

基于人机交互的 知识管理系统

Knowledge Management System Based
on Human-Computer Interaction

滕明岩 郑 福 苏亚坤/著



科学出版社

基于人机交互的知识管理系统

滕明岩 郑 福 苏亚坤 著

科学出版社

内 容 简 介

语义网技术的发展,使管理海量规模的语义网数据成为巨大挑战。本书介绍如何使用人机交互技术更为有效地进行知识获取、集成和检索,从而更好地构建基于人机交互的知识管理系统。主要内容有三部分:基于人机交互的知识获取及集成、基于人机交互的知识检索和基于人机交互的知识管理系统原型。

本书可供计算机理论与应用、知识管理和知识工程以及其他相关专业的高年级本科生、研究生和科研工作者参考。

图书在版编目(CIP)数据

基于人机交互的知识管理系统/滕明岩,郑福,苏亚坤著. —北京:科学出版社, 2017.10

ISBN 978-7-03-054494-0

I. ①基… II. ①滕… ②郑… ③苏… III. ①人-机系统-应用-企业经营管理-知识管理-管理信息系统 IV. ①F270.7

中国版本图书馆 CIP 数据核字 (2017) 第 224530 号

责任编辑:张 震 姜 红/责任校对:贾伟娟

责任印制:吴兆东/封面设计:无极书装

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

北京厚诚则铭印刷科技有限公司

科学出版社发行 各地新华书店经销

*

2017 年 10 月第 一 版 开本:720×1000 1/16

2018 年 1 月第二次印刷 印张:7 3/4

字数:120 000

定价:75.00 元

(如有印装质量问题,我社负责调换)



前 言

以微电子、信息技术为基础，计算机、通信和互联网等为核心的技术革命，以及由此引起的经济和社会的发展产生的信息革命，正在迅速向前发展。互联网（包括移动终端、移动互联网）全球化普及成为信息革命的重要标志。互联网的高度发展，允许广大网民非常便捷地制造和消费网络数据，在众多领域积累起规模越来越庞大的数据，于是产生了大数据。其本质是当数据的规模和复杂性超过现有数据处理平台的软硬件能够处理的范畴时，人们迫切寻求在有限的成本下高性能的大数据处理解决办法。云计算的出现，推动了大数据研究的发展。本书从系统的角度解决一类重要类型的大数据即语义网大数据的管理和应用问题。

由于自然语言处理技术的进步，语义网大数据使人们可以从互联网数据中抽取规模庞大的知识条目，不断积累形成大规模开放资源描述框架语义网知识库。它能够为用户提供如语义检索和信息推荐等重要的知识服务，在现代搜索引擎中发挥着越来越重要的作用。为了提高知识服务的质量，知识库需要不断丰富和完善来自不同领域的知识内容，需要从众多知识源头获取新知识，并将新老知识编织（集成）到一起。随着资源描述框架（RDF）语义网知识库规模的不断壮大，它具备了典型的大数据系统特征：规模庞大、类型多样、不断变化。因此，人们将其称为语义网大数据，或者知识大数据。尽管人们不断提高知识抽取的精度，但受制于机器在形象思维和语义辨别方面的能力，总有一些带有瑕疵的知识被抽取并融入知识库系统中，加之知识源头众多、领域宽泛、用户需求多样化，RDF 语义网知识库成为开放复杂巨系统，其有效管理和利用成为巨大挑战。本书从知识大数据管理和应用出发，利用开放复杂巨系统中定性定量集成方法的思路，研究在知识大数据的管理和应用过程中，通过人的参与和反馈，更好地解决知识大数据在获取、集成和检索方面面临的一些难题。

本书的主要研究工作和创新点如下:

(1) 在知识获取和知识集成方面, 研究如何在现有知识挖掘和知识抽取技术的基础上, 高效并且准确地从互联网上获取新知识, 并将这些知识集成到知识库中。提出具有自感知能力的知识获取方法和自底向上的知识集成方法, 通过引入人工反馈来校验知识集成的准确性和评价知识源数据质量的重要性, 用以提高知识库系统完成知识获取和集成的能力。整个知识获取和集成的流程采用多线程流水线设计, 根据不同环节的处理能力, 动态调度系统资源, 协调知识获取和集成各个环节高效地进行。

