



测绘地理信息科技出版资金资助

CEHUI DILI XINXI KEJI CHUBAN ZIJIN ZIZHU

RESEARCH ON INTERPRETING MECHANISM OF NATURAL SPATIAL QUERY LANGUAGE

李 滨 著

自然空间查询语言 解译机制研究



测绘出版社

测绘地理信息科技出版资金资助

自然空间查询语言解译机制研究

Research on Interpreting Mechanism of
Natural Spatial Query Language

李 滨 著



测绘出版社

· 北京 ·

© 李滨 2012

所有权利(含信息网络传播权)保留,未经许可,不得以任何方式使用。

内容简介

本书是作者攻读博士学位期间针对自然空间查询语言的研究成果,包括空间查询的分类及分析、自然空间查询语言空间知识(空间词典知识、空间句法知识、空间语义知识)梳理及空间知识库的构建、基于理解的空间分词算法、基于规则的句法分析算法、基于框架语义的空间语义分析算法、基于 SQL 扩展的空间对象存储与访问等。通过本书,读者能够理解自然空间查询语言的解译机制并掌握实现这一机制的核心技术。

本书可作为地理信息科学专业研究生及高年级本科生的教材或教学参考书,也可作为本领域相关理论、技术和应用科研人员的参考书。同时,本书对自然语言等相关领域科研教学人员也具有一定的参考价值。

图书在版编目(CIP)数据

自然空间查询语言解译机制研究 / 李滨著. — 北京: 测绘出版社, 2012.12

ISBN 978-7-5030-2717-8

I. ①自… II. ①李… III. ①自然语言处理—研究
IV. ①TP391

中国版本图书馆 CIP 数据核字(2012)第 222043 号

责任编辑	贾晓林	封面设计	李 伟	责任校对	董玉珍	责任印制	喻 迅
出版发行	测绘出版社			电 话	010—83060872(发行部)		
地 址	北京市西城区三里河路 50 号				010—68531609(门市部)		
邮政编码	100045				010—68531160(编辑部)		
电子信箱	smp@sinomaps.com			网 址	www.chinasmp.com		
印 刷	北京天顺鸿彩色印刷有限公司			经 销	新华书店		
成品规格	169mm×239mm						
印 张	9.25			字 数	176 千字		
版 次	2012 年 12 月第 1 版			印 次	2012 年 12 月第 1 次印刷		
印 数	0001—1200			定 价	28.00 元		

书 号 ISBN 978-7-5030-2717-8/P·612

本书如有印装质量问题,请与我社门市部联系调换。

序

自然空间查询语言是自然语言学与地理信息科学交叉融合的新兴研究领域。作为实现空间信息查询的重要语言形式,自然空间查询语言使得人们可以通过自然语言(如汉语)形式查询所需的地理空间信息。这不仅符合人们的认知习惯和语言习惯,而且可以实现人机的紧密结合,使地理信息系统更具智能性。自然空间查询语言解译机制的研究就是要解决从自然语言中自动抽取空间信息、消除结构化空间信息与自然语言符号化空间信息之间的语义障碍这两道难题,其相关理论研究具有科学性、方法实践具有应用性,因此是一项充满挑战却又很具实用价值的工作。从研究现状来看,自然空间查询语言解译机制的研究仍处在起步阶段,理论与技术体系尚未建立,相关应用成果也不够成熟。本书系统阐述了自然空间查询语言的解译机制,揭示了其空间分词、句法分析、语义分析、用户空间查询意图解译、查询结果反馈等完整的解译过程。因此,本书是关于自然空间查询语言解译机制研究的一本不错的学术专著。

