

# 基于 OGC 标准的地理信息服务语义标注研究

程文超<sup>1</sup>, 苗立志<sup>2,3</sup>, 周 亚<sup>1</sup>, 焦东来<sup>2,3</sup>

(1. 南京邮电大学 通信与信息工程学院, 江苏 南京 210003;

2. 南京邮电大学 地理与生物信息学院, 江苏 南京 210023;

3. 南京邮电大学 地理信息与生物信息处理研究所, 江苏 南京 210023)

**摘 要:** 地理信息服务数量的剧增使服务发现、集成和使用变得愈加困难, 用户也难以获取真正所需要的目标数据。如何自动、快速、准确地发现所需的地理信息数据服务, 成为地理信息服务语义检索研究中亟需解决的问题之一。然而, 目前大部分地理信息服务查询还主要是基于关键字匹配的检索, 语义歧义或语义缺失则会导致用户无法获取其真正所需的目标服务。文中基于 OGC 标准建立了 OGC 网络服务能力描述文件的语义标注方法及其流程, 并设计开发了基于语义标注的 OGC 地理信息服务语义搜索原型系统。通过验证表明, 文中语义标注方式可提高 OGC 地理信息服务检索的查准率和查全率。

**关键词:** OGC 网络服务; 属性约简; 语义标注; 地理本体; 语义查询

中图分类号: TP311; P208

文献标识码: A

文章编号: 1673-629X(2015)08-0108-05

doi: 10.3969/j.issn.1673-629X.2015.08.023

## Research on Geographic Information Services Semantic Annotation Based on OGC

CHENG Wen-chao<sup>1</sup>, MIAO Li-zhi<sup>2,3</sup>, ZHOU Ya<sup>1</sup>, JIAO Dong-lai<sup>2,3</sup>

(1. College of Telecommunications & Information Engineering, Nanjing University of Posts  
and Telecommunications, Nanjing 210003, China;

2. College of Geographical and Biological Information, Nanjing University of Posts and Telecommunications,  
Nanjing 210023, China;

3. Institute of Geographical and Biological Information Processing, Nanjing University of Posts  
and Telecommunications, Nanjing 210023, China)

**Abstract:** As geographic information services increase dramatically, it makes more difficult to discover, integrate and utilize them for users. Also, users cannot obtain the objective data really needed. So, how to get the necessary GIS data service automatically, quickly and accurately is a challenge which needs to be solved urgently in geographic information semantic search field. However, much more service search is still based on the keyword matching. Semantic ambiguity and semantic deficiency bring out that users cannot get the really needed service. In this paper, the semantic annotation method and workflow of OGC (Open Geospatial Consortium) Web Service (OWS) capability files is generated, and a prototype system based on semantic annotation is designed and developed. It is verified that semantic annotation method can enhance the precision ratio and recall ratio of OWS discovery.

**Key words:** OGC Web service; attribute reduction; semantic annotation; geographic ontology; semantic query

## 0 引 言

地理信息有着数据源广、发布形式多以及应用多元化的特点, 导致地理信息数据表达形式的多样化。为了实现地理信息资源的充分共享和广泛使用, OGC

(Open Geospatial Consortium, 开放地理空间联合会) 面向地理信息数据的集成与互操作相关研究分别提出了 WMS (Web Map Service, 网络地图服务)、WFS (Web Feature Service, 网络要素服务)、WCS (Web

收稿日期: 2014-09-30

修回日期: 2014-12-31

网络出版时间: 2015-07-21

基金项目: 国家自然科学基金资助项目(41101359, 41101358)

作者简介: 程文超(1990-), 男, 硕士研究生, 研究方向为地理信息服务智能推理; 苗立志, 博士, 副教授, 研究方向为地理信息共享与空间数据互操作、分布式地理空间信息处理; 焦东来, 博士, 副教授, 研究方向为地图符号共享。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150721.1453.070.html>