(2) 在知识检索方面, 重点解决因为用户查询的语义歧义性造成的知识检索结果质量不高的问题。提出在知识检索过程中使用谓词-关键字对来实现对查询的语义限制, 通过设计高性能的谓词-关键字对的组合推荐算法, 在尽可能少的用户反馈的基础上, 更高效地排除关键字查询中的歧义, 从而搭建起现有知识检索解决方案中结构化查询与关键字查询之间的桥梁。

(3) 将系统工程中人机交互的思想应用在知识大数据的获取、集成和检索问题上, 以解决机器在管理大规模知识库过程中知识质量低、检索结果差的问题。实现一个基于人机交互的知识获取和集成、知识检索的原型系统, 在实践中进一步检验所提出的人机交互技术的有效性, 加强人机交互在大规模知识管理系统中的重要性。

本书在编写和出版过程中得到了辽宁省数学一流重点学科和国家自然科学基金(61403043、11201037)的支持, 在此表示由衷的感谢。本书的部分内容是作者在中国航天科工集团第二研究院攻读博士学位时取得的成果, 因此特别感谢中国航天科工集团第二研究院和北京信息控制研究所在本人攻读博士学位期间给予的支持。

本书是作者对近年来研究工作进行系统整理的初步尝试, 由于作者学识水平有限, 本书可能存在不足之处, 希望广大读者批评指正。

作 者

2017年5月

目 录

前言

第 1 章 绪论	1
1.1 研究背景与意义	1
1.1.1 RDF 语义网知识库	3
1.1.2 RDF 语义网知识库知识管理的挑战	4
1.1.3 基于人机交互的知识管理	8
1.2 主要内容	10
第 2 章 知识管理的基础知识	14
2.1 知识库及知识管理	14
2.2 知识获取与集成	15
2.3 知识检索	17
2.4 关键字检索	19
2.4.1 图上关键字检索	20
2.4.2 RDF 知识库上的关键字检索	22
2.5 交互式信息检索	23
2.6 基于综合集成方法的知识管理	24
2.7 本章小结	25
第 3 章 基于人机交互的知识获取与集成	26
3.1 大规模知识库的知识获取与集成	26
3.2 具有自感知能力的知识获取	28
3.2.1 知识库丰度的自感知	28
3.2.2 知识源重要性的自感知	30

3.3	自底向上的知识集成	31
3.4	基于人机交互的知识获取与集成整体方案	34
3.5	自底向上知识集成方法实验	36
3.5.1	实验环境	36
3.5.2	实验结果及分析	39
3.6	本章小结	41
第4章	基于人机交互的知识检索	43
4.1	知识库的检索技术和存在的问题	43
4.2	基于关键字查询的知识检索	46
4.2.1	关键字检索基本定义	46
4.2.2	结果相关度评价	48
4.2.3	计算关键字匹配节点	50
4.2.4	图扩展和发现匹配树	52
4.2.5	复杂度分析	53
4.3	基于语义限制的知识检索的基本概念	54
4.4	以人机交互为基础的知识检索	57
4.4.1	整体流程	57
4.4.2	人机交互的结构感知算法	59
4.5	人机交互实验研究	63
4.5.1	实验设定	63
4.5.2	纯粹关键字查询实验	66
4.5.3	人机互动知识检索实验	70
4.5.4	关键字查询系统的性能	78
4.5.5	人机交互性能实验总结	80
4.6	本章小结	83
第5章	基于人机交互的知识管理系统原型	85
5.1	系统构架	85

5.1.1 细节实现.....	89
5.1.2 交互界面.....	91
5.2 知识存储与管理	92
5.2.1 数据管理.....	92
5.2.2 字典管理.....	95
5.2.3 索引管理.....	95
5.3 知识检索.....	97
5.3.1 交互界面设计.....	97
5.3.2 知识检索执行引擎.....	99
5.4 面向人机交互的系统设计和实现	101
5.5 本章小结.....	102
第 6 章 总结与展望	103
参考文献	106