本书系统地梳理了地理信息系统领域目前常见的空间查询分类方式并分析了各种分类方式的相关关系,这使得有关空间查询的研究更加条理化;归纳和构建的地理实体名称、空间词汇和查询词汇等三类空间词典知识为自然空间查询语言的研究提供了坚实的词汇基础;总结描述的句法规则及常用空间查询句型模板不仅提高了空间分析的效率还有效支撑了句法分析;采用框架语义对自然空间查询语言涉及的内部空间语义、外部空间语义和空间查询语义等三类空间语义的详细描述以及基于这些框架空间语义设计的语义分析算法,有效解决了自然空间查询语言解译中语义解译这一核心问题;基于 SQL 扩展了空间数据类型和空间函数,实现了地理空间信息的对象存储与访问。这些研究内容均具有很高的学术价值和创新性,这都是在国内外学者的专著中很少看到的。

作者历时六年专注于将语言学、人工智能学等其他领域的理论与方法引进到地理信息系统领域,以解决自然空间查询语言这一专业难题的工作精神是十分值得肯定的。与此同时,作为一名高校教师,将自己的研究方向融入到本科学生的教

学实践中,培养出多名优秀本科毕业生,这种教学相长的态度也是值得赞许的。本人相信,本书的出版能够为地理信息系统领域学者,特别是自然空间查询语言的相关理论、技术和应用科研教学人员提供很高的参考价值,同时对相关领域的科研教学人员也具有一定的指导意义。

A handwritten signature in black ink, appearing to read '王宪' (Wang Xian), written in a cursive style.

中国工程院院士

2012年4月5日

前 言

随着地理信息科学与计算机科学的不断发展与融合,空间数据库和地理信息系统(geographic information system, GIS)在理论及技术领域都取得了长足的发展。尤其是随着 Google 地图服务、基于 Web 2.0 的基于位置的服务、导航定位服务等一系列标志性空间信息技术的广泛应用,标志着空间信息时代已悄然来临。空间信息服务已从原有的面向专业地理信息系统人员发展到现在的面向公众,而要进一步推广空间信息服务应用,就需要考虑普通公众的知识背景,提高各种信息服务的易用性。空间查询是各种空间信息服务的基础性功能之一,自然空间查询语言则是实现空间查询的重要语言形式。自然空间查询语言由于其使用方法简单、空间需求描述符合人们语言习惯的特性引起了研究人员的关注,针对自然空间查询语言的研究已成为当前地理信息科学的一大研究热点。自然空间查询语言的解译是自然空间查询语言得以实现和应用的前提,解译过程涉及空间知识、空间分词、句法分析、语义分析等众多研究内容,是一项很具挑战性的工作。本书正是在这样的背景下对汉语自然空间查询语言解译机制进行系统的研究,向读者揭示汉语自然空间查询语言从空间分词、句法分析、语义分析,到实现空间查询函数,最终给用户提供了所需查询结果的完整的解译过程。

本书共分 6 章:第 1 章为绪论,第 2 章至第 6 章是重点内容。具体安排如下:

第 1 章绪论。介绍了自然空间查询语言解译机制研究的背景和核心问题,总结了目前自然空间查询语言相关研究的现状和研究内容。

第 2 章空间查询与自然空间查询语言。本章首先在对空间查询的分类方式进行总结分析的基础上,针对目前分类方式角度单一、分类标准不清晰等问题,从空间概念、地理信息、查询语言和表现形式四个角度对空间查询内容进行了总结和分类,并分析了分类方式之间的关系。然后总结了自然空间查询语言的文法规则,在此基础上对自然空间查询语言的受限性进行了研究,并制订了少量句法限制规则。最后介绍了自然空间查询语言的解译过程。

第 3 章空间知识库。本章对自然空间查询语言的核心支撑——空间知识和空间知识库作了详细的论述。具体内容包括:系统归纳了地理实体名称、空间词汇和查询词汇等空间词典知识;定义了常用的空间查询句型模板及句法规则;采用框架语义的方式对自然空间查询语言涉及的空间语义进行了详细描述。

第 4 章空间分词、句法分析及语义分析。本章详细介绍了自然空间查询语言解译的重要步骤——空间分词、句法分析及语义分析。研究内容包括:在说明空间分词规范的基础上研究了基于字符串匹配、基于理解和基于统计的分词算法思

