

空间信息检索系统中的 语义本体技术

The semantic ontology technology in spatial information retrieval system

孙胜涛 ©著



北京邮电大学出版社
www.buptpress.com

国家自然科学基金青年基金项目(61303130)和
河北省自然科学基金面上项目(F2014203093)资助

空间信息检索系统中的 语义本体技术

孙胜涛 著



北京邮电大学出版社
www.buptpress.com

内 容 简 介

本书系统地介绍了语义本体技术在空间信息检索应用中的新技术和方法,在充分调研和分析语义本体技术在智能信息检索的应用现状和局限性基础上,主要从语义本体对自然语句中用户检索需求的理解、本体对不确定性领域知识的描述、本体操作对近似性检索和推理的支持等几个方面展开论述和研究,尝试对语义本体的描述机制和推理能力进行扩展或改善。力图从语义分析角度,解决空间信息自然语言检索中各类关键技术问题,构建并实现具有语义检索能力的空间数据服务语义化检索系统。

本书可作为高等学校计算机专业研究生和相关教师的参考用书,也可供从事语义 Web、语义信息检索、智能空间信息检索等方面研究人员的参考资料。

图书在版编目(CIP)数据

空间信息检索系统中的语义本体技术 / 孙胜涛著. — 北京:北京邮电大学出版社, 2016.5
ISBN 978-7-5635-4719-7

I. ①空… II. ①孙… III. ①卫星通信系统—研究 IV. ①TN927

中国版本图书馆 CIP 数据核字(2016)第 054719 号

书 名: 空间信息检索系统中的语义本体技术

著作责任者: 孙胜涛 著

责任编辑: 满志文

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路 10 号(邮编:100876)

发行部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 北京鑫丰华彩印有限公司

开 本: 720 mm×1 000 mm 1/16

印 张: 13.5

字 数: 261 千字

版 次: 2016 年 5 月第 1 版 2016 年 5 月第 1 次印刷

ISBN 978-7-5635-4719-7

定 价: 38.00 元

• 如有印装质量问题,请与北京邮电大学出版社发行部联系 •

前 言

各类空间数据的飞速增长和应用扩展,使得空间信息检索表现出越加重要的作用和地位,传统的空间信息检索常常采用与地理术语相关的组合式检索技术,其专业性和复杂性往往难以为大多数非专业人员所掌握。在空间信息检索中存在的时空分析、数据多样性、数据相似性检索、知识不确定性等特点,需要借助本体技术来提供对应的语义化信息检索方式与匹配方法。

虽然下一代互联网技术引入了语义本体技术并定义了应用框架,可以提供基础的语义描述方法和知识推理技术,但是基于现有研究成果和水平,在空间信息自然语言检索应用中,本体技术在自然语言理解、不确定性知识描述、相似性检索和不确定性推理等方面,均存在一定的局限性。这些问题的解决对于本体在空间信息检索中的应用将提供有力的支持,对于实现高效、智能的空间信息检索具有重要的研究意义和应用价值。

本书以空间信息的智能化检索为应用背景,从本体对自然语句中用户检索需求的理解、本体对不确定性领域知识的描述、本体对近似性检索和非精确推理的支持等几个方面进行论述,介绍和说明语义本体在空间信息检索系统中的关键技术问题和应用情况,并通过具有语义检索和推理能力的空间数据服务自然语言检索原型系统来展示本体的具体实施技术和应用效果。

第1章对目前基于语义的信息检索技术发展现状进行简要介绍,针对空间信息检索的特殊需求,总结出基于语义的空间信息检索面临和需要解决的关键问题,为后继章节内容的展开提供基础。

第2章针对日益被大众用户所接受的基于自然语言的语义检索方式,说明了本体技术在空间信息自然语言检索中的应用情况和发展现状,分析并指出了存在的问题和不足。

第3章对空间信息系统的发展历程进行简要介绍,并针对空间信息系统中存在的特殊语义问题进行了分析,提出了需要面临和解决的关键问题,为后继各项研究工作的展开提供依据。

第4章介绍了一种基于层次约束的多层本体设计和构建方法,并构建了具有三层结构的空间信息本体,实现了对各类空间信息服务资源的本体化描述,为后继的研究工作提供了本体知识库基础。