Coverage Service,网络栅格服务)、CSW(Catalog Service for the Web,网络目录服务)和 SOS(Sensor Observation Service,传感器观测服务)等多个 OWS(OGC Web Service,OGC 网络服务)实现规范。数据拥有者基于 OGC 的多种地理信息数据相关的网络服务实现规范,可以进行网络在线发布供用户使用,极大促进了地理信息数据的共享。但是,OGC 数据服务实现规范只是通过标准接口着重强调技术数据的互操作性,仍然存在空间数据共享的语义异构问题。目前,在海量的共享地理信息服务中查找自身所需的数据时,通常是基于关键字并结合空间过滤条件来查询 OGC 目录服务以获取相关的地理信息服务。由于 OGC 数据服务能力文档中缺乏明确的语义信息,致使语义歧义或语义缺失,使得用户无法获取其真正所需的服务,这也是地理空间数据自动发现与检索的主要限制之一。比如在搜索关键字“北京”时,传统搜索引擎只能从词形上进行匹配,无法检索到“中国首都”、“beijing”、“pe-king”一类具有“北京”相同意义词汇的相关信息,即传统检索方式无法理解关键字的语义,致使在海量的地理信息数据和服务中,往往会受查全率和查准率不高的困扰,亟需寻找一种方法帮助用户找到其所想查询的地理信息服务。

传统的地理信息数据侧重于表达空间实体的几何成分,忽略了其语义关系,语义标注的任务就是建立非语义服务描述与相应的语义描述的本体部分的关联,最终使信息检索方式从传统的关键字检索转变为基于标注的语义检索<sup>[1-2]</sup>。目前国内外学者已进行了一些相关研究,如 Zhao Peisheng 等<sup>[3]</sup>提出了基于语义网和本体的服务知识标注及组合策略研究地理信息数据和服务的描述和推理;Zhang Chuanrong 等<sup>[4]</sup>提出了基于本体的 WFS 发现与组合算法用于地理要素数据在语义层面的互操作;梁汝鹏<sup>[5]</sup>创建面向服务语义标注的地理本体三层架构,开发了面向语义标注的领域本体与应用本体;柳佳佳等<sup>[6]</sup>引入服务接口依赖关系提出支持接口多态性的本体语义地理信息服务输入输出匹配方法。

但是,上述研究是以非自动方式实现模式概念与地理本体语义的映射,同时也未针对元数据进行专门标注,很难满足海量地理信息服务的语义标注要求。并且,上述标注方式并没有统一的标准或规范,在实现语义标注后导致次生的共享障碍。

文中基于 OGC 语义标注标准,对 OGC 地理信息服务元数据能力描述文件进行半自动化的知识语义标注,在不破坏 OWS 能力文档原有机构的基础上实现海量 OWS 基于语义层面的检索,实现检索的查全率和查准率的双重提高。

## 1 OGC 语义标注标准

OGC 语义标注标准<sup>[7]</sup>是由 OGC 于 2008 年提出的,该标准的提出不是为了引入新的标准,而是希望拓展已有的 OGC 标准的相关概念,并像 W3C(The World Wide Web Consortium)一样,由应用级标准转变为 OGC 标准。OWS 相关实现规范提供了定位和传输地理空间数据以及调用这些方法的标准,但是使用的资源往往被限制在一个很小的用户群组中<sup>[7]</sup>。为了使地理空间数据服务资源能够被广泛地发现、访问和使用,OGC 语义标注标准为服务发布者提供了一种面向服务的元数据进行标注和描述服务的方式,即语义标注,以便于更多的用户可以读取、理解和应用。

OGC 语义标注标准针对地理信息服务的不同应用层面,分别在元数据层、数据模型层、实例层进行语义标注。通过对 OWS 的能力描述文件的属性字段和功能类型进行语义标注,可以增加服务网络可读性、有用性和语义描述性,扩大用户的地理信息数据和服务的查询和检索,同时返回的结果也更加精确,即在查全率和查准率两方面同时得到提高,语义相似度也更高。