第1章 绪 论

1.1 研究背景与意义

现代电子信息技术的高速发展引发了一场信息革命，促进了互联网的迅速发展，使互联网的全球化普及成为信息革命的重要标志。互联网产业（包括移动终端、移动互联网）的迅猛发展推动了知识经济成为继农业经济、工业经济之后一种新的经济发展方式。它是一种以知识在生产中占主导地位，建立在知识和信息的生产、分配和使用基础上的经济形态。近年来，在信息化浪潮的推动下，随着大数据时代的到来，很多国家开始陆续迈向知识经济时代。知识经济时代有很多有别于工业经济时代的特征^[1]：知识与信息成为企业、机构乃至国家重要的无形资本；以知识经济为导向的知识密集型产业在国民经济中所占的比重越来越大；对从事知识经济产业的劳动者的知识水平和技能要求越来越高；以知识和信息为主体的创新越来越成为企业发展的关键。例如，互联网搜索引擎公司（谷歌和百度等）、社交网站（脸谱和微博等）、电商网站（亚马逊和淘宝网等）的迅猛发展，中关村传统电子卖场的没落，这些都能看出知识经济时代的来临势不可挡。

互联网之所以能够推动知识经济的发展，与大数据的出现密切相关。在互联网时代，广大网民可以非常便捷地制造和消费网络数据，在众多领域积累起规模越来越庞大的数据，于是产生了大数据的概念。大数据的种类有很多，如互联网网页数据、日志数据、社交网络数据、科学研究数据等，这些数据大部分与互联网相关。这些大数据的本质是当数据的规模和复杂性超过现有数据处理平台的软硬件能够处理的范畴时，人们在有限的成本下迫切寻求高性能的大数据处理解决办法。大数据通常蕴含着很多重要的价值，而其中更为重要的是知识。借助自然语言处理技术的进步，人们可以从互联网大数据中抽取出规模庞大的知识条目，

不断积累而形成大规模的、开放的资源描述框架(resource description framework, RDF)语义网知识库。这些面向开放领域的知识库,能够在很多与知识经济相关的产业中发挥重要作用,为众多的应用领域提供如语义检索和信息推荐等知识服务。

在知识经济模式下,知识的积累、利用和创新越来越重要。随着企业、机构和政府等在数据和知识方面的不断积累,如何有效地对知识的获取、更新和使用过程进行系统的管理,就显得尤为重要,这就是知识管理。其目的是系统地处理、寻求、理解和使用知识,以从知识资产中得到回报、创造价值^[2]。在知识管理体系中,知识生命周期的概念非常重要,McElroy^[3]早在1999年就提出把知识生命周期分为知识产生、知识声明、知识验证、知识整合、知识评价反馈五个阶段,这五个阶段构成一个完整的闭环系统。随着知识管理研究的进展,知识生态系统的概念越来越受到人们的重视,它既包括知识管理的各要素(人、知识、组织、技术、知识流程等),又包含这些要素的相互作用机制。它是动态的、开放的系统,各要素之间既相互联系又相互影响,随着知识环境的改变而不断进行演化^[4]。总体来讲,尤其是在互联网环境下,知识生态系统具有适应性、系统性、开放性和动态性等复杂巨系统的一些特征^[5]。

在知识和信息迅速膨胀的今天,针对如何从技术手段上有效地实现知识管理,产生了知识工程这门学科。它主要研究设计和实现知识库系统及其应用系统的理论、方法和技术,是研究知识获取、知识表示、知识管理和知识运用的一门学科^[6]。知识工程的发展借助了人工智能、数据库技术、数理逻辑、认知科学和心理学等多学科的交叉发展。传统信息产业的数据管理主要针对的是结构化数据,而知识远比结构化数据复杂,会涉及很多非结构的信息,从而造成了知识管理的高复杂性和跨学科的特点。在这样的背景下,文献[7]和[8]提出了利用系统工程的思想和方法来综合研究知识管理中的问题,形成了知识系统工程这样一门研究对知识进行组织和管理 的学科。文献[9]和[10]指出在系统综合集成方法的指引下,把人、知识和工具结合起来,把定性方法和定量方法、理论和实践、动态过程和静态过程、宏观层次和微观层次、集中状态和分散状态等集成起来;它是从整体

