基于地理本体时空特征的全文语义检索研究

JIYU DILI BENTI SHIKONG TEZHENG DE QUANWEN YUYI JIANSUO YANJIU

宋佳 著

河南大学出版社
HENAN UNIVERSITY PRESS

ISBN 978-7-5649-3421-7
9 787564 934217 >
定价：32.00 元

基于地理本体时空特征的全文语义检索研究

宋　佳　著

河南大学出版社
·郑州·

图书在版编目(CIP)数据

基于地理本体时空特征的全文语义检索研究/宋佳著. —郑州:河南大学出版社,2018.7

ISBN 978-7-5649-3421-7

Ⅰ.①基… Ⅱ.①宋… Ⅲ地理信息学—语义信息—信息检索—全文检索—研究 Ⅳ.①P208.2

中国版本图书馆 CIP 数据核字(2018)第 172002 号

责任编辑 郑 鑫
责任校对 田丽贞
封面设计 郭 灿

出 版 河南大学出版社
地址:郑州市郑东新区商务外环中华大厦 2401 号 邮编:450046
电话:0371-86059701(营销部) 网址:www.hupress.com
排 版 郑州市今日文教印制有限公司
印 刷 北京虎彩文化传播有限公司
版 次 2018 年 8 月第 1 版 **印 次** 2018 年 8 月第 1 次印刷
开 本 787mm×1092mm 1/16 **印 张** 9.25
字 数 213 千字 **定 价** 32.00 元

前　言

地球信息科学经过几十年的发展，在地理认知、地理信息平台以及地理信息共享方面都有了长足的发展。从这三者的需求和发展趋势上来看，语义逐步成为了当前地理信息系统的研究热点之一。近年来，国内外许多学者都意识到地理本体的研究对实现语义层次上的地理信息系统具有重要意义，对促进地球信息科学的进一步发展也具有重要作用。但是地球信息科学及其相关领域的本体研究尚处于初级阶段，理论研究多而应用研究少，广度研究多而深度研究少。本书以地学数据资源共享中的数据检索为应用背景，着重于运用本体理论及技术，在本体多个应用方向中的一个——全文语义检索上作深入的研究和实践。本书从理论技术、模型框架、方法实现和实例应用几个层次，开展了基于地理本体时空特征的全文语义检索研究，主要工作体现在以下几个方面：

在理论基础部分，参阅了国内外学者现有的大量研究成果，从本体理论、地理本体理论、时空推理理论三个方面进行了研究综述。本体理论方面，从来源、定义、分类三个层次对本体概念进行了分析和总结，并对当前存在的多种本体描述语言和本体构建工具进行了比较。地理本体理论方面，首先给出并分析了广大学者从不同角度给出的地理本体的定义，然后在地理本体的构成和层次结构方面做了分析和总结，最后介绍了地理本体理论的研究进展。空间推理相关理论方面，重点介绍并比较了拓扑关系的描述模型和推理方法，明确了两个简单区域间的八种面一面拓扑关系。时间推理相关理论方面，阐述了时间的表达方式及时态表示模型。这些相关理论是本文开展应用研究的基础和前提，对指导本体支持下的语义检索研究具有重要的意义。

在模型框架部分，从两个方面展开。首先，以地理标记语言(GML)规范为参考基础，提出了地理时空本体模型的框架，并研究了构建地理时空本体模型的核心构建方法。从要素模型、几何模型、空间关系模型三个部分实现了空间本体模型；从时间点模型、时间段模型两个部分实现了时间本体模型。然后，基于传统的全文检索系统，结合语义检索技术和要求，提出了基于领域本体的全文语义检索框架模型，详细阐述了模型的组成和功能，为开发实现领域本体支持下的全文语义检索系统提供了框架参考。