第5章分析了本体应用于自然语言理解的关键问题和难点,分析了概念层次

网络 HNC 模型与本体模型结合的可行性,提出了语言认知模型和本体描述模型相结合的机制,实现了自然检索语句中用户需求的语义解析和本体化描述。

第 6 章分析了本体模型中语义不确定性度量的关键问题和难点,重点研究可能性逻辑和概率逻辑的原理和特点,分析两者相结合的优势和可行性,提出了基于可能性理论和概率统计方法的语义定量化度量机制 SRQ-PP,并给出空间信息领域中各类不确定性语义关系的定量化度量方法,实现了层次化空间信息本体知识库中各类不确定语义关系的定量化描述。

第 7 章分析了包含有不确定性度量的本体检索和推理过程的关键问题和难点,重点研究了激活扩散算法在语义定量化本体结构中进行知识检索和推理分析的适用性,提出了基于激活扩散算法的本体相似性检索和非精确推理的机制,实现了用户需求和空间信息资源间的关联性检索和近似性匹配。

第 8 章基于以上研究成果,设计并构建了基于本体的空间数据服务语义化查询系统,给出了该系统的总体结构设计和操作流程描述,对系统的各模块功能和关键技术进行了说明,实现了空间数据服务的智能语义化检索,并通过该原型系统的应用效果,对研究成果进行了验证和总结。

在本书的编写过程中,作者参考了大量国内外出版物和网上资料,在此谨向各位作者表示由衷的敬意和感谢。本书的有关研究工作得到国家自然科学基金项目(P2P-Grid 环境中分布式不确定本体模型的研究,编号 61303130)和河北省自然科学基金面上项目(面向高性能数据密集型地学计算的 IO 中间件的方法研究,编号 F2014203093)的资助,北京邮电大学出版社的领导和编辑对本书的出版也给予了大力支持,作者在此深表谢意。

本书的内容及编写得到了中国科学院遥感与数字地球研究所李国庆研究员、刘定生研究员、王力哲研究员、于文洋副研究员等的悉心指导和帮助,燕山大学任家东教授、宫继兵副教授也对本书提出了建设性的建议,燕山大学知识工程实验室的张琳、李芸、陶双男、许钦亚等研究生也参与了本书中部分研究工作,特别是我的妻子吴爱芝女士为本书校稿付出了的辛勤努力,作者在此一并表示感谢。

由于语义信息检索相关理论、技术及应用发展迅速,加之作者水平有限,书中难免存在不足之处,恳请各位专家和读者批评指正。

孙胜涛

燕山大学 信息科学与工程学院

河北省计算机虚拟技术与系统集成重点实验室

2015 年 11 月

目 录

第 1 章 绪论	1
1.1 空间信息语义检索发展现状	1
1.2 空间信息语义检索关键问题	3
1.3 本书主要研究内容和特点	6
1.4 本章小结	8
本章参考文献	8
第 2 章 本体技术在语义检索中的应用现状	9
2.1 自然语言检索的研究现状	9
2.2 本体的含义以及在自然语言检索中的作用	12
2.3 基于本体的空间信息自然语言检索研究现状	16
2.4 当前研究的不足和存在问题	19
2.5 本章小结	20
本章参考文献	20
第 3 章 空间信息系统中的语义问题分析	22
3.1 空间信息系统的发展历程	22
3.1.1 空间信息服务的发展历程	22
3.1.2 空间信息服务和管理模式的发展	27
3.1.3 空间信息系统存在的问题分析	30
3.2 空间信息系统的特点	31
3.2.1 空间数据的特点	31
3.2.2 空间信息系统的特点	33
3.3 空间信息系统的语义问题分析	35
3.3.1 空间信息系统中语义异构性问题	36
3.3.2 空间信息系统中领域知识的形式化问题	37