## 2 基于 OGC 标准的语义标注

### 2.1 地理本体库建立

地理本体把有关地理学科领域的知识、信息和数据抽象成由一个个具有共识的对象,按照一定的关系组成的体系,同时进行概念化处理和明确的定义,最后进行形式化的表达<sup>[8]</sup>。地理本体描述在空间语义服务共享中起到了至关重要的作用。面向服务的语义标注的本体层次可以分为三层:需要哲学和认知科学研究人员参与建立的顶层本体、需要领域专家参与建立的领域层本体及服务发布者参与建立的应用层本体。文中应用开源项目 DOLCE<sup>[9]</sup>(Descriptive Ontology for Linguistic and Cognitive Engineering,语言学与认知工程描述型本体)作为顶层本体,领域层参考地质矿产术语分类标准 GB/T 9649.32-2009,应用层参考 OGC WMS 规范,采用 WMS 服务本体<sup>[10]</sup>,利用可视化构建本体的开源软件 Protégé<sup>[11]</sup>构建基于 OWL(Web Ontology Language,万维网本体语言)标准<sup>[12]</sup>的本体库模型,选取固体矿产普查与勘探一个分支建立对应地理本体,如图 1 所示。

如图 1 所示,金属矿产包含有色金属矿产、黑色金属矿产、稀有金属矿产等,比如有色金属矿产包括铜矿、镍矿、铝矿等,若使用有色金属矿产关键字语义查询时,将返回包括铜矿、镍矿、铝矿在内的所有要素。

### 2.2 基于粗糙集的约简信息提取

OGC 能力描述文件是以 XML 文件的形式存在的,其中的所有字段都是基于特定的 XML schema 文

件描述的,如 WMS 的能力描述文件基于最新的标准<sup>[13]</sup>中包含的字段有 Service、Capability、GetMap、Name、Title、Abstract、KeywordList 等标签。然而,能力描述文件的知识并非是同等重要的,有些甚至是冗余的,为提高信息查询的准确度,需对能力文件所包含的词汇按重要程度进行约简。文中将能力描述文件进行解析,将相关描述词汇构成属性集合,采用粗糙集中基于属性重要度约简的方式,得到最优约简属性集合。通过将最优约简属性集合的词汇与已建立的本体库本体知识进行比对,删除词汇集中冗余的知识,最终得到最约简的知识集合。