在方法实现部分，基于理论和模型框架研究的成果，首先，对开源的传统全文检索工具包 Lucene 进行了改进，形成了基于 Lucene 改进的全文检索工具包 ELucene。然后，研究了语义检索系统框架中提到的语义标注和本体推理模块实现的关键技术细节：提出并实现了自动语义标注的算法，设计了查询推理引擎接口并提供了接口方法的默认实现，以及进行了语义检索下的相关性算法研究。最后，将实现了的语义标注和本体推理模块同 ELucene 相结合，形成了一个基于领域本体的全文语义检索工具包(Semantic ELucene)。Semantic ELucene 为进行面向具体应用的语义检索案例开发提供了工具，使应用开发的

周期大大缩短。

在上述理论、模型、方法研究的基础上，论文最后以“国家科技基础条件平台——地球系统科学数据共享网”为应用背景进行了元数据的区域语义检索实践。在我国典型区划本体库的建设和基于 Semantic ELucene 的元数据区域语义搜索系统实现两个方面进行了研究，构建了我国行政区划本体模型及行政单元本体实例，研究了其他区划本体的建立方法，实现了一个元数据区域语义搜索的原型系统，并进行了运行测试，测试结果表明元数据的区域语义搜索比现有的元数据搜索系统在查准率和查全率指标上均有了一定的提高。

应用地理本体进行语义检索研究对深化国家科学数据共享工作，满足国家重大科研需求方面具有现实意义。地理本体的构建研究对地学领域的数据互操作及集成等其他方面的应用也打下了良好的基础。

目　录

前言……（1）

第1章　绪论……（1）
1.1　研究背景 ……（1）
1.2　立题依据及背景 ……（3）
1.3　研究目的及意义 ……（5）
1.4　研究思路及章节安排 ……（7）

第2章　本体与地理本体理论及技术……（9）
2.1　本体理论基础 ……（9）
2.2　地理本体基础及研究进展 ……（15）
2.3　小结 ……（19）

第3章　地理时空本体模型构建的理论技术基础……（20）
3.1　空间推理相关理论 ……（20）
3.2　时间推理相关理论 ……（26）
3.3　地理标记语言 GML ……（27）
3.4　地理时空本体基本问题 ……（29）
3.5　小结 ……（30）

第4章　基于 GML 的地理时空本体模型构建研究 ……（31）
4.1　地理时空本体模型的构建思路 ……（31）
4.2　地理时空本体模型框架及构建方法 ……（32）
4.3　空间本体模型 ……（39）
4.4　时态本体模型 ……（54）
4.5　小结 ……（55）

第5章　基于领域本体的全文语义检索框架研究……（57）
5.1　传统全文检索技术 ……（57）
5.2　语义检索技术 ……（63）
5.3　基于领域本体的全文语义检索模型 ……（65）

5.4 小结 …… (70)

第 6 章 基于领域本体的全文语义检索开发实现 …… (71)
6.1 开源开发工具 …… (71)
6.2 基于 Lucene 改进的全文检索工具包开发 …… (74)
6.3 基于本体库的语义标注 …… (81)
6.4 基于本体推理机的查询扩展 …… (87)
6.5 语义检索下的相关性算法研究 …… (97)
6.6 小结 …… (102)

第 7 章 本体库支持下的全文语义检索实践及应用 …… (104)
7.1 应用项目背景 …… (104)
7.2 中国典型区划本体库的开发 …… (106)
7.3 地球系统科学数据共享网元数据区域语义搜索 …… (120)
7.4 小结 …… (128)

第 8 章 结论与展望 …… (129)
8.1 主要研究成果与创新 …… (129)
8.2 研究展望 …… (131)

参考文献 …… (133)