想,详细介绍了基于空间查询模板和空间词典的空间分词和词性标注算法以及未登录词的识别;在总结分析基于规则和基于统计的句法分析方法及其优缺点对比的基础上,详细介绍了基于规则的句法分析算法——线图分析算法;在介绍基于框架语义的自然空间查询语言的空间语义分析过程的基础上,通过实例介绍了空间语义框架模拟人的理解方式进行语义分析的过程。

第5章基于SQL扩展的空间对象存储与访问。本章对自然空间查询语言的底层空间数据库支撑进行了介绍。具体内容包括:介绍了关系数据库对面向对象技术的支持以及基于这些技术对SQL进行空间数据类型和空间函数扩展的方法;详细论述了基于OpenGIS简单要素规范的几何对象模型对SQL进行空间数据类型和空间函数扩展的方法;说明了利用基于SQL扩展的空间数据类型和空间函数构建对象关系空间数据库进而实现空间数据的对象存储与访问的过程。

第6章自然空间查询原型系统实现与评测。本章介绍了自然空间查询系统实现的硬软件环境及实验数据,并展示了自然空间查询系统的部分功能,以及对分词正确率等各种性能指标进行了分析评测。

本书系统归纳的空间知识不仅可以应用于自然空间查询语言研究,也可以用于自然空间关系解译、空间决策支持系统等领域的研究。自然空间查询语言解译机制的细致探讨,为自然空间查询语言的实现提供了一种解决方案。本书作为学术专著主要适用于对自然空间查询语言的相关理论、技术和应用进行研究的科研教学人员。由于自然空间查询语言的研究涉及自然语言解译、空间语义分析、空间数据存储等方面的研究,对相关领域的科研教学也具有一定的参考价值。

此书是在本人博士论文的基础上修改完成的。回首六年博士求学之路,可谓颇多感慨。感慨选题之不易,感慨研究之坎坷,感慨写作之艰辛。在此过程中如果没有师长的鼓励、家人的关心、同事的支持和学生的帮助,顺利完成论文是不可想象的,本人谨对他们表示衷心的感谢。尤其要感谢我的恩师王家耀院士(博士生导师)和王青山教授(硕士生导师)传授给我做人的道理和做学问的态度;感谢崔铁军教授、王光霞教授、张文诗教授对本书研究的关注与支持;感谢信息工程大学测绘学院我所指导的学生们,他们为本书的研究积累了丰富实验成果,在指导他们学习的过程中,我深刻地体会到了教学相长的意义。本书的出版得到了“2110工程:作战环境学重点建设学科领域”专著出版经费以及测绘地理信息科技出版资金的资助,特于此谨致我的谢意。

在这里还要提前感谢将来阅读本书的同行们,希望你们能对书中的错漏及不足之处提出宝贵意见(本人邮箱:zb35libin@sina.cn)。毕竟本书的完成不是研究的结束,而是新阶段研究的开始,希望与大家的交流和共同努力,为本领域的研究尽绵薄之力。

目 录

第 1 章 绪 论	1
1.1 自然空间查询语言解译机制研究的背景与核心问题	1
1.2 自然空间查询语言解译机制的研究现状、内容及目标	2
第 2 章 空间查询与自然空间查询语言	10
2.1 空间查询的概念	10
2.2 空间查询的分类	10
2.3 自然空间查询语言及其受限性	19
2.4 自然空间查询语言的解译过程	25
2.5 本章小结	28
第 3 章 空间知识库	30
3.1 知识及知识库	30
3.2 空间查询知识及知识库	32
3.3 空间查询词典	35
3.4 空间查询句型模板库	41
3.5 空间语义库	45
3.6 本章小结	74
第 4 章 空间分词、句法分析及语义分析	75
4.1 空间分词	75
4.2 基于句法规则的句法分析	84
4.3 基于框架语义的语义分析	93
4.4 本章小结	94
第 5 章 基于 SQL 扩展的空间对象存储与访问	96
5.1 对象关系空间数据库	96
5.2 OpenGIS 规范几何对象模型	97
5.3 基于 ADT 的 OpenGIS 几何对象类描述	99
5.4 基于 SQL 的空间数据类型和空间函数的扩展	104

