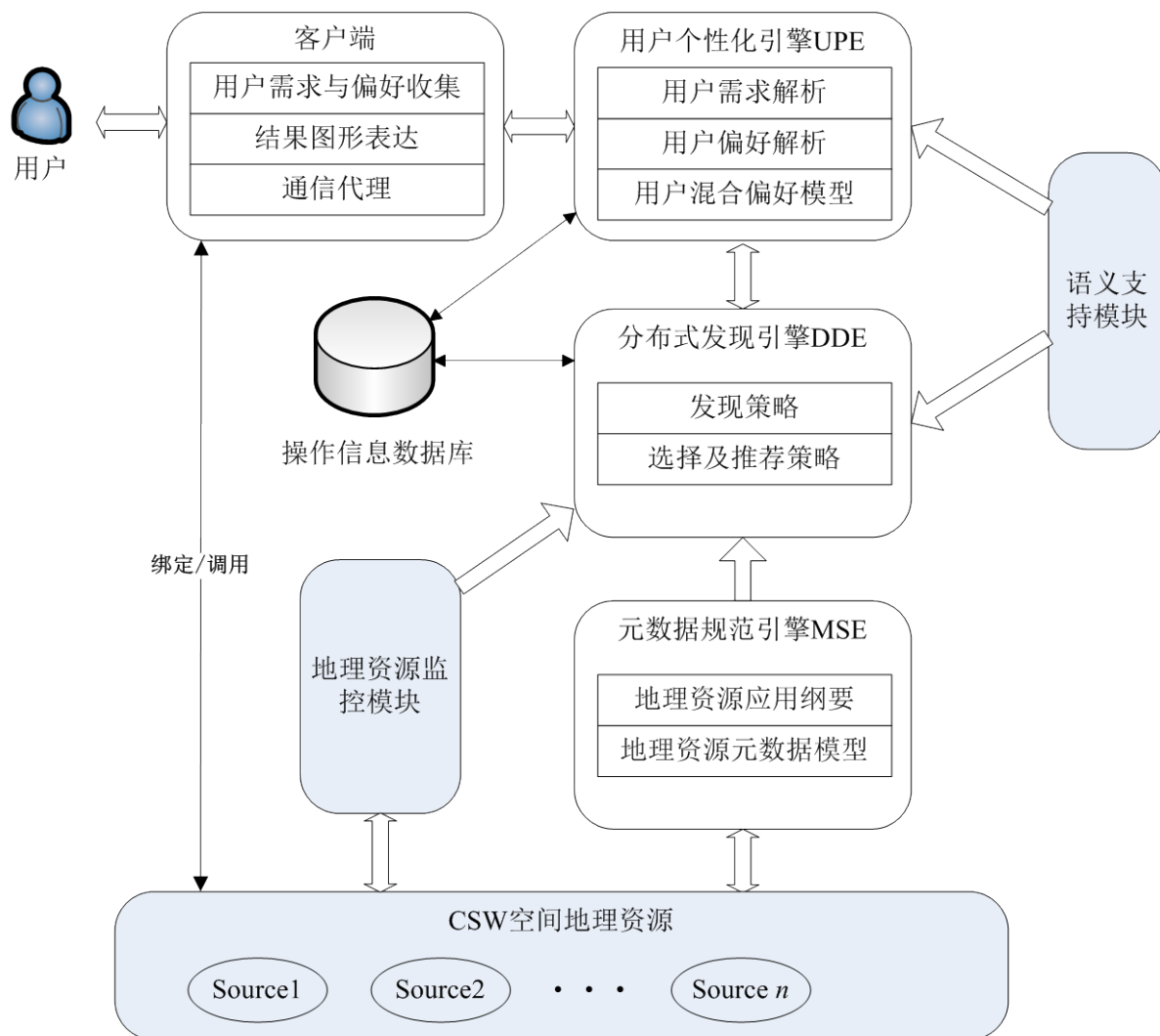


Figure 1. Architecture diagram of personalized geospatial resource discovery system