第1章 绪 论

1.1 研究背景

1.1.1 语义成为地理认知理论的重要内容

几十年来，地理认知的发展经历了从地理数据到地理信息，再到地理知识的发展要求和过程。地理数据是指表征地理圈或地理环境固有要素或物质的数量、质量、分布特征、联系和规律的数字、文字、图像和图形等的总称(邬伦等，2001)。地理信息是在认识和分析地理数据的基础上形成的对地理实体的性质、分布、状态等特征的描述，是对地理数据提炼、总结后的信息，具有空间位置、时序和多层次的专题特征。而地理知识是对地球表层系统的地理概念、原理、现象、过程、实践等的体系化的，语义层次上的人类认识的总和。地理知识不同于地理信息的一个明显特征是更强调信息的语义性，也比地理信息更具系统性。地理知识重视的是地理信息、概念、原理等各方面的相互关系和内在联系。地理知识是我们今后在地理认知发展上的重要方向，陈述彭(2004)院士曾指出：信息系统发展到现阶段，现有的地理信息系统软件包括 Arc/Info 里的空间分析软件包，数据挖掘的深度都远远不够，没有充分地把数据变成信息，把信息变成知识。

1.1.2 GIS 平台的发展要求语义支持

地理信息平台的发展经历了从桌面 GIS(地理信息系统)到网络 GIS，并将向网格 GIS 迈进的发展过程。网络 GIS 的发展成熟，无疑是 GIS 平台发展上里程碑式的成就，但是随着遥感、对地观测等空间信息技术的飞速发展，当前的计算机硬件处理能力和计算机网络组织在面对海量地理数据的分析和提炼上面临巨大挑战，而网格技术的出现为 GIS 平台的发展提供了新的发展方向。

自 1998 年，Forster(1998)首次提出网格的概念后，有关网格的研究在国内外逐步开展起来，网格的概念被进一步深化。网格 GIS 包含了各种资源全面共享的内涵，要共享就不可避免地需要考虑语义上的解决方案。孙九林(2002)院士曾针对网格技术下地学数据的共享问题提出：地学数据共享中数据网格的研究重点是如何消除信息孤岛和知识孤岛，实现信息资源和知识资源的智能共享。实现智能共享根本上需要语义层次的网格技术。此外，李德仁(2004)院士在论文《空间信息语义网格》中首次将本体引入网格系统，力

图构造一个具有语义共享的空间信息语义网格系统。

但不论是当前发展成熟的网络 GIS，还是处于刚刚起步的网格 GIS，都面临有异构环境下，地理数据、地理信息、甚至地理模型的集成共享问题。解决这个问题，建立统一的标准规范当然是一个重要的方面。但是，近年来，越来越多的学者主张应当从语义层次上解决异构环境信息的集成访问问题。特别是本体的提出和发展，已为我们解决平台异构带来的信息集成问题，提供了强有力的基础和支持。地理信息平台的发展暴露出了异构环境下的资源共享问题，语义无疑是推动各种资源无缝集成的重要途径。

1.1.3 语义互操作是地理信息共享的发展方向

地理信息早期共享的主要方式是数据格式转换，这是在早期 GIS 软件相互独立和各自为政发展的情况下而提出的。近年来，随着 WebGIS 的发展和成熟，GIS 互操作成为地理信息共享研究的热点之一。GIS 互操作的发展可以由两个层次来看：语法互操作和语义互操作。语法层次的 GIS 互操作除了依赖于各种分布式计算等技术外，更主要的是依赖于各种开放的、互操作的规范和标准。因此一些致力于研究制定 GIS 互操作相关标准、规范的团体和机构应运而生。例如：国际标准化组织地理信息技术委员会（ISO/TC211），开放地理信息系统协会（Open GIS Consortium Inc.，OGC）等。近年来，国内政府部门、科研单位和学者在 GIS 互操作和共享方面也研究和制定了很多标准和规范。国家科技部领导的科学数据共享工程从指导标准、通用标准和专用标准方面制定了《标准体系及参考模型》、《元数据标准化基本原则和方法》等 32 项标准（GB/T 2260－2002，2002. 中华人民共和国行政区划代码。国家质量监督检验检疫总局. 北京：中国标准出版社。