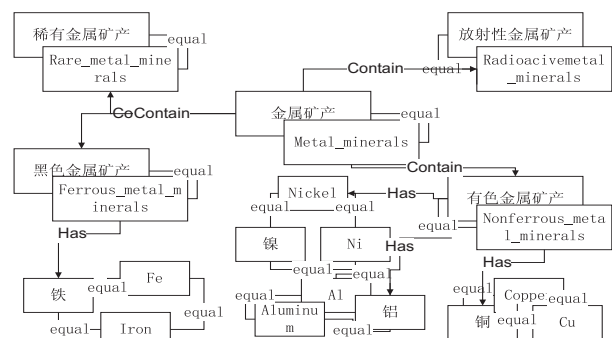


图 1 金属矿产本体库模型

基于粗糙集的知识约简的流程如图 2 所示。

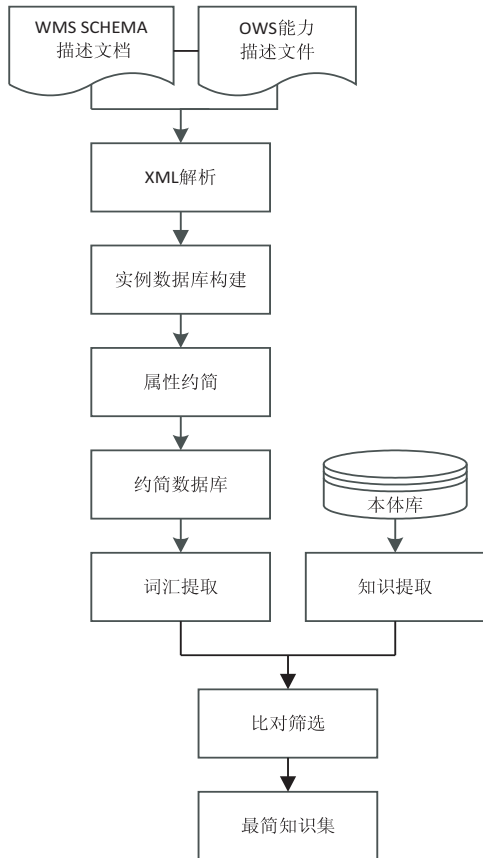


图2 约简知识提取流程

(1)解析并提取 OWS 模式文件和 OWS 能力描述文件中必需的标签建立实例数据库。

(2) 基于 Pawlak 属性重要度的粗糙集属性约简<sup>[14-16]</sup>方法对属性集进行约简,得到约简数据库。

(3) 将从约简数据库中提取到的词汇比对本体库中的知识,进行词汇的语义去同化。

(4)生成每个 OWS 能力描述文件对应的最简知识集。

基于上述方法,获取 OWS 能力描述文件中需要进行标注的知识词汇。由于文中是以基于矿产领域为例进行该方法的研究,因此,能力描述文件仅保留与该领域有关的知识。

### 2.3 语义标注

文中采用的 OWS 语义标注策略是建立在 OGC 语义标注标准中的第一应用层面,即元数据层进行标注。相关字段在各 OWS 服务实现规范中遵从其相应的模式,具有一致性,语义标注流程如图 3 所示。

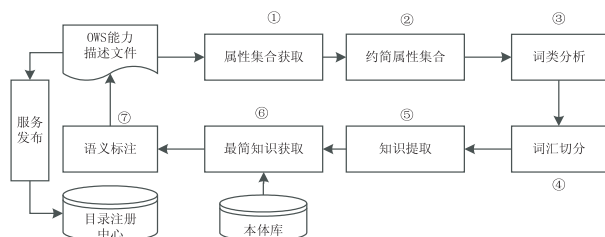


图3 语义标注流程

OWS 语义标注流程如下:

①解析 OWS 元数据能力描述文件,提取相关描述词汇组成属性集合;

②基于 Pawlak 属性重要度的属性约简算法对获取的属性集合进行约简；

③将约简的属性集合中的属性值集合进行词类分析,如果含有中文,则转入中文切词模块,否则转入标准切词模块;

④基于切词模块将属性值集合中的非单个词汇切分成单个词汇或简单词组：

⑤将切词完成后的词汇进行知识描述,组成一个知识集合:

⑥与专业领域本体库进行比对,删除在本体库中未知的知识,并将知识集合中等价的知识进行删减;

⑦将匹配成功的最简语义知识依据 OGC 标注标准对元数据能力描述文件进行标注；

⑧将语义标注后的元数据能力描述文件发布到目录注册中心。

其中,词类分析设计的目的在于识别属性值集合中是否含有中文,如果语义标注的内容含有中文,则标注的内容会涉及多种语言,那么建立本体时须考虑建立多语言的本体库,以保证能力描述文件中的字段能够被准确标注。切词模块目的在于词汇集合中的值可能不是一个词或简单词组,而是多个词或词组构成的



短语,为了保证不丢失短语的信息,对可能存在的短语基于地学词典<sup>[17]</sup>切分成多个词或简单词组。词汇集合中多数是单个词汇或简单词组,为了便于优化,需先进行预判,如果与本体库中的知识一致则直接进入标注操作,无需经过分词系统。

文中以矿产领域的相关网络地图服务为样本数据,数据来源于 USGS(美国地质勘探局)提供的在线空间数据<sup>[18]</sup>。下面是依据上述流程对一个 WMS 能力描述文件进行标注前后的示例,如图 4、5 所示。