5.5	基于 SQL 的空间数据类型和空间函数扩展及应用实例·····	108
5.6	本章小结 ·····	116
第 6 章	自然空间查询原型系统实现与评测·····	117
6.1	系统实验环境 ·····	117
6.2	原型系统功能 ·····	117
6.3	各种指标评测 ·····	124
6.4	本章小结 ·····	130
参考文献 ·····		131

Contents

Chapter 1 Introduction	1
1.1 Background and Key Problems of Research on Interpreting Mechanism of Natural Spatial Query Language	1
1.2 Current Situation and Contents of Research on Interpreting Mechanism of Natural Spatial Query Language	2
Chapter 2 Spatial Query and Natural Spatial Query Language	10
2.1 Concepts of Spatial Query	10
2.2 Classification of Spatial Query	10
2.3 Natural Spatial Query Language and Its Restriction	19
2.4 Interpreting Process of Natural Spatial Query Language	25
2.5 Chapter Summary	28
Chapter 3 Spatial Knowledge Base	30
3.1 Knowledge and Knowledge Base	30
3.2 Spatial Query Knowledge and Spatial Knowledge Base	32
3.3 Spatial Query Dictionary	35
3.4 Sentence Patterns Base of Spatial Query	41
3.5 Spatial Semantic Base	45
3.6 Chapter Summary	74
Chapter 4 Spatial Segmenting, Parsing and Semantic Analyzing	75
4.1 Spatial Segmenting	75
4.2 Parsing Based on Grammar Rules	84
4.3 Semantic Analyzing Based on Frame Semantic	93
4.4 Chapter Summary	94
Chapter 5 Storing and Accessing of Spatial Objects Based on SQL Extended	
.....	96
5.1 Object-relation Spatial Database	96

5.2	OpenGIS Geometry Object Model	97
5.3	Describing of OpenGIS Geometry Object Classes Based on ADT	99
5.4	Spatial Data Types and Spatial Functions Extended by SQL	104
5.5	Application Instances of Spatial Data Types and Spatial Functions Extended by SQL	108
5.6	Chapter Summary	116
Chapter 6 Realizing and Evaluating of Prototype Natural Spatial Query System		116
6.1	System Experiment Environment	117
6.2	Functions of Prototype System	117
6.3	Evaluating Several Performance Indexes	124
6.4	Chapter Summary	130
References		131

第1章 绪 论

1.1 自然空间查询语言解译机制研究的背景与核心问题

1.1.1 研究背景

人类最关心的两个基本物质向量是空间和时间,生活中的每一天都离不开空间位置的需求(景贵飞 等,2004)。据统计,人类所拥有信息的80%都与空间信息相关。空间信息的应用涉及人类生产生活的各个领域,不仅在国土资源、测绘、城市规划、地质勘探等传统空间信息需求领域有广泛而深入的应用,而且不断在农业、环保、交通、气象、公安、电力、电信等领域得到广泛应用。空间信息的应用逐渐趋于社会化,也正日益成为我们日常生活的一部分。作为我们克服空间障碍、了解所处空间信息及出行辅助的重要工具,定位服务得到了迅速发展和应用,并具有良好的成长性。特别是在人类活动范围不断扩大的情况下,导航定位产品已经成为大众化的消费电子产品,融合互联网、无线移动通信定位和GIS的基于位置的服务(location based service, LBS)也在现代生活中发挥越来越大的作用。

空间查询是人们获取空间位置信息最重要也是最直接的方式。用户提出的关于空间信息的问题都是以空间查询的形式表现的。因此,不论是GIS还是空间数据库,空间查询都是其最基本和使用最频繁的功能(李滨 等,2003)。作为空间查询的主要手段,空间查询语言成为GIS和空间数据库的一个核心要素。空间查询语言不仅是用户方便地访问、查询和处理空间数据的重要支撑,也是实现空间数据的安全性和完整性控制的重要途径(黄波 等,1999)。