标准规范研究，2005）。孙九林院士主持领导的“国家科技基础条件平台——地球系统科学数据共享网”项目不仅在共享平台建设实践方面取得重要成果，更形成了一套与地球系统科学数据共享相关的标准、规范和条例（地球系统科学数据共享网规范，2005；孙九林等，2002，2003）。还有国内 GIS 专业的博士研究生也在数据共享方面作了大量的研究。比如，王卷乐（2005）就地学数据共享中，元数据的标准、结构、分析设计方面做了卓有成效的研究。诸云强（2006）前瞻性的探讨了面向 e-Geoscience 的地学数据共享的方向、进展和意义。目前学者关于 GIS 语法互操作方面的研究基本成熟，并已着手开始了语义层次的 GIS 互操作研究。

由于地理信息语义比较复杂，使得语义层次的互操作成为 GIS 研究中的一个难题。语义 GIS 互操作关键要弄清两个问题，一是 GIS 数据所表达的地理意义到底是什么，二是 GIS 数据描述这一地理意义的方式。前者涉及到了地理学理论问题，后者涉及到了语义学理论问题（陈常松等，2000）。

综上所述，从地理认知、地理信息平台、地理信息共享方面的需求和发展趋势上来看，都充分体现出了语义成为地理信息表达、集成和互操作的关键。建立在本体理论上的地理信息语义网，甚至地理信息语义网格是今后 GIS 的主要发展方向。作者在景东升（2005）的博士学位论文中归纳的“以语义为核心的地理空间信息技术发展趋势图”的基础上作了简单修改，展现出了语义为核心的 GIS 研究方向，如图 1-1。GIS 在语义层次上的突破已成为当前 GIS 研究的热点之一。

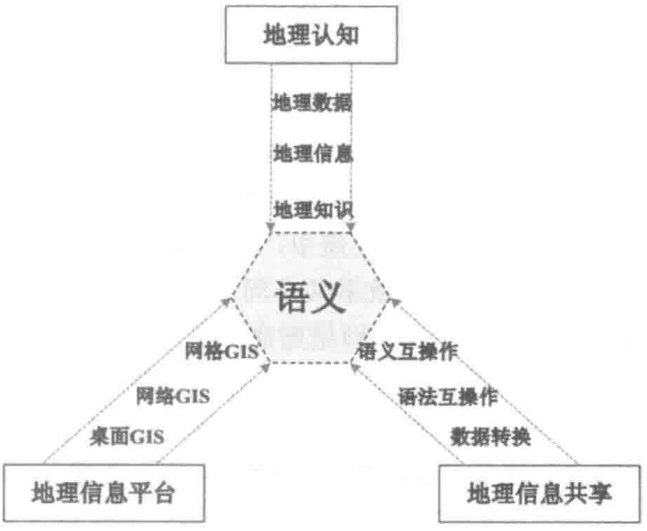

图 1-1 语义为核心的 GIS 研究方向

1.2 立题依据及背景

近年来出现的对地理本体的研究及应用对实现语义层次上的地理信息系统具有极大的推动作用，对促进地球信息科学的进一步发展也具有重要意义。

本体原本是一个哲学概念，即哲学中的本体论(Ontology)。近十几年来被引入了计算机科学的人工智能领域，并提出了信息本体的概念，即信息领域谈到的本体(ontologies)。基于地理本体的研究应用围绕地理数据语义异质和地理信息的智能化处理表现在三个方面：(1)地理信息的集成与互操作；(2)地学数据语义检索查询；(3)其他 GIS 时空信息智能处理及决策。地理信息的集成与互操作包含三个层次：数据层、语法层和语义层。数据层和语法层局限在解决由数据结构、概念模型和软硬件环境的差别引起的数据异构问题，而语义层次的地理信息集成与互操作是基于地理本体解决由不同团体、行业等对地学知识主观理解上的异构问题。地学数据语义检索查询是基于地理本体开展应用的另一个重要方面。因为地学数据具有时空特征，与时空概念结合的地学数据语义检索可以明显的改善数据检索效果，发现更多的相关数据，在地学数据发现与共享方面有极大的应用空间和应用价值；同时，地学数据的发现及查询检索也是地学数据集成应用的前提条件。从本质上来看，建立本体最终目的是要实现计算机与人之间或计算机系统与之间的相互理解，最终实现智能的人机交互、计算机系统之间的互操作和计算机系统中的知识重用(金芝，2001)。依据地学领域知识建立的地理本体可以使计算机理解蕴含在地理信息内部的人所理解的知识，使计算机对地理信息具有智能处理并进行自动管理及决策的能力。比如，智能交通 GIS 中基于本体的道路选择、公交换乘等应用对推动交通 GIS 的发展具有重要意义。由此可见，地理本体在地理信息挖掘、集成、智能处理等方面具有极大的应用价值，可以实现网络信息资源语义层上的互联共享，并为实现基于知识的智能处