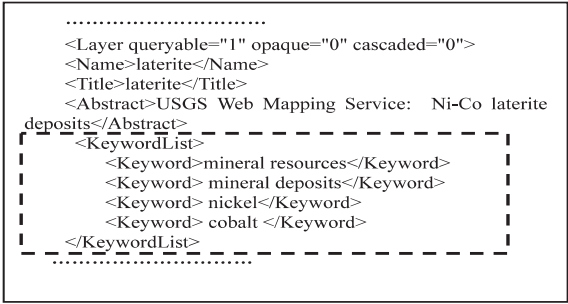


图 4 标注前 OGC 能力描述文件

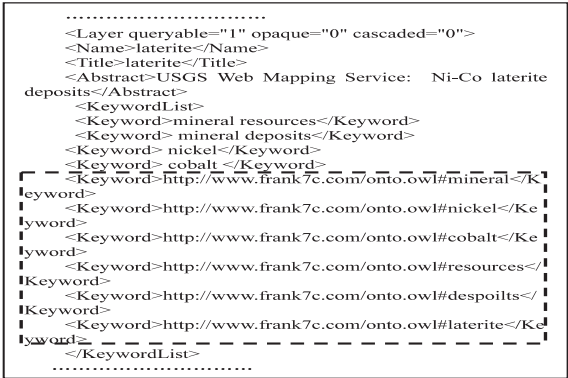


图 5 标注后 OGC 能力描述文件

标注后,Keyword 标签中的一部分字段指向领域本体库,使得相关字段具备本体中所描述的语义关系,能够使该能力描述文件不仅支持关键字查询,同时也支持语义推理。

3 原型系统实现

3.1 体系结构

为实现对 OGC 地理信息服务的语义标注和语义检索,文中设计了基于 OGC 语义标注标准的地理信息服务语义检索系统框架体系结构及其工作流程,如图 6 所示。

- 整个体系结构分为三层:
- (1)用户界面层:是该体系结构与用户交互的部分,主要包括查询模块和可视化模块。主要用于用户构建满足自身需求的特定查询条件,以及对响应结果 OGC 网络服务数据的基于 OpenLayers 的二维可视化。
  - (2)事务处理层:是该体系结构的核心部分,主要处理来自用户的查询请求、调用底层的数据及返回相

关检索结果数据服务,包括了目录注册中心和语义查询引擎。其中目录注册中心用于保存可用服务的相关数据服务信息,语义查询引擎是基于开源软件 Jena<sup>[19-20]</sup>和 Lucene<sup>[21-22]</sup>进行语义推理的查询模块。

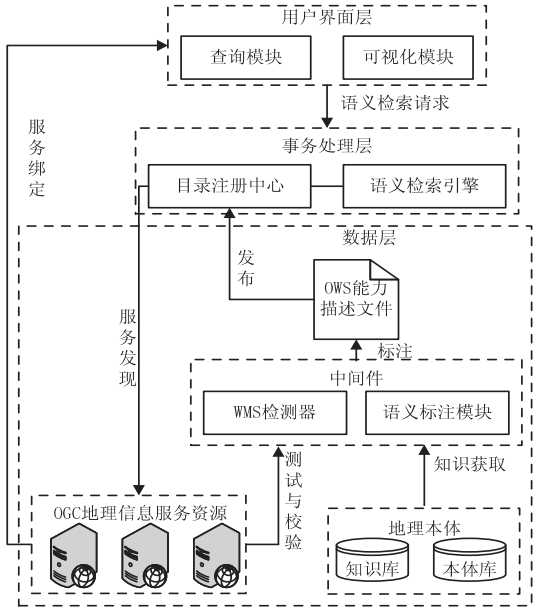


图 6 基于 OGC 语义标注标准的语义查询系统框架

(3)数据层:包含了地理信息资源层、地理信息本体和元数据层,是整个系统的基础层。数据层中内嵌了一个中间层<sup>[23]</sup>,包含语义标注工具和 WMS 检测器。WMS 检测器通过比对时间戳的方式保证同一个 WMS 服务最新。