图 1. 个性化地理空间资源发现系统架构图

3.4. 插件式图形表达

为提高系统的人机交互性, 增强可视化功能, 在客户端借助第三方工具和方法来实现插件式图形表达与交互方式, 显示地理资源的发现过程及结果, 通过图形化形式来给用户资源相关度、对象服务参数、资源热度和用户反馈等大量有用信息, 引导其做出最终选择。

3.5. 地理资源监控模块

地理资源监控模块包含一个资源注册表, 该模块定时监控列表中已注册地理资源的性能参数和用户评价等, 为分布式发现引擎和用户的选择不提供依据, 同时也在用户离线时自动添加新的候选资源。

4. 系统实现

4.1. 原型系统构建

本文使用文献[12]中提出的方法来构建用户个性化模型, 其可以表示为:

$$UserProfile = (ImPM, ExPM) \quad (1)$$

$$ImPM = (C, A^C, R, A^R, H, X, INS) \quad (2)$$

$$ExPM = \{KB, UPO, PCT\} \quad (3)$$

其中 $ImPM$ 为隐式偏好子模型, 它是一个基于本体技术构成的七元组; 而 $ExPM$ 为显式偏好子模型, 它由基于严格偏好关系的偏好构造函数组成。

同时基于文献[13]中设计的地理资源统一模型 GRUM 来描述用户的服务请求, GRUM 的 UML 图如图 2 所示。一个用户服务请求 R 可以表示为:

$$R = (N, O, E) \quad (4)$$

N 是一个由本体概念组成的概念集, 用来归纳和表达用户对目标服务的功能需求, 其内容与 GRUM 中的 $sIdentifier$, $sTitle$, $sAbstract$, $sKeyword$, $sType$ 等参数对应; O 用来在语义层面上指定用户对目标服务中操作的输入和输出的要求, 与 GRUM 中的 I/O 参数对应; E 是用户的实时偏好, 表明用户对所需服务 QoS 等参数的需求。

原型系统中基于前期研究中提出的差分式服务发现策略(Differential Service Discovery Strategy, DS2) [14]来设计服务匹配策略, 其匹配函数可以表示为:

$$RSim_i = \alpha \times NSim(s_i, sn, R, N) + \beta \times FSim(s_i, O, R, O) \quad (5)$$

$$(\alpha, \beta > 0, \alpha + \beta = 1)$$

$RSim_i$ 代表候选服务 s_i 与服务请求 R 的匹配程度, 相似度计算函数 $NSim(\cdot)$ 与 $FSim(\cdot)$ 采用文献[8]中提出的方法。