上思考和解决问题的系统方法论。本书的研究也是受此思想启发,针对大规模知识库的获取、集成和检索过程,研究如何借鉴系统工程的一些思想,更好地利用和发挥知识库的价值。

1.1.1 RDF 语义网知识库

无论是知识管理还是知识系统工程,一个重要的研究内容就是知识的表达、获取、集成、存储和检索。在传统意义上,知识库中的知识源于领域专家,它是求解问题所需领域知识的集合,包括基本事实、规则和其他有关信息^[6]。知识表示就是对知识的描述,其采用一组特定的符号将知识编码成便于计算机处理的数据结构。知识表示是人工智能和知识工程中使用知识库求解问题的基础,主要以本体的形式存在。近年来,随着语义网技术的发展,在以互联网之父 Tim Berners-Lee 为代表的计算机科学家的推动下,知识的表示开始走向标准化、开放化和互联网化^[11]。其中一个典型代表就是 RDF 的提出,它是国际万维网(world wide web, WWW)提出的互联网上知识和信息描述的基本框架,如今已成为很多互联网和 RDF 语义网知识库用来表达信息实体的描述规范^[12]。RDF 的基本原理是采用主语(subject)、谓词(predicate)、宾语(object)的三元组形式来表达信息实体(主语)在某个谓词上所具有的属性值(宾语),其中,主语和部分宾语一般采用统一资源标识符(uniform resource identifier, URI)的形式来唯一标识一个信息实体。例如,要表示 Albert Einstein 出生在德国,可以使用如下形式的三元组:

```
<http://www.w3.org/rdf#Albert_Einstein,  
    http://www.w3.org/rdf#bornin,  
    http://www.w3.org/rdf#Germany>
```

其中, `http://www.w3.org/rdf#Albert_Einstein` 代表一个人物的信息实体的 URI; `http://www.w3.org/rdf#bornin` 表示的是一个谓词,代表一个人的出生地; `http://www.w3.org/rdf#Germany` 则对应于宾语——德国。由于三元组为主语和宾语建

立了语义关联,所以数据有了更为丰富的语义信息。RDF 采用的这种信息描述方式很好地表达了互联网和语义网实体的属性和实体之间的关联关系,也为不同应用程序之间在不丧失语义的情况下进行数据交换奠定了基础。

在这样的背景下,RDF 成为语义数据描述的一种典型的标准,被广泛应用于描述知识库中的数据。尤其是近年来,随着信息抽取和数据集成等技术的发展,出现了规模越来越大的用 RDF 表达的知识库,如 DBpedia^[13]、Freebase^[14]、YAGO^[15]等。这些数据库采用 RDF 来表达网络资源及其之间的联系。例如,原语义网技术公司 Metaweb (2010 年被谷歌收购)维护的 Freebase 知识库中,可以用 RDF 表示包括电影、体育、化学、生物、地理、生物医学等众多领域的知识条目信息。随着更多 RDF 数据集的出现,人们还提出了关联数据(linked data, LD)的概念,用于将不同组织机构发布的数据关联起来,形成规模更为庞大的 RDF 数据集。截至 2012 年 3 月,LD 收集的三百多个数据集已经包含了超过 520 亿条 RDF 三元组,从 LD 的发展趋势上可以看到,整个互联网上关联数据的规模急剧膨胀^[16]。