理、决策方面的应用奠定基础。但是 GIS 及其相关领域的本体研究尚处于初级阶段，主要体现在理论研究多而应用研究少，广度研究多而深度研究少。所以本文着重于运用本体理论及技术，在本体多个应用方向中的一个——全文语义检索上作深入的研究和实践。

全文语义检索研究在科学数据共享中已表现出了很大的需求背景。科学数据是人类活动的产物，它代表人类的文明和社会的进步，对它的开发应用又可以进一步推动社会向前发展。科学数据应该在充分的传播和流通中，“把珍珠串成项链”，让全社会的人去利用(孙九林，2002)。国外很早就开始重视数据共享问题，建立了许多部门和行业数据中心，包括 1990 年美国国家航空航天局(NASA)组建的分布式最活跃数据档案中心群，地学空间数据总站(GOS)，世界数据中心(WDC)等。我国自上世纪 80 年代末开始逐步推动科学数据共享，包括 1982 年中国科学院提出的“科学数据库及其信息系统”建设项目，2002 年科技部着手启动的国家科技基础条件平台建设等。科学数据共享涉及到政策规章、技术支撑、标准规范、资源评价等多个方面。在技术层面，形成了以元数据为核心的数据管理、发现、使用方式。在数据发现层次上，由于当前元数据搜索是基于关键词的简单匹配方式，而数据资源可能是海量的，所以当前的搜索引擎在检索效果上远未能令人满意，特别是针对地学领域存在的大量蕴含时空特征的数据，国内尚没有面向地学时空特征数据开展语义检索的深入研究。所以，论文以“国家科技基础条件平台——地球系统科学数据共享网”项目为支撑和应用，针对传统元数据检索中单纯以关键字字面匹配的不足，从语义入手，对检索词进行有依据的推理扩展来返回检索结果。举例来说，当检索与“长江中下游植被分布”相关的数据，如果单纯从字符串匹配的角度，而不考虑语义层次的检索，像“长江三角洲地区植被分布”或“南京市植被分布”这样字面上没有“长江中下游”这几个字，但在语义上却与“长江三角洲地区”密切相关的数据，就很难被找出来。

本文的核心是进行语义检索的研究，但与计算机领域谈到的语义检索相比，主要侧重于解决地球信息科学及相关领域中与时空特征有关的语义层次上的全文检索。这首先还是由地理信息本身具有的空间特征、属性特征和时态特征决定的。另外，从地理科学及相关地学领域研究的角度来看，区域性是一个非常重要的特征。对此，我国地理学家林超(1981)曾谈到：“区域概念是地理学的基本观点，区域地理是地理学的核心”。基于空间分布和空间关系上的推理检索是本文的重要组成部分。另外，时空结合的语义检索也是很有必要的，如地籍变更、海岸线变化、土地城市化、道路改线，环境变化、行政界限变迁等应用研究领域都体现出了对时间或时空结合概念上的语义检索要求。本文的研究体现在考虑地理信息中蕴含的空间、时态特征，在时空概念上进行一定的推理扩展，改善检索性能中的两个基本评价指标：查准率(precision)和查全率(recall)。

1.3 研究目的及意义

1.3.1 研究目的

基于地理本体时空特征的全文语义检索研究将地理本体理论和技术与计算机领域的搜索引擎技术相结合，开展应用型研究。这类研究尚处于发展阶段，相关研究比较少，而且不够深入。