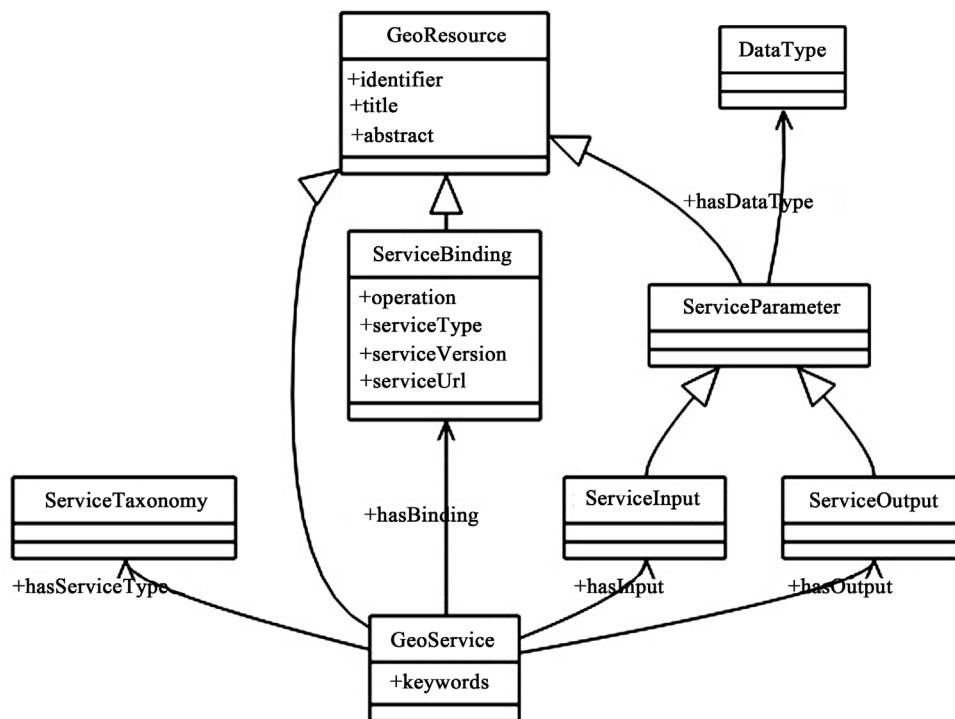


Figure 2. UML diagram of GRUM

图 2. GRUM 模型的 UML 图

原型系统主要基于 Java 开发。其中 Web 系统采用 Spring MVC 框架搭建; 前端页面基于 CSS 框架 Cube 搭建; 数据库使用 MongoDB; 本体操作则使用 Jena 框架; 使用 Lucene 架构对服务进行文本处理。

4.2. 系统实例

GWSC 是文献[13]中基于 Eclipse GEF 框架开发的一个服务组合流程建模设计器。为验证本文设计的系统性能, 使用 GWSC 对全国气温分布制作进行工作流建模。

建模过程中, 系统自动获取逐小时的气温数据, 基于气温数据做插值分析; 同时获取全国的行政区划图数据, 然后合并做矢量数据切割栅格数据处理, 最后将结果导出为常用图片格式。在此过程中, 自动气象站数据和数据处理功能均作为空间信息服务发布到 CSW, 每个流程节点通过本文提出的系统检索并绑定某一具体空间信息服务, 从而实现从地面观测数据到气象产品输出的自动化制作流程。每个节点绑定的具体空间信息服务决定了最终导出结果的质量。

图 3 和图 4 分别展示了露点温度分布产品服务链和空气温度分布产品服务链的执行结果。可以看到, 服务链构建成功后, GWSC 流程设计器可以选择执行该空间信息服务链, 执行服务链成功后, 可以得到我国大陆地区的逐小时露点温度分布产品和空气温度分布产品, 并能最终通过原型系统展示服务产品结果。这些结果表明, 本文提出的方法能够有效检索出优质的地理空间服务资源, 可以较好地应用于实际空间服务应用中。

5. 总结

本文通过建立一个基于本体技术与偏好查询理论的用户个性化模型, 采用适应偏好查询的地理资源统一模型 GRUM, 提出基于语义的地理资源发现方法, 构建了一个基于 CSW 的个性化地理空间资源发现系统框架。通过原型系统实例证明, 本文提出的方法能支持多源 CSW 的统一检索, 有效增强 CSW 对地理资源的检索能力。