由于很多海量 RDF 数据集包含大量来自不同领域的实体以及实体之间的关联信息,所以也常被称为 RDF 语义网知识库。一些应用开始借助 RDF 语义网知识库所能提供的知识,支持实体检索、语义检索、问答系统等应用,谷歌的知识图谱^[17]就是其中的一个例子。如今,以 RDF 语义网知识库为主体的知识图谱受到越来越多的搜索引擎公司的重视,在其检索产品中发挥着越来越重要的作用。随着互联网技术的不断发展,尤其是在以语义为核心的 Web3.0 相关技术领域^[18],存储海量 RDF 数据的 RDF 语义网知识库成为基于语义的信息查询和知识检索等重要应用的基础平台。RDF 语义网知识库在大数据的背景下发展起来,是具有很多大数据特征的、开放的、复杂的巨系统。这些特点必然为其基础上的知识管理带来很多新的挑战,这也是本书所要研究的核心问题。

1.1.2 RDF 语义网知识库知识管理的挑战

RDF 语义网知识库的技术发展,可以使来自不同领域的知识通过 RDF 的表

达形式集成起来,提供面向众多领域的知识服务。为了能够提供高质量的知识服务,知识库需要不断丰富和完善知识内容。因此,开放性是 RDF 语义网知识库的一个重要特征,它可以兼容来自不同领域、不同数据源的知识条目,并利用信息抽取技术,从众多知识源头不断获取新知识,再通过这些知识条目的相互引用,将更多的知识编织(集成)到一起。所以,RDF 语义网知识库包含了海量的、来自众多领域(知识库子系统)的实体信息。各个子系统知识库的知识条目可以互联关联,形成不同子系统之间的信息关联和信息交互。

随着 RDF 语义网知识库规模的不断壮大和知识获取途径的不断增多,相对于传统意义上专家系统中的知识库,RDF 语义网知识库规模非常庞大(可以包含上亿个节点)。如果考虑知识条目的来源是大规模的互联网网页,整个知识库系统就完全符合大数据的特征^[19]:数据量大,数据多样性强(既有结构化的三元组信息,又有非结构化的网页信息),数据是动态和开放的,结构复杂(实体之间关联关系的多样性)。因此,RDF 语义网知识库中的数据具备典型的大数据特征,可以看成知识大数据。

从复杂巨系统的角度来看,面向开放领域的 RDF 语义网知识库,作为知识大数据的载体,其本身也是开放的复杂巨系统。其开放性,更多地体现在系统与外界的信息交换。一方面,系统需要不断从外界(互联网上)获取新知识,子知识库系统之间也通过 RDF 数据搭建的桥梁,频繁地信息交换与融合,有着紧密的联系。另一方面,从用户的角度出发,知识库需要在知识服务中和人发生交互,需要为不同领域、不同需求的用户提供知识服务,并且还可以利用一些专家的交互和反馈,完善知识库系统的信息质量。知识大数据的复杂性也体现在多方面。首先,知识库包含众多领域的信息实体,这些信息实体之间又有各种各样的关联关系,从整个知识图谱的角度来看,知识库本身非常复杂,尤其是在知识检索过程中,如何准确地理解用户的检索意图,如何在复杂的知识网络中定位和用户查询相关的信息片段,具有非常大的挑战。其次,知识库融合了多知识库的信息,信息种类多且这些信息之间存在冗余重复、不一致等因素,互相之间所采用的数据源的差异性也很大,这都增加了系统的复杂性。再

次,不同知识库子系统的结构也随着知识的不断获取和更新而不断发生演化,是一个动态的复杂结构。从上述角度来看,RDF 语义网知识库是不折不扣的具有大数据特征的开放复杂巨系统。

大数据的这些特征也给 RDF 语义网知识库的知识管理带来了很多新的挑战,本书从知识管理的两个重要方面即知识获取和集成以及知识检索入手,论述目前的研究工作存在的挑战。

1. 知识获取和集成的挑战

RDF 语义网知识库之所以应用范围广、重要性高,源于其所具有动态性和开放性的大数据特征。知识大数据需要从众多的知识源头获取新知识,不断地将新知识融入系统现有的知识条目中,以持续不断地维护知识库中知识条目的新颖性、丰富性,并且不断增大知识库的规模,涵盖更多的信息实体。现有的 RDF 语义网知识库知识获取和更新的渠道不尽相同,有全自动化的知识抽取方法^[20],也有使用基于众包思想^[21]的人工编辑技术。