3.3.3 空间信息系统中信息检索方式问题	38
3.3.4 空间信息系统中信息检索策略问题	38
3.4 本章小结	39
本章参考文献	39
第4章 层次化空间信息本体构建方法研究	41
4.1 本体构建方法的对比和分析	41
4.1.1 本体构建的基本原则	41
4.1.2 本体构建的主要方法	42
4.1.3 本体构建方法的对比分析	45
4.2 空间信息本体的特点	45
4.3 层次化空间信息本体构建方法	48
4.4 层次化空间信息本体的构建	51
4.5 层次化空间信息本体的管理	54
4.6 层次化空间信息本体的对比评价	57
4.7 本章小结	59
本章参考文献	60
第5章 基于本体的自然检索语句语义化解析	62
5.1 自然检索语言的语义化解析研究现状	62
5.1.1 自然语言处理的主要技术	62
5.1.2 自然语言的表述方法和语言模型	64
5.1.3 概念层次网络 HNC 模型	66
5.2 本体模型和 HNC 模型的结合	68
5.2.1 本体模型和 HNC 模型结合方式的分析	69
5.2.2 用于自然检索语言理解的应用本体构建	71
5.2.3 自然语句中用户检索需求的本体化描述	75
5.3 自然检索语句的解析实例展示和结果分析	76
5.4 本章小结	83
本章参考文献	83
第6章 基于本体的不确定性知识描述方法	85
6.1 基于本体的不确定信息描述研究现状	85

6.1.1 空间信息领域中的不确定性表现	85
6.1.2 基于本体的不确定性知识描述研究现状	87
6.1.3 对空间信息领域中各类不确定性的分析	
6.1.4 各种描述逻辑扩展方法的对比	92
6.2 不确定性表达上可能性逻辑和概率理论的结合	95
6.2.1 可能性逻辑在不确定性知识表达中的问题	95
6.2.2 可能性逻辑中 Nec 度量的语义解释	97
6.2.3 可能性逻辑和概率统计方法的结合	100
6.2.4 基于 SRQ-PP 方法的不确定性语义关系度量实例	104
6.3 空间信息服务中各类不确定性的定量化描述	109
6.3.1 语义属性描述的不完整性	109
6.3.2 语义属性描述的侧重性	111
6.3.3 属性取值的多值性和语义关联的多解性	114
6.3.4 概念间以及概念和实例间从属关系的部分性	119
6.3.5 推理结果满足程度的可比性	121
6.4 空间信息本体中不确定性描述的综合实例展示和结果分析	125
6.5 本章小结	128
本章参考文献	128
第 7 章 基于本体的关联性检索和不确定性推理技术	132
7.1 基于本体的关联性检索技术	132
7.1.1 原有本体检索机制和存在问题	133
7.1.2 本体中概念和术语间关系的精确化描述	144
7.1.3 基于激活扩散算法的本体关联性检索	152
7.1.4 基于 SRQ-PP 的激活扩散算法检索策略	165
7.2 基于本体的非精确推理技术	169
7.2.1 已有本体推理机制和存在问题	169
7.2.2 基于激活扩散算法的本体推理方法	172
7.3 本章小结	176
本章参考文献	177
第 8 章 基于自然语言的空间数据服务语义化查询系统	180
8.1 空间数据服务检索系统的发展现状	180

8.2	基于本体的空间数据服务语义化查询系统总体设计	181
8.3	基于本体的空间数据服务语义化查询系统关键技术说明	183
8.3.1	检索需求的解析和本体化描述	183
8.3.2	空间数据服务的本体化描述	184
8.3.3	基于激活扩散算法的空间数据服务检索和匹配	187
8.4	基于本体的空间数据服务语义化查询系统运行结果	198
8.5	本章小结	205