空间查询语言的主要语言形式有扩展关系数据库查询语言和自然语言两种。基于结构化查询语言(structured query language, SQL)及SQL扩展的人工空间查询语言由于其面向问题和结构化的良好特征,适合表达复杂的空间查询,被认为是目前解决空间查询的最好方式(杨林 等,2007)。但扩展关系数据库查询语言是结构化很强的查询语言,使用扩展关系数据库查询语言语句必须具有关系数据库和GIS等相关领域的专业知识,适合GIS和空间数据库软件专业开发和应用人员使用,不易被普通用户理解和操作。反观自然语言形式的空间查询语言,目前受到汉语的复杂词汇、语法以及解译困难的限制,往往在一些情况下只作为其他查询方式的补充。但自然语言最符合人们的认知习惯和语言习惯,能提供最为理想的空间查询方式:用户只需用某种自然语言(如汉语)来表达自己的操作意图,然后由空间信息系统自动解译用户的意图并执行相应的空间操作,最后以自然语言或地图的

方式给出查询结果,即通过自然空间查询语言进行空间查询。自然空间查询语言使得用户更容易表达自己的空间查询需求,对空间查询语句的理解更加容易,操作更加简便。随着自然语言学相关研究的不断深入,汉语自动分词技术、句法分析和语义分析方面取得了长足进步,本书正是在这一背景之下对自然空间查询语言的解译机制进行研究和探讨。需要说明的是,本书所指的自然空间查询语言特指汉语自然空间查询语言。

1.1.2 核心问题

自然空间查询语言由于具备符合人们认知习惯和语言习惯的自然语言特质,势必会随着智能 GIS 的不断发展,实现用户与 GIS 自然语言交互这一最终目标。自然空间查询语言实现用户与 GIS 的自然语言交互需要解决以下两个核心问题。

(1) 自然空间查询语言中空间信息的识别问题。空间信息识别是指从自然空间查询语句(可视为一种简化自然语言文档)中找到空间信息的过程,是实现自然空间查询语言的首要核心问题。空间信息识别问题解决的关键支撑是空间查询词汇以及一系列空间信息识别规则(或模式)等空间知识。自然空间查询语言空间信息识别问题的解决可以借鉴自然语言处理领域中信息抽取(information extraction, IE)技术,尤其是可以重点借鉴从自由式文本(如新闻稿件等)中找到特定信息的信息抽取技术。

(2) GIS 结构化空间信息与自然语言非结构化空间信息的语义障碍问题。自然语言中的空间信息是以空间词汇或者特定句法模式等非结构化的形式存在的。例如在“离测绘学院最近的医院”空间查询语句中,“最近”这一空间词汇就蕴含着两个空间实体间距离最小的空间信息;在“测绘学院一公里范围内的医院有哪些”空间查询语句中地理实体(“测绘学院”) + 数词(“一”) + 量词(“千米”) + 空间词汇(“范围内”) + 实体种类(“医院”)这一特定句法模式则蕴含着叠置分析的空间信息。而 GIS 中的空间信息是以空间坐标(如经纬度)、空间函数(如缓冲区分析)等结构化的形式存在。这两种形式的空间信息之间转换存在的最大障碍就是语义问题。解决这一问题需要在空间知识的支撑下进行空间语义的分析和转换。