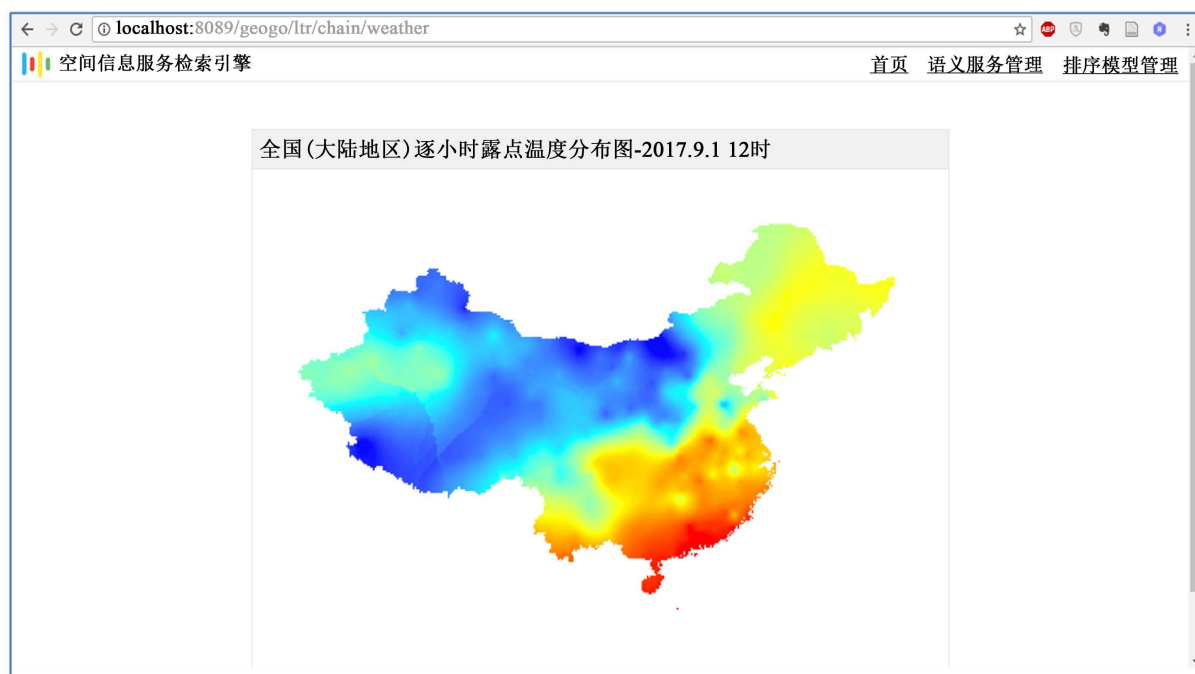


Figure 3. Distribution map of dew temperature

图 3. 露点温度分布图

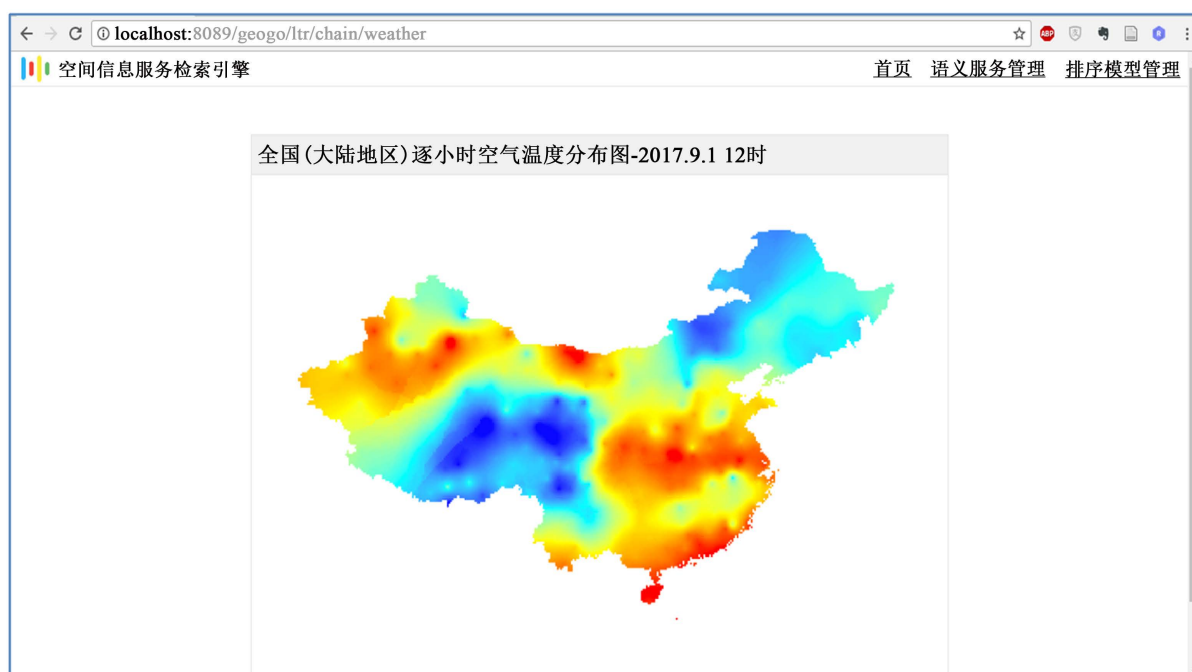


Figure 4. Distribution map of air temperature

图 4. 空气温度分布图

致谢

本文受到湖北省自然科学基金项目(2016CKB714)资助。

参考文献 (References)

- [1] Aalders, H.J.G.L. (2005) An Introduction to Metadata for Geographic Information. International Cartographic Association, Oxford, 3-27.
- [2] Yue, P., Di, L., Wei, Y. and Han, W. (2013) Intelligent Services for Discovery of Complex Geospatial Features from Remote Sensing Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, **83**, 151-164. <https://doi.org/10.1016/j.isprsjprs.2013.02.015>
- [3] 徐生磊, 李松年, 孙亚琴. 基于 CSW 的全球地表覆盖信息共享[J]. 测绘通报, 2017(1): 74-78.
- [4] 杨中, 李国庆, 于文洋, 等. 基于 CSW 的空间资源统一化发现服务设计实现[J]. 计算机工程与设计, 2014, 35(5): 1816-1822.
- [5] Stock, K., Kristin, M., Atkinson, R. and Higgins, C. (2010) A Semantic Registry Using a Feature Type Catalogue instead of Ontologies to Support Spatial Data Infrastructures. *International Journal of Geographical Information Science*, **24**, 231-252. <https://doi.org/10.1080/13658810802570291>
- [6] Stock, K., Stojanovic, T., Reitsma, F., Ou, Y., Bishr, M., Ortman, J. and Robertson, A. (2012) To Ontologies or Not to Ontologies: An Information Model for a Geospatial Knowledge Infrastructure. *Computers & Geosciences*, **45**, 98-108. <https://doi.org/10.1016/j.cageo.2011.10.021>