第1章 绪论

1.1 空间信息语义检索发展现状

在空间数据服务中,随着空间信息的日益增多和广泛应用,空间信息的检索越加发挥出重要的作用,同时非专业人员在这一检索活动中的比重也大幅提高。这些用户一般采用简单的关键词进行检索,其特点是要求简单、实现容易,但实际检索结果常常与用户所期望的结果间存在着很大的差距。为了使检索的结果更为精确,人们进一步构造了可对各种关键词进行多种组合的检索方式,该种方式可获得更好的检索结果,但大多数用户由于没有经过检索培训,难以很好地利用这种组合检索方式来构造复杂的检索策略,使得这类复杂的组合式检索难以在大多数非专业人员中使用。如何为用户提供一种基于自然语言检索的方式,使用户可以利用规范的自然语言检索到全面精确的相关信息,是国内外信息检索界广泛关注和着力解决的关键问题之一。由于空间数据的特点、空间数据服务的异构性、用户需求表达的模糊性等问题,空间信息的自然语言化检索、查询和获取存在着较多的难题。

在空间数据服务中,空间信息检索具有显著的时间和空间特性,它是信息检索的一个特殊应用领域,其特殊性主要表现在相似性检索、时空关联检索、知识的不确定性等方面^[1]。传统的基于关键词字符串匹配的信息检索技术已不能完全满足这些要求,需要更高层次的基于语义的信息检索和匹配方法。

语义网中的本体技术作为一种新型的知识组织和表达方式,具有良好的概念组织结构和对逻辑推理的支持,在信息检索领域特别是在基于知识的语义化检索方面得到了飞速的发展和广泛的应用。近几年来,国际上各科研团体在不同领域都进行着本体技术的应用研究工作,并取得了卓有成效的研究成果,语义本体技术在智能系统中对各类语义关系的形式化描述和处理方面都发挥了重要的作用^[2]。基于本体的语义描述和推理方法,为空间数据服务中用户需求的解析和理解、空间信息资源的语义化描述和匹配、领域知识的形式化描述和推理分析等自然语言检

索中关键技术问题的解决提供了可行的途径和有力的工具,将有效地提高空间数据服务的检索效率和智能化程度。

然而在各类空间信息服务系统不断涌现的同时,本体在空间信息智能化服务系统中的知识表达和推理能力却没有得到相应的改善,所面临的问题主要表现为:以自然语言描述的检索需求存在一定的模糊性,以本体形式描述的领域知识存在一定的不确定性,信息检索和匹配过程存在一定的近似性,本体推理过程存在一定的非精确性。本体在语义处理能力上局限于对确定性精确化知识的表达和推理,对于以上不确定性因素难以理解和处理,这种局限性极大地影响了其语义处理能力的灵活性和适应性。在自然语言解析、资源的描述和查询匹配过程中都存在一定的不确定性,在空间信息自然语言检索中尤为突出。如何实现基于本体的不确定性领域知识表述和非精确推理,对于增强当前空间信息语义系统的描述和推理能力,提高空间数据服务的智能化程度都具有重要的研究意义和应用价值。

随着本体技术的发展和广泛应用,本体技术于 20 世纪 90 年代被引入到地学领域中,地学领域专家对本体进行了相关的研究,并构建了各类应用目的的地理本体^[3],初步解决了时空关系推理、语义异构匹配、空间信息服务语义检索等语义问题。但由于本体对语义描述和理解能力的局限性,本体在空间信息领域所发挥的作用也是有限的,特别是针对空间信息的多样性、不确定性等特点,本体技术在空间信息领域知识的表达和分析中仍有若干关键问题没有很好地解决,主要表现在以下几个方面。