以上两个问题正是自然空间查询语言解译机制必须研究和解决的核心问题。

1.2 自然空间查询语言解译机制的研究现状、内容及目标

1.2.1 研究现状

1. 数据库中文自然语言查询研究现状

数据库自然语言查询是自然语言研究的一个重要分支领域。旨在便于非数据库专业人员更加方便、简单和自然地使用数据库管理、存储、分析和应用其领域数

据,智能数据库需要一种便于非数据库专业人员理解和使用的查询语言,即数据库自然查询语言。数据库自然查询语言及其实现自然查询的接口已经受到了数据库领域研究人员的广泛关注和研究。国外该领域的研究起步于20世纪60年代,经过长期的研究和应用实践取得了很多进展,最为大家所知的就是微软SQL Server数据库6.5和7.0提供的English Query。它是一个基于英语自然语言的数据库查询接口。中文数据库自然语言查询的研究起步于20世纪70年代末,经过近40年的发展,在语言模型、语言受限性、语言接口、SQL转换等相关理论和应用领域都取得了很大的进展,这些进展都对自然空间查询语言的研究有着很高的借鉴价值。现将该领域的研究成果汇总如下。

吴照林等(1991)在对大量自然查询语料进行分析研究的基础上,提出了基于结构分析的语言模型(concept dependence on structure analysis, CDSA),并推出了基于该模型的关系数据库自然语言接口实验软件(natural language query interface, NLQI)。

孟小峰等(1997, 1999, 2001, 2001a, 2001b)对中文数据库自然语言查询和界面(natural language interface to database, NLIDB)进行了大量研究,提出了适合中文数据库自然语言查询处理的很多方法,包括数据库查询语言的中文表述方式、多语句查询处理技术、中文数据库自然语言接口的领域知识、受限自然查询语言语义依存树向SQL的转化算法等,设计实现了中文数据库自然语言查询系统。

许龙飞等(1997, 1998, 1999, 2002)对基于受限汉语的数据库自然查询接口(natural language(Chinese) query interface, NLCQI)技术进行了大量研究,设计了从受限汉语查询语句向SQL语句的转换策略,提出了NLCQI的系统模型和框架,并通过相关系统实验证明了基于受限的汉语查询语句表达方式更为自然。

胡晓军(2006a, 2006b)对数据库汉语接口进行了较深研究,总结了基于自然语言理解的数据库查询接口的特点、优势、发展情况和现状,对数据库汉语接口的工作原理进行了论述,尤其是对分词算法及消歧、中间语言和生成SQL语句等重点内容进行了研究,设计了一个数据库自然语言汉语查询接口模型。

赵剑威等(2005)、郑逢斌等(2004, 2007)对数据库自然语言查询条件及祈使句、特指问句、是非问句、正反问句、选择问句等五类句型进行了研究,建立并研究了复合概念、逻辑概念与标准概念的内在联系,将查询目标概念分解为直接查询目标、逻辑推理目标和比较判断目标三个阶段,研究了这三个阶段的关系。

顾国良等(1990)、吕光楣等(1991)、黄祥喜等(1993)、赵元正(1999)、李保利等(1999a, 1999b)、余正涛等(2004)、于龙(2005)、谢晓飞等(2006)、胡明耀(2006)、韩涛等(2008)对数据库汉语自然语言查询接口进行了大量研究和实验,在分词算法、语义分析及应用方面取得了很多进展。

较为成熟的数据库自然语言查询系统除了上面介绍的由武汉通信指挥学院吴

照林、高广峰设计的基于 CDSA 模型的 NLQI,中国人民大学孟小峰、王珊等提出的中文数据库自然语言查询系统(NCHQL),暨南大学许龙飞设计的基于受限汉语的数据库自然查询接口(NLCQI)等以外,还包括北京大学崔宗军、唐世渭等提出的基于受限汉语和 E-R 语义模型的数据库查询语言原型 RCHQL(崔宗军等,2000),河南大学高校信息自然语言查询系统(张占英 等,2005)等。尤其是目前在国家自然科学基金重点项目支持下,中国人民大学、香港中文大学和北京大学开展了“中文数据库系统及其语言和界面”研究,在查询语言分析、受限处理、界面管理和领域知识自动提取等多个领域进行深入的研究,取得了很大进展。