基于全自动化的知识抽取方法^[20]采用自然语义处理中信息抽取方面的技术,自动化地分析网页中蕴含的知识条目,如 YAGO^[15]和 DBpedia^[13]。其优点是知识获取的效率高,能够在很短的时间内从大量的网页中收集到很多知识,不足之处是知识抽取的精度有限。尽管目前很多知识抽取方法在高质量的知识源上能够达到 95% 以上的抽取精度,但对于质量要求非常高的 RDF 语义网知识库,这样的精度还有很大的提升空间。为了获取更多的自动知识抽取精度比较低的不常见的实体,尤其是在需要增加知识召回率的前提下,对知识抽取的精度要求就更为重要。在知识更新方面,这些方法不适合主动地获取新知识,没有采取增量地管理和维护知识库中知识条目。另外一种知识获取和集成方法采用的是人工编辑的方式,即集群众智慧获取新知识,如 Freebase^[14]和 Wikidata^[22]。然而,由于这种方法基于纯粹手工编辑的方式,知识积累的效率比较低,而且由于用户水平参差不齐,知识的质量也很难得到很好的控制。

传统的数据集成方法普遍建立在结构化数据表格间的模式匹配基础之上^[23]。然而,一方面,在 RDF 语义网知识库的背景下,知识库中的信息条目并不存在数据模式信息,也就是不存在数据表的概念;另一方面,从互联网上抽取的数据可以是数据表格的形式,也可以是基础实体关系对的形式,这都与传统的数据集成方法不同。因此,需要研究知识库与互联网上抽取的知识条目之间的知识集成问题。

2. 知识检索的挑战

随着大量 RDF 知识库和数据集的出现,如何让用户高效地从知识大数据中查询和检索到自己感兴趣的数据或知识片段,成为越来越棘手的研究问题。最近一段时间内,针对这一问题,学术界和工业界都开展了大量的研究工作。例如,德国的马克斯-普朗克研究所(Max Planck Institutes, MPI)提出了一个称为 RDF-3X 的 SPARQL 查询引擎^[24]、Sindice^[25]和 Swoogle^[26]等语义搜索引擎。然而,目前的解决方案存在的一个较大问题是缺少既表达能力强又简单易用的 RDF 数据查询和检索方法。

目前,常见的 RDF 数据查询检索方法有两种:使用关键字查询 RDF 数据或者使用 SPARQL 查询语言检索 RDF 数据。关键字查询的目的是在 RDF 知识库中,找到包含所有关键字的、结构紧凑的子图(或者树结构)^[27]。其虽然灵活度大、实用性强,却很难保证查准率和查全率,并且关键字查询的语义表达能力弱,不能对 RDF 图数据给出结构上的约束。结构化的 SPARQL^[28]查询语言力图在知识库中找到满足 SPARQL 查询条件的子图,其有着较为复杂的语法定义,用户需要熟悉它的语法规则并了解 RDF 数据的模式信息(如谓词和前缀等),才能使用该语言查询 RDF 数据。这对于包含简单模式的一些垂直应用尚可,但对于谓词数量繁多的、面向开放领域的海量 RDF 知识库,SPARQL 语言对于普通用户甚至专业开发人员都不具备良好的实用性。面对结构复杂、规模庞大的 RDF 知识库,用户通常很难明确自己的信息需求,也很难通过简单的查询检索到理想的结果。

1.1.3 基于人机交互的知识管理

RDF 语义网知识库在知识管理上所面临的挑战与其本身具有大数据的特征密切相关。在现有的大数据管理和分析解决方案中^[29], 为了追求数据分析和数据处理的性能, 普遍采用自动化的方式。然而, 众所周知, 尽管计算机处理信息的速度非常快, 但由于其在形象思维和语义理解方面的不足, 自动化大数据处理的精度必然受限。在大数据分析和挖掘领域较为普遍的观点是: 精度不足可以通过数据量的增加, 从统计意义上予以弥补^[30]。然而, 对于精度高度敏感的知识大数据的应用, 很多情况下, 精度的不足是无法容忍的。那么, 面对开放的、复杂的、动态的知识大数据系统, 如何既能高性能地实现知识管理, 又能以较高的精度要求管理知识和利用知识, 成为要解决的一个重要难题。