本文的研究目的是在地理本体及时空推理理论研究的基础上，开展基于地理本体时空特征的应用研究——全文语义检索研究。解决当前数据共享中检索数据单纯基于关键词检索的不足。基于地理数据资源中蕴含的区域、时间信息进行语义层次的推理查询扩展，改善衡量检索效果和质量的两个指标：查全率(Recall)和查准率(Precision)，达到一定的智能检索的效果。

1.3.2 研究意义

本文通过采用以地理本体为核心的新理论和技术，解决当前地理数据检索中的不足，满足用户对检索智能性的新需求，为科学数据共享等应用领域服务。具体包括以下几个方面：

(1) 对地理本体的应用发展具有重要意义

近年来，与地理本体有关的学术讨论日趋增多，这些研究在地理本体基本理论和进展、本体构建工具和构建方法等方面讨论得比较多，偏重于介绍和综述，真正深入的讨论如何建立地理本体的实现方案比较少。本文在地理本体理论、构建工具和方法等基础理论技术研究的基础上，重点探讨如何在计算机中构建、实现一套地理时空本体，从应用和工程的角度具体的研究地理本体的构建方法及基于本体如何完成推理的实现细节，对深入研究地理本体理论及应用具有重要意义。

(2) 促进时空推理理论在地理本体中的应用

时空推理是GIS理论中一个非常重要的研究方向。空间推理和时态推理各自都有着完整的理论体系，在时间、空间的表示与推理方面的研究都比较成熟的时候，非常自然的想法是把时间与空间结合起来研究。目前，已有不少时空数据模型和推理方面的研究，但将时空推理理论应用在地理本体中尚属起步阶段，相关的实践研究也比较少。本文将时空理论与地理本体理论相结合，探讨如何应用地理本体来组织和表达地理要素、现象和过程等，体现他们在空间上、时态上的特征或关系，对时空推理理论的应用有一定意义。

(3) 对地球信息科学理论研究具有重要意义

地球信息科学经过多年的发展，取得了很多突破和成绩，但也有许多理论问题未能解决，包括地理空间数据的语义表达和处理、空间数据在语义层次上的集成和互操作、地理空间认知、空间的多尺度处理等。有许多学者意识到了地理本体研究对解决这些前沿问

题具有重要意义，提出了基于地理本体的各种解决方案。有些学者(Mark，et. al.，1999)已经将地理本体研究提升到地球信息科学的基础理论这样的高度，可见其对地球信息科学理论研究的重要性。

(4) 对 GIS 系统的发展具有重要作用

新一代地理信息系统应该实现语义层次上的互操作，并且应该是一种智能化、大众化和分布式的系统。显然，现有的 GIS 系统大多数都还是一种非常专业的信息系统，用户必须具备 GIS 方面的专业知识，才能够理解和操作 GIS 系统，而且对于常识性的问题很难做出回答。地理本体对地理概念及其之间关系的形式化说明以及对推理能力的支持，使得其在自然语言理解和处理、地理概念比较与匹配、检索结果的知识化表达等处理步骤中起着关键的作用，是智能化地理信息检索的核心组件。

(5) 推动地理信息系统的社会化应用

地理信息系统的应用目前已经逐步从最初仅仅局限于个别专家到学术圈再到政府部门和大型企业的过渡，虽然迄今为止 GIS 已经应用到大多数与地理相关的行业和部门，但其社会化应用程度仍很有限。其主要原因：一是使用 GIS 所需的地理知识的普及教育程度低，且比较专业和抽象，同时 GIS 表达和处理常识地理知识的能力还很有限；二是 GIS 用户界面的友好程度低，回答一个简单的常识问题，需要经过许多操作步骤才能完成，而不能通过简单的按键来自动化的智能实现。地理本体使地理知识的共享和传播成为可能，地理本体对推理能力的支持，使之可以实现复杂的、智能的地理推理计算，对于 GIS 社会化具有重要意义(王敬贵，2005)。