2. 基于 SQL 扩展的空间查询语言研究现状

结构化查询语言自 1974 年提出后,经过不断发展和应用已成为实现关系数据库核心理论——关系模型的重要途径。SQL 所使用的数据定义语言(data definition language, DDL)、数据操纵语言(data manipulation language, DML)、数据查询语言(data query language, DQL)和数据控制语言(data control language, DCL)充分实现了关系数据模型中关系代数、关系演算、范式等众多理念,成就了关系数据库应用的辉煌。基于 SQL 扩展的空间查询语言研究起步于 20 世纪 90 年代,20 多年的研究取得了很大的进展。基于 SQL 扩展的空间查询语言是空间查询语言的一种重要的语言形式,它的研究进展与成果对于自然空间查询语言的研究起到很好的借鉴作用。

Egenhofer(1994)对 SQL 在空间应用方面的扩展进行了探讨,指出空间数据查询语言(geometric structured query language, GSQL)的基本要求就是要采用更贴近人们对空间理解的概念,为空间数据提供更高层次的抽象。

黄波(1996)总结了语气算子、模糊化算子和判定化算子等模糊单词,提出了模糊查询的扩展 SQL 表达,详细介绍了模糊化查询扩展 SQL 的两种实现方法。他还与林琿(1999)合作对 POINT、IMAGE 等简单空间数据类型和空间操作算子进行了总结,提出了将标准 SQL 扩展成 GeoSQL 的空间查询表达形式,设计了便于用户操作的交互式 GeoSQL 界面及 GeoSQL 的实现过程。

宋金刚等(2000)介绍了基于特征的空间查询语言(feature-based spatial query language, FSQL)的基本功能,定义了 FSQL 的语法,尤其是对多种选择方式的描述,简述了 FSQL 到任意数据库查询语言(x-query language, XQL)的映射。

郑玉明等(2002)设计了一个面向用户的 GSQL 语言,对 SQL 扩充了分别表示点、线、面以及复合体类型的四种空间数据类型,介绍了 OpenGIS 规范对空间关系和空间分析等空间算子的描述,简单构造了一个 GSQL 查询语句。

刘瑜等(2005)定义了栅格点(pixel)、栅格区域(rasterRegion)、栅格覆盖(rastercoverage)三个与栅格相关的抽象数据类型,基于此设计了支持栅格数据访问的 GSQL-R,用关系数据库扩展的用户自定义类型对栅格抽象数据类型进行了

扩展,并作了查询试验。

黄舟等(2006)说明了空间数据分布式查询流程及 GSQL 的全局解析模型,详细介绍了 GSQL 全局解析的基本方法,涉及查询树构造、GSQL 语句优化器、GSQL 解析器,并作了解析实例。

万波等(2006)介绍了在标准 SQL 的基础上引入扩展空间数据类型和空间操作函数的 GSQL 以实现空间特征的查询,对 GSQL 的查询过程和解译器的设计进行了说明,详细介绍了 GSQL 解译器的词法分析器和句法分析器的设计,并在此基础上进行了空间查询实验。

余亮等(2006)将原生 XML(extensible markup language)数据查询语言进行空间扩展得到 XML-GSQL,支持 XML 路径、表达式、空间对象、空间运算及空间关系符,描述了 XML-GSQL 的查询过滤过程,并进行了一些空间查询试验。

杨林等(2007)基于 MapGIS 7.0 面向实体的空间数据模型设计了一种空间查询语言,其语法规则是在 SQL 语言上扩展了面向空间实体,能够清晰地表达 MapGIS 7.0 面向实体的空间数据模型中空间数据的层次和组织,提供基本的数据操纵功能,并介绍了 GSQL 的体系架构和解译器实现原理。

徐翠玲等(2007)对空间数据操纵语言的实现体系结构进行了研究,在关系数据库和 GIS 应用之间引入中间层以实现空间数据的操纵,并描述了 GIS 应用发出的数据查询处理过程。

崔建军等(2007,2008)提出了一种语法树生成和实现的原理和具体方法,说明了中间件封装空间操作的 SQL 词法分析、SQL 语法分析的具体实现,通过语法树的生成对关系操作和空间操作进行分解以简化空间数据操作的复杂度。