(1) 本体的主要特征是基于概念空间来描述语义,其实施重点在于表达和理解特定领域中专业词汇间的语义关系,它在解决名称异构(多词同义)、概念异构(一词多义)等问题上发挥了显著的作用。然而,由于其知识结构是按照领域中名词术语的层次从属关系来组织的,其语义表达过程完全围绕概念间各类关系来进行,使得其语义理解和处理能力局限于对自然语句中名词术语间关系的分析,缺乏对自然语言中动词的描述与解析能力。特别是在空间信息自然语言检索中,动词对应用目的、数据用途等需求信息的获取具有重要的作用,不同的动名词搭配关系反映了不同的应用需求,仅依靠名词术语的理解所获得的语义是具有一定片面性的。因此,如何在概念知识结构上实现对自然语句中动词的描述和理解,是解决空间信息检索语句语义理解的关键问题。该技术问题的解决将为自然检索语句中用户空间信息检索需求的理解提供有效的分析方法,可为后继空间信息检索的全面性和准确性提供有力的保证。

(2) 本体模型的理论基础是描述逻辑(Description Logic),描述逻辑具有很强的表达能力和可判定性,在解决领域知识的形式化表达和可判性推理问题上具有一定的优势。本体对知识的描述机制依赖于描述逻辑的理论基础,其知识的表达

和语义的描述都针对的是静态知识和确定性信息。但由于现实世界中客观上存在的随机性、模糊性以及事物或现象暴露得不充分性,导致人们对各类事物的认识往往是不精确、不完全的,大多都具有一定程度的不确定性,该问题在空间信息领域尤为突出,使得空间信息领域知识的表达过程中存在较多的不确定性因素和关系。因此,如何扩展和改善本体的描述能力,解决空间信息本体知识库中各类不确定性领域知识的定量化描述,是解决空间信息本体对不确定性信息理解和处理的关键问题。该问题的解决将为智能空间信息检索系统中不确定性领域知识的表达提供有效的手段,并可有效地保证信息检索过程中各类不确定信息的理解和处理能力。

(3) 在空间信息检索和推理分析过程中也存在一定量的不确定性和非精确性,本体技术及其推理机制在处理不确定性信息时同样表现出较为明显的不适应性和局限性。因而如何在采用布尔运算规则(定性化方法)的描述逻辑基础上,引入对各类空间信息语义关系的定量化描述机制,并基于其推理过程中不确定性的推算来实现非精确推理,是解决空间信息相似性检索和非精确推理的关键技术问题。该类问题的解决将为本体中各类知识的相似性检索提供有效的方法,并可基于相关性检索和不确定性推理机制,实现空间信息检索过程中用户需求和资源描述间的关联性匹配,可在一定程度上改善空间信息检索的效果。

1.2 空间信息语义检索关键问题

空间信息获取与应用技术的不断发展,带来了各类空间数据的飞速增长和应用领域的不断扩大,使得空间信息的检索表现出越加重要的作用和地位。随着越来越多非专业人员利用空间信息检索获取所需信息,传统的空间信息检索常常采用的与地理术语相关的组合式检索技术,其专业性和复杂性往往难以为大多数非专业人员所掌握。能否以自然语言形式来描述空间信息检索需求,成为国内外空间信息检索界广泛关注和着力解决的关键技术之一。

从信息检索角度,自然语言检索方式近年来在专家系统、情报检索、办公自动化系统的人机接口应用中获得了大多数用户的认可和使用^[4],同时基于自然语言的信息检索方式在实际应用中,需要面临不同应用领域中信息检索的特殊性要求。从技术实现角度,基于语义本体的检索技术是自然语言检索的关键技术之一。在空间信息检索中存在的时空分析、数据多样性、数据相似性检索、知识不确定性等特点,需要在自然语言检索的语义本体层次,提供对应的信息检索方式与匹配方法。从技术研究角度,这就要求本体技术在自然语言理解、领域知识描述、相似性检索和不确定性推理等方面,提供较好的解决方案与性能。

虽然下一代互联网技术引入了语义本体技术并定义了应用框架,可以提供基

础的语义描述方法和知识推理技术。但是基于现有研究水平,在空间信息自然语言检索应用中,本体技术在自然语言理解、不确定性知识描述、相似性检索和不确定性推理等方面,均存在一定的局限性。这些问题的解决对于本体在空间信息自然语言检索中的应用将提供有力的支持,对于实现高效、智能的空间信息检索具有重要的研究意义和应用价值。