3.2 应用实例

基于上述体系结构,依据 OGC 标准语义标注方式,构建了面向 OGC 地理信息服务的语义搜索原型系统。用户可以在搜索条件设置区域输入相关条件,如关键字“有色金属矿产”并结合空间过滤条件“美国”进行语义查询的结果,最终在服务目录注册中心发现 4 个 OGC WMS 数据文件,并通过能力描述文件加载显示相应的图层。

检索“有色金属矿产”时,若基于关键字的查询,只能查找到包含“有色金属矿产”词汇匹配的相关地理信息服务数据,无法得到其语义信息相匹配的数据,返回结果数据条目为 0 个。而本例中通过基于语义标注的搜索原型系统对“有色金属矿产”进行检索时,解析本体知识,获取“有色金属矿产”的语义知识,即“有色金属矿产”包含“镍矿”“铜矿”“铝矿”以及“Nickel”“Copper”“Aluminum”等,其语义检索返回结果数据条目为 4 个,从而能够根据语义信息查询到所需要的地理信息服务数据,在语义层面提高了查全率和查准率。

4 结束语

针对日益增多的 OWS 数据查全率和查准率不高

的问题,文中提出了一种基于 OGC 标准的语义标注方法,对 OWS 能力描述文件进行半自动化的标注,对现有基于关键字的 OWS 检索进行改进,解决了语义标注在检索中文时无法获取相应英文服务的问题,即多语言社区问题;解决了基于关键字匹配检索带来的回调率不高的问题,即多词同义的问题。

文中设计开发的语义标注工具可作为现有目录注册中心的中间件,其作用是标注现有的 OWS 能力描述文件,使其具备语义匹配和发现的能力,提高服务发现的准确率,有效解决 OGC 地理信息服务淹没的问题和语义异构的问题,可作为地理信息服务语义检索的一种有益参考。

参考文献:

[1] Gurrin C, He Yulan, Kazai G, et al. Recent developments in information retrieval[ C ]//Proceeding of 32nd European conference on IR research. [ s. l. ]:[ s. n. ],2010.

[2] 梁汝鹏,李宏伟,李文娟,等. 空间语义学与地理信息语义服务研究进展[J]. 测绘科学,2013,38(6):19-22.

[3] Zhao Peisheng, Di Liping, Yu Genong, et al. Semantic Web-based geospatial knowledge transformation[ J ]. Computers & Geosciences,2009,35(4):798-808.

[4] Zhang Chuanrong, Zhao Tian, Li Weidong, et al. Towards logic-based geospatial feature discovery and integration using web feature service and geospatial semantic web[ J ]. International Journal of Geographical Information Science,2010,24(6):903-923.

[5] 梁汝鹏. 基于语义标注的地理信息服务关键技术研究[ D ]. 郑州:解放军信息工程大学,2013.

[6] 柳佳佳,葛文. 基于本体语义的地理信息服务发现[ J ]. 测绘工程,2013,22(6):9-13.

[7] Maue P, Schade S, Duchesne P. Semantic annotations in OGC standards[ S/OL ]. 2012. [http://portal.opengeospatial.org/files/?artifact\\_id=34916](http://portal.opengeospatial.org/files/?artifact_id=34916).

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

[9] Masolo C. Wonder Web project: D18 ontology library[ EB/OL ]. 2003-01-12. <http://wonderweb.man.ac.uk/>.

[10] 梁汝鹏,李宏伟,李文娟,等. 基于知识标注的地理信息语义服务框架研究[ J ]. 地理与地理信息科学,2012,28(3):1-6.

[11] Stanford Center for Biomedical Informatics Research. Protégé[ CP/OL ]. 2011. <http://protege.stanford.edu/>.

[12] 胡鹤,刘大有,王生生. Web 本体语言 OWL[ J ]. 计算机工程,2004,30(12):1-2.

[13] WMS schema capabilities[ EB/OL ]. 2014. [http://schemas.opengis.net/wms/1.3.0/capabilities\\_1\\_3\\_0.xsd](http://schemas.opengis.net/wms/1.3.0/capabilities_1_3_0.xsd).