王飞(2008)对基于 XML 语言进行空间查询进行了探讨,在比较多种 XML 语言后提出使用 XQuery 实现空间查询的优势,并根据开放地理信息系统协会(OpenGIS Consortium,OGC)简单要素规范的 SQL 标准将 XQuery 扩展成基于地理标记语言的查询语言(geography markup language-query language,GML-QL)以支持空间查询操作。

另外,还有方裕等(1999)、李海华等(2000)、张连蓬等(2002)、汪汇兵(2006)等人也对基于 SQL 扩展的空间查询语言进行了相关研究。

3. 中文自然语言空间信息查询的研究现状

随着自然语言的不断发展,自然空间查询语言的研究也有了一定的进展,有很多研究人员已经开始关注和研究这一领域的理论、技术和方法,在空间知识(空间查询词汇、空间查询语句句法等)、解译算法(空间分词、语义分析等)、查询系统(包括查询接口)以及空间基础理论与技术(空间关系等)等方面取得了一些有意义的成果,主要包括以下几个方面。

1) 自然空间查询语言涉及的词法、句法、语义内容及方法的研究

马林兵(2003a, 2003b, 2003c)研究了空间关系的自然语言查询形式及相应的空间语义, 设计了空间信息自然语言查询接口(spatial information natural language query interface, SINLQI), 并对相应的词典、分词方法、文法规则进行了探讨。其博士学位论文以自然语言理解的相关理论和方法为基础, 根据空间信息查询的特点, 系统地研究了空间信息查询的词法、文法、语义以及形式化问题, 探讨了自然语言空间信息查询的应用需求和领域。

付胜博(2007)研究了利用基于心理语言学原理的词典——WordNet 扫描空间数据表名和字段名, 建立同义词集, 构造从自然语言转化为标准 SQL 语言的规则库和模板库, 以实现自然语言理解技术检索空间数据库的方法。

徐爱萍等(2006a, 2006b, 2006c, 2006d, 2006e, 2007a, 2007b, 2009)对 GIS 中文查询语句的词典、表层语义以及中文查询语句与 SQL 语句的转换问题进行了研究, 提出了基于扩展 E-R 空间数据库环境的全匹配分词算法。基于系统词典对词的表层语义识别算法进行了研究, 提出了句子链栈的存储结构, 并通过分别扫描系统词典和单词词典来确定表链和句子链, 研究了栈式中间语言到 SQL 语句的转换, 从而实现 GIS 中文查询语句与 SQL 语句的转化。

乐小虬等(2005, 2006)利用地理空间描述模型中的相关概念扩展了自然语言中的空间语义角色, 定义了实体、属性、拓扑关系、方向关系、度量关系等空间语义角色, 并在其基础上构建了空间实体、空间关系、时空过程三类空间概念结构单元。根据这些角色对空间词义进行了扩展并介绍了描述方法, 通过对这些语义角色的标注、短语识别以及模式匹配来提取复杂的空间概念, 识别非受限文本中深层空间语义。

张雪英等(2007)对自然语言中空间词汇、空间关系描述的句法模式、自然语言空间关系与 GIS 中计算模型的语义转换等进行了研究。在此基础上探讨了汉语中描述各种空间关系的空间词汇及其句法模式, 以及自然语言空间关系查询请求表达的解析方法。

吴振南等(2007)对基于系统语料库的 GIS 中文查询语句的未登录词识别算法进行了研究。通过人机交互的方式将识别出来的未登录词加入语料库, 实现查询语句分词结果的自动记忆功能, 为 GIS 中文查询语句的正确理解奠定了基础。

李霖等(2008)基于形式本体的原理, 提出了一种对地理信息形式化语义分析的范例方法。

初宝华等(2008)对时态空间查询语言进行了研究, 设计了一种混合式时态空间查询语言, 对空间查询语言中的空间关系和时态关系进行了定义。

2) 自然空间查询语言空间关系语义与细化的研究

空间关系语义与细化是自然空间查询语言空间语义研究的重点。用自然语言