对此,本书以空间信息的智能化检索为应用背景,针对本体技术在以上几个语义描述和推理能力方面的局限性,从本体对自然语句中用户检索需求的理解、本体对不确定性领域知识的描述、本体对近似性检索和非精确推理的支持等几个方面展开深入的研究,尝试对本体描述机制和推理能力进行扩展或改善。力图从语义描述和分析角度,解决空间信息自然语言检索中各类关键技术问题,构建并实现具有一定语义检索能力的空间数据服务语义化检索系统。

基于以上本体对语义描述和理解能力的局限性分析,采用现有的本体描述和推理机制来进行自然语言的理解和处理,主要存在以下几类关键问题:

(1) 以数值形式描述的概念属性具有量化的值,但是这些数值对于机器而言很难理解其数学含义,对于自然语言理解可提供的帮助非常有限。而且基于这些数值进行推理和分析,又需要采用程序代码来描述如何将这些数值转换成语义的推理规则,这将直接影响语义理解的实现难度和灵活性。

(2) 以概念间关系来表述其语义关联,可以比较形式化地描述概念间关系,但是计算机很难能直接理解这些语义关系名称(标签)的含义,例如理解语义关系“canUse”代表“利用、使用”的含义,这使得计算机很难直接获知这些语义关联的含义,同样需要依靠用户编写程序代码或设置对应的推理规则来形式化地表述出这些语义。而且在汉语自然语句的理解上,动词是语义的重要表达,但是本体只注重概念(名词)的语义表达,缺失对动词所表达含义的描述。

(3) 对本体中概念结点可以用自然语句或关键词来描述其语义属性,但是这些自然语言符号都是人类容易阅读的,计算机很难直接理解。而目前很多本体概念的匹配和映射都是依赖于这些描述信息来进行的,使得语义理解和解析能力受到传统词法匹配技术的影响,降低了语义理解和分析的级别。

(4) 在领域知识中包含有大量的不确定性因素,自然语言中也包含有一定量的模糊性,这些不确定性信息在本体化描述时只能采取定性化处理方法来实现,缺乏对这些语义属性和关系的量化度量机制,使得本体中语义关系过于规范化和单纯化,难以适应自然语言理解中不确定性语义关系理解的需要,限制其语义分析的灵活性和适应性。

以上这些问题,总体来说就是目前本体所描述的语义信息要么过于数值化,要么过于语言化,没能给计算机一种适当的形式化描述形式,使得目前本体关于知识

的描述大都是词汇级的,没有真正达到语义级。这些问题极大限制了本体在自然语言理解中应用的范围和效果,也成为目前限制本体继续深入发展和广泛应用的瓶颈所在。

基于本体的工作原理可知,语义本体在空间信息自然语言检索中应用具有一定的技术优势,可有效地满足空间信息检索的特殊需求,主要表现在:用户使用自然语言来表达数据检索意图和目的,可有效地避免专业性较强的传统检索界面对用户背景知识的要求,用户以习惯和方便的方式来表达需求,可提高信息检索的便捷性和灵活性;基于本体知识库中对时间和空间关系的语义形式化描述,可进行时空推理和分析,自动地为检索过程提供时间和空间需求的语义化解析,可增强信息检索的高效性和智能性;基于本体知识库中对领域知识的形式化描述,可采用不确定性推理和分析算法来提供相似性检索和匹配分析,在用户需求和所需资源间建立模糊语义化关联,保证信息检索的全面性和有效性。基于上面对现阶段本体局限性的分析,可以发现目前在基于本体的自然语言检索中,仍未很好解决并且值得研究的关键问题主要有如下几方面:

(1) 现有的用于支持自然语言检索的语言知识库,大多从词法和语法上进行自然语言的描述,缺乏语义分析的基础,限制了语言理解的能力。而已有针对语义关系建立的语言知识库,大多只针对某种语言环境和限定条件进行语义的分析和理解,没有从语言学角度来描述语义关系的知识库。因而需要选择适用于汉语特点的自然语言描述模型和语义认知模型,将这些模型和本体模型相结合来改善本体对自然语言的理解能力,以提高本体对自然检索语句中用户需求的语义分析能力。

(2) 描述逻辑作为本体知识表示的形式化基础,具有很强的知识表示和推理能力,但是描述逻辑通常只能处理含义明确的概念知识,在处理非单调的、不完备的知识时,却无能为力。在基于本体的信息检索过程中存在大量的非精确信息和不确定性因素,在本体知识的表述中也存在较多的不确定性知识,这就需要对原有的本体模型进行改造和扩展,以支持在本体中对不确定性知识的描述和表达,并提供不确定性语义推理和分析能力。

(3) 目前国内外都建成有特定用途的地理本体,这些本体大都将研究的主要精力放在了如何从半结构或无结构的文本中自动获取空间实体和关系以及实例上,而对基于这些本体知识库进行语义分析和推理的研究相对的投入力度很不够,使得目前地理本体的作用没有得到有效地发挥,地理本体中的领域知识没有得到充分的发现和利用。因而,需要针对本体知识库中知识组织的特点,选择和设计适用的语义分析和推理方法,并对本体的推理机制进行改进和扩展,充分发掘其中有用的语义关联信息,用于辅助空间信息的检索、处理和融合等。

总体来说,目前本体的描述语义主要侧重于对概念从属关系、概念间相似关系的描述,这两方面的信息对于语义的描述和理解都停留在了概念关系级别,仅可用于描述各种歧义概念间的映射和匹配关系,使得其所描述的语义变得很有限。此外,现有语义 Web 的实现也主要依赖于对各类 Web 资源进行概念级的语义标记,用于实现对这些资源所提供功能的语义理解。因此,要将本体更好地应用于自然语言检索中,需要解决的关键问题是如何在本体中形式化地描述自然语言理解所需要的语义知识,真正地实现语义层面的自然语言分析和理解,而不是仅仅停留在词法或语法层次,需要寻求一种适用的语义描述机制或模型,同时还需要提供相应的不确定性知识描述和推理机制,解决自然语言理解和语义检索中的相似性检索和非精确推理等问题,尽力增强其语义解析能力,以改善基于本体的自然语言检索的应用效果和适用范围。

1.3 本书主要研究内容和特点

本书针对以上空间信息语义检索中的关键问题,从以下几个方面进行深入的分析 and 研究,主要工作和贡献如下:

(1) 针对本体对自然语言中动词解析能力的局限性,研究如何利用语言认知模型来扩展本体模型的语义描述能力,借鉴语言模型中对动词和名词搭配关系的描述机制和分析方法,试图从句类的表达和分析角度,解决基于本体的动词词汇表达和组织、动词搭配关系的描述和解析、用户检索意图的获取和本体化描述等技术问题,以实现空间信息自然检索语句中用户检索需求的语义理解和形式化表达。

提出了本体模型和语言模型相结合的自然检索语句解析方法。针对本体对自然检索语句中动词理解能力的局限性,尝试将语言认知 HNC 模型和本体模型相结合,试图利用 HNC 模型中对谓词的表达和解析方法,从句类分析角度来解决自然语句中动词和名词不同搭配的含义理解问题,获得以语义关联图形式描述的用户检索需求,实现自然检索语句到本体形式化描述语句的转变,为基于本体的信息检索和推理提供初始条件和检索依据。

(2) 针对本体对不确定性语义关系描述能力的局限性,研究如何利用非经典逻辑来扩展本体模型的不确定性知识描述能力,借鉴非精确逻辑中不确定性知识描述机制和度量方法,试图从语义关系的定量化度量角度,解决不确定性空间信息领域知识的本体化描述、语义关联强度的定量化度量、语义关系度量值的推算和传递等技术问题,以实现空间信息领域中各类不确定性知识的本体化描述和定量化度量。