在系统工程领域著名的综合集成方法中, 于景元等^[31]很早就提出了以人为为主的人机交互的思维方式和研究方式。他强调, 采取机帮人、人帮机的合作方式, 机器能做的尽量由机器去完成, 极大地扩展人脑逻辑思维处理信息的能力。通过人机交互以人为为主, 实现信息、知识和智慧的综合集成, 通过人机交互、反复比较、逐次逼近, 实现从定性到定量的认识, 然后再提出新的经验性假设, 继续进行定量研究, 这是一个循环往复、不断深化的螺旋式上升过程。受这一思想的深刻启发, 在 RDF 语义网知识库的知识管理过程中, 也可以采用人机交互、人机互补、逐次逼近等思路。就实质而言, 定性、定量相结合的综合集成方法是通过专家、数据和各种计算机技术的有机结合, 实现多学科的科学理论与人的经验知识的结合。其思想很适合管理具有大数据特征的 RDF 语义网知识库开放复杂巨系统, 它一方面充分发挥出计算机处理能力强的特点, 部分解决复杂和巨大方面带来的挑战; 另一方面充分发挥专家的特长, 利用人的智慧解决计算机所难以自动解决的开放和复杂方面的挑战。

在综合集成方法下, 计算机系统通过仿真运算、系统优化, 在复杂的系统背景下往往只能得到次优的或者并不能让人非常满意的策略。在这样的定量分析基

基础上,如果引入专家的科学经验和知识,通过理性的分析和感性的判断,就可以对计算机得出的结果进行修正和补充。当然,用户或者专家的反馈也未必是完全正确的,计算机系统在接收到用户的反馈后,可以通过调整模型或者参数,重复进行优化计算,通过这样的定性分析和定量分析相结合的办法,解决开放复杂巨系统面临的难题。现在的大数据研究普遍重视机器算法方面的研究,适合解决一些直观的、简单的问题。由于没有很好地借鉴系统综合集成方法,在面对一些和开放复杂巨系统相关的难题时,纯粹依赖计算机的方式处理这些问题就会面临很大的挑战。

实际上,已经有人将人机交互应用在知识处理中。朝乐门^[32]提出了人机交互知识处理的理念,并提出人机交互知识管理是指在知识管理,尤其是在基于语义网的知识处理过程中,强调人与计算机的分工与合作,通过人对知识处理前端控制,降低计算机知识处理的难度,在人与计算机之间寻找最佳的结合状态,推动在知识的创造、表示、存储、检索、推理、验证、抽取、再现、集成能力上超过人与计算机的超知识处理系统的出现。在这种超知识处理系统中,人与计算机共同感知、共同决策、相互学习、相互监督,共同完成知识管理任务。

通过借鉴上述研究的思路,本书将人机交互的理念应用在知识管理的两个重要问题上:知识获取和集成,知识检索。从解决大数据面临的一些复杂问题的角度,围绕 RDF 语义网知识库这个开放复杂巨系统,探索由人、专家系统及智能机器相结合构建而成的人机交互系统。并试图通过人机的互相协调配合,在人的指导和决策下,完成复杂的知识管理方面的难题,进而形成人机结合的、和谐的知识库系统。

在知识检索方面,随着大数据研究的兴起,以交互式的查询和检索数据逐渐引起人们的关注^[33]。本书认为这种交互式检索非常适合为存放海量 RDF 数据的 RDF 语义网知识库提供一种新的、支持交互性的数据检索接口,通过人的参与与反馈,更深层次地理解用户的查询语义,提高知识检索的精度。然而,人参与其中的信息检索应用要求系统具备非常强的交互性能,与交互相关的查询处理时间要能够达到亚秒级别。由于当前 SPARQL 查询处理系统的性能^[34],交互性信息检