[14] 苗夺谦,李道国. 粗糙集理论、算法与应用[ M ]. 北京:清华大学出版社,2008.

[15] Stoilos G, Stamou G, Pan J Z. Fuzzy extensions of OWL: logical properties and reduction to fuzzy description logics[ J ]. International Journal of Approximate Reasoning,2010,51(6):656-679.

[16] 纪滨. 粗糙集理论及进展的研究[ J ]. 计算机技术与发展,2007,17(3):69-72.

[17] DWiKi: earth science dictionary[ DB/OL ]. 2014. <http://www.geology.com.cn/wiki/>.

[18] Mineral resources spatial data: OGC WMS services[ EB/OL ]. 2013. <http://mrdata.usgs.gov/wms.html>.

[19] Apache: Jena ontology api[ CP/OL ]. 2014. <http://jena.apache.org/documentation/ontology>.

[20] 谢谏. 基于 Lucene 的 XML 索引与检索[ D ]. 广州:华南理工大学,2011.

[21] 何来坤,缪健美,刘礼芳,等. 基于 Ontology 与 Jena 的研究综述[ J ]. 杭州师范大学学报:自然科学版,2013,12(5):467-473.

[22] 郑榕增,林世平. 基于 Lucene 的中文倒排索引技术的研究[ J ]. 计算机技术与发展,2010,20(3):80-83.

[23] Su Xiaolu, Meng Xianxue, Hu Haiyan. Research on development and application of OWL ontology search middleware[ J ]. Mathematical and Computer Modelling,2013,58(3-4):614-618.

(上接第 107 页)

alignment in multi-secondary users cognitive radio system[ C ]//Proc of IEEE ISCC. Cappadocia, Turkey: IEEE,2012:785-789.

[10] Du H, Ratnarajah T. Robust joint signal and interference alignment for MIMO cognitive radio network[ C ]//Proc of IEEE WCNC. Paris, France: IEEE,2012:448-452.

[11] Mitola J, Maguire G Q. Cognitive radio: making software radios more personal[ J ]. IEEE Pers Comm,1999,6(4):13-18.

[12] Khaled N, Thoen S, Deneire L. Optimizing the joint transmit

and receive MMSE design using mode selection[ J ]. IEEE Trans on Comm,2005,53(4):730-737.

[13] Sampath H, Stoica P, Paulraj A. Generalized linear precoder and decoder design for MIMO channels using the weighted MMSE criterion[ J ]. IEEE Trans on Comm,2001,49(12):2198-2206.

[14] Yetis C M, Gou T, Jafar S A, et al. On feasibility of interference alignment in MIMO interference networks[ J ]. IEEE Trans on Signal Pro,2010,58(9):4771-4782.

基于OGC标准的地理信息服务语义标注研究

作者：

[程文超](#)，[苗立志](#)，[周亚](#)，[焦东来](#)，[CHENG Wen-chao](#)，[MIAO Li-zhi](#)，[ZHOU Ya](#)，[JIAO Dong-lai](#)

作者单位：

[程文超, 周亚, CHENG Wen-chao, ZHOU Ya \(南京邮电大学 通信与信息工程学院, 江苏 南京, 210003\)](#)，[苗立志, 焦东来, MIAO Li-zhi, JIAO Dong-lai \(南京邮电大学 地理与生物信息学院, 江苏 南京 210023; 南京邮电大学 地理信息与生物信息处理研究所, 江苏 南京 210023\)](#)

刊名：

[计算机技术与发展](#)

英文刊名：

[Computer Technology and Development](#)

年，卷(期)：

2015(8)

引用本文格式：[程文超](#). [苗立志](#). [周亚](#). [焦东来](#). [CHENG Wen-chao](#). [MIAO Li-zhi](#). [ZHOU Ya](#). [JIAO Dong-lai](#) [基于OGC标准的地理信息服务语义标注研究](#)[期刊论文]-[计算机技术与发展](#) 2015(8)