(6)在满足国家重大科研需求和建设科研平台方面具有现实意义

2004 年，科技部、国家发展与改革委员会、教育部、财政部联合制定颁布了《2004——2010 年国家科技基础条件平台建设纲要》，指出国家科技基础条件平台建设是充分运用信息、网络等现代技术，对科技基础条件资源进行的战略重组和系统优化，以促进全社会科技资源高效配置和综合利用，提高科技创新能力。其建设原则突出共享，包括了自然科技资源共享平台、科学数据共享平台等六个重点建设平台。另外，中国科学院早在 2003 年已确定了“全面推进科研环境手段和方法的信息化，建设 e－Science，全面推进科研管理的信息化，建设 ARP(Academy Resource Planning)，最终建设数字化科学院(Digital CAS)”的信息化发展总目标(阎保平，2003)。“十一五”期间，e－Science 的建设工作将全面展开和推进，它是一种信息化的基础设施，提供了一种信息化的科学研究环境和平台，使得不同学科领域的研究和科研活动能够有针对地开发特定的科学研究与应用。基于地理本体时空特征的全文语义检索研究以解决数据或资源检索的质量和效果为目标，对加强数据和资源共享平台建设(特别是地学领域的科学数据资源共享)，构建 e－Science 科研平台具有一定的现实意义。

1.4 研究思路及章节安排

1.4.1 研究路线

基于地理本体时空特征的全文语义检索研究从逻辑上可分为两个阶段(图 1-2)。第一阶段是具有时空特征地理本体库的设计和实现阶段;第二阶段是运用建立好的地理本体库,作为语义推理的基础,设计实现基于领域本体的全文语义检索系统。

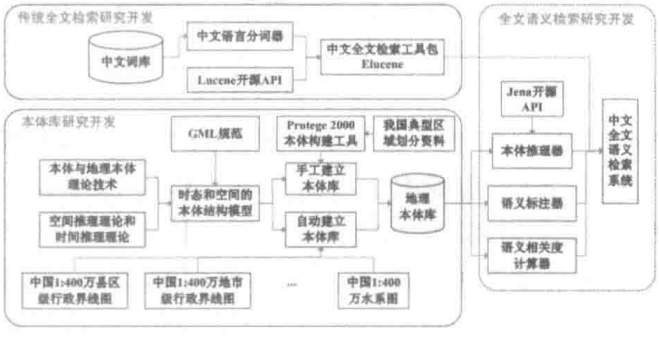

图 1-2 研究技术路线图

1.4.2 章节安排

围绕着基于本体库的全文语义检索这个核心内容,本文将组织成如图 1-3 所示的完整体系。

第 1 章首先论述本文的研究背景,从三个角度给出语义为核心的 GIS 研究方向是当前的热点之一,然后引出论文所要研究的具体问题及其背景,并明确研究目的和意义,最后给出论文的研究思路及章节组织。

第 2 章分为本体理论和地理本体理论技术两大块。本体理论部分主要阐述和总结国内外学者在本体理论研究上的观点和进展。首先,阐述本体起源于哲学领域的背景情况;然后,综述要研究的信息领域本体的含义和分类,对学者关于本体的各种定义和分类进行分析和总结;最后,比较和总结本体的主要描述语言和构建工具,确定本文研究采用的本体描述语言和本体构建工具软件。

地理本体理论技术部分主要集中在地理本体的基础概念和对地理本体的研究进展。首先,从地理学的学科特点出发,阐述地理概念,特别是地理学中的空间概念和时间概念,明确其空间关系;接着,总结分析学者们关于地理本体的各种定义,讨论其逻辑构成及蕴含在它们当中的层次关系;最后,讨论地理本体的构建方法。

第 3 章给出地理时空本体模型构建的理论技术基础。理论部分包括空间推理相关理论和时间推理相关理论。空间推理相关理论以空间关系中的拓扑关系描述和推理为核心;时间推理相关理论主要是研究了时间的表达及其表示模型。因为下一章研究的是基