建立了基于可能性逻辑和概率统计方法的语义关系定量化描述机制。针对本体对不确定性知识描述的局限性,尝试将反映主观经验的可能性度量和代表客观特征的概率统计方法相结合,试图综合利用可能性逻辑的人性化和概率统计的全局性特点,解决各类不确定性语义关系的定量化描述问题,实现语义描述机制从定性到定量的转变,为本体的相关性检索和不确定推理提供语义强度的定量化度量手段。

(3) 针对本体对非精确信息检索和推理分析的局限性,研究如何利用启发式检索和推理算法来实现空间信息本体的相似性检索,综合利用启发式算法的关联性检索方法和推理规则的潜在语义表达能力,试图从检索过程的定量化度量和语义化关联角度,解决基于本体的相关性信息检索、语义关联性推理、检索过程的不确定性度量、检索结果的定量化对比等技术问题,以实现空间信息的非精确检索,以及检索过程中不确定性的定量化度量和推算。

提出了基于激活扩散 SA 算法的本体相关性检索和不确定推理方法。针对定量化语义知识空间中信息检索的近似性和推理的非精确性,尝试利用语义关系的分类度量方法来对激活扩散算法进行改进,将本体相关性检索分为多个阶段来进行,试图提高激活扩散过程的针对性和导向性;尝试将激活扩散算法和本体推理规则相结合,利用激活扩散的路径检索能力和本体推理规则的潜在语义关系描述机制,试图在激活扩散获得的语义关联网中依据推理规则发现隐含的语义关联,实现更多相关信息结点的推理发现,为检索需求和匹配资源间的相似性检索提供匹配和推理机制。

(4) 综合利用上述研究成果,设计并实现了空间数据服务的语义化查询原型系统。该系统以包含有检索需求的自然语句为输入,以带有定量化度量值的数据服务检索列表为输出。本书对该系统中所涉及的检索需求的本体化描述、空间数据服务的本体化描述、基于激活扩散算法的空间数据服务检索和匹配等关键技术进行了说明。最后通过两个典型检索案例的试运行结果来分析该系统的优势和不足,并对本书的相关研究成果进行了总结和评价。

本书针对空间信息服务系统智能化检索中难题,探索基于语义本体的解决方案,并对其关键技术问题进行深入的研究和探讨。针对本体对自然语言表达和理解的局限性,将语言认知模型和本体模型相结合,尝试解决空间信息自然检索语言的解析和用户检索需求的语义化描述等问题;针对空间信息检索过程中存在的不确定性因素,采取不确定语义关系的本体定量化描述机制,并采用启发式关联语义检索算法和非精确推理技术,尝试解决用户检索需求与空间信息资源间的相关性匹配等问题,力图提高空间数据服务中信息检索的效率和智能化程度。

1.4 本章小结

本章从知识工程和语义检索的角度,调研了空间信息检索的发展现状,分析了空间信息检索的特点和特殊需求,论述了空间信息服务系统中存在的语义问题,并对这些问题的解决给出了初步的设想和研究思路,最后对本书的主要研究工作和特点进行了介绍,便于帮助读者从总体上把握本书的脉络和结构。

本章参考文献

- [1] 陆桑璐,周晓方,陈贵海,谢立. 空间信息检索及其数据库概化技术[J]. 软件学报, 2002,13(8):1534-1539.
- [2] 李善平,尹奇(韦华),胡玉杰,郭鸣,付相君. 本体论研究综述[J]. 计算机研究与发展, 2004,41(7):1041-1052.
- [3] 黄茂军. 地理本体的关键问题和应用研究[M]. 合肥:中国科学技术大学出版社, 2006.
- [4] 宣瑄,刘建波. 遥感数据自然语言检索技术研究[D]. 北京:中国科学院大学, 2013.