- [7] 王东旭, 诸云强, 潘鹏, 罗侃, 侯志伟. 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.
- [8] Lutz, M. and Klien, E. (2006) Ontology-Based Retrieval of Geographic Information. *International Journal of Geographical Information Science*, **20**, 233-260. <https://doi.org/10.1080/13658810500287107>
- [9] 许泉立, 易俊华, 杨昆. 基于地理本体的空间信息检索机制初探[J]. 测绘地理信息, 2015, 40(10): 65-68.
- [10] Fitzner, D., Hoffmann, J. and Klien, E. (2011) Functional Description of Geoprocessing Services as Conjunctive Data-log Queries. *GeoInformatica*, **15**, 191-221. <https://doi.org/10.1007/s10707-009-0093-4>
- [11] Janowicz, K., Schade, S., Bröring, A., Keßler, C. and Maué, P. (2010) Semantic Enablement for Spatial Data Infra-

structures. *Transactions in GIS*, **14**, 111-129. <https://doi.org/10.1111/j.1467-9671.2010.01186.x>

- [12] 田浩. 以用户为中心的 Web 服务发现方法及其在金融服务中的应用研究[D]: [博士学位论文]. 武汉: 武汉大学, 2014.
- [13] 杜武. 基于排序学习的空间信息服务检索关键技术研究[D]: [博士学位论文]. 武汉: 武汉大学, 2017.
- [14] 田浩, 樊红, 杜武. 基于用户社群关系的 Web 服务发现研究[J]. 通信学报, 2015, 36(10): 28-36.

知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询
2. 打开知网首页 <http://cnki.net/>
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: csa@hanspub.org