于地理标记语言(GML)的地理时空本体模型,所以本章最后给出了 GML 的介绍,作为地理时空本体模型构建的技术基础。第 2 章与第 3 章部分构成了论文研究的理论技术基础。

第 4 章这一章的核心是参考 GML 的规范,构建地理时空本体模型。首先针对构建中面临的关键问题,提出解决思路。然后,研究并给出了地理时空本体模型框架,并归纳出核心构建方法。最后以 GML 中的模型为参考,用 OWL 语言设计实现了空间本体模型和时间本体模型。

第 5 章着手研究在传统全文检索技术体系的基础上,设计并提出基于领域本体的全文语义检索模型。首先,介绍传统全文检索相关概念、模型、结构特点等;然后,讨论了语义检索的概念、研究思路等;最后,给出了基于领域本体的全文语义检索模型,并详细描述了其组成。这一章为后面全文语义检索系统的开发实现提供了清晰的功能模块组成和总体设计框架。

第 6 章从软件方法实现的角度,阐述基于领域本体全文语义检索系统的实现。首先,搜集并评述了支持基于领域本体全文语义检索开发的开源开发工具;然后,基于开源的 Lucene 工具包,设计实现一种改进的全文检索引擎工具包 ELucene(EnhancedLucene),在此基础上,加入执行语义检索功能的核心模块:语义标注器、基于 Jena 的语义推理机、相关度计算器,最终形成一套完整的基于领域本体的全文语义检索工具包(Semantic ELucene)。

第 7 章实践及应用。以"地球系统科学数据共享网"为应用背景,在本体实践上研究了中国典型区划本体库的开发。基于这套本体库,运用 Semantic ELucene 开发实践了元数据区域语义搜索系统,并进行了运行测试,对检索效果进行了分析及评价。

第 8 章对全文进行总结。给出论文的主要研究成果和创新点,总结研究过程中的难点及体会,对今后的研究方向进行展望。

第 2 章　本体与地理本体理论及技术

2.1　本体理论基础

2.1.1 本体的来源

本体原本是一个哲学概念，在哲学领域称为本体论(Ontology)。最早有关本体的解释是公元前四世纪古希腊哲学家亚里士多德的描述："对世界上客观存在物的系统地描述，即存在论"(Tim, et al., 2001)。17 世纪，"本体论"的概念被西方哲学家明确提出来，它研究客观事物存在的本质，是对客观存在的一个系统解释或说明，属于形而上学理论的分支，与研究主观认知的认识论相对。本体论与认识论、方法论共同构成哲学的三大基本问题。

近十几年来，哲学上的本体论思想和方法被引入人工智能领域，其最终目的是为了解决知识重用和共享，而且语义互联网、语义建模和语义集成需要本体作为支持发展的基础理论。

另外，由于中文对 Ontology 一词的译法不一，众多中文文献中出现了"本体论"、"本体"和"本体理论"几个术语。王敬贵(2005)在参考了国内许多学者比较公认的看法的基础上，在他的博士论文中予以了规范：术语"本体论"(Ontology)只用在哲学研究中，其英文单词的首字母大写且为不可数名词；在信息科学研究中使用术语"本体"(ontologies，其英文单词的首字母小写且为可数名词)和"本体理论"(ontology theory)。"本体"与"本体理论"之间的关系类似于"数据库"和"数据库理论"之间的关系。即"本体"是"本体理论"的研究对象，而"本体理论"则是研究"本体"的系统理论和方法。

2.1.2 本体定义

在信息科学领域，学者对于本体的概念和思想的讨论研究非常活跃。Neches 等(1991)是最早在人工智能领域使用"本体"这个术语，并将本体定义为"组成主题领域的词汇表的基本术语及其关系，以及结合这些术语和关系来定义词汇表外延的规则"。

1993 年，Gruber(1993)给出一个关于本体的最为流行的定义，即本体是概念化的明确的规范说明。Gruber 采用概念化的形式定义＜D，R＞结构，把本体解释成"共享概念化的形式的明确地规范"，其中 D 是领域，R 是 D 中相关的关系集合，因此该定义能够很