信息管理
新视野论丛

Knowledge Retrieval in Semantic Libraries:
语义图书馆知识检索的方法与应用
Methodology and Practice

楼 雯 著

上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

Knowledge Retrieval in Semantic Libraries:

语义图书馆知识检索的方法与应用

Methodology and Practice

楼 雯 著

上海科学技术文献出版社
Shanghai Scientific and Technological Literature Press

图书在版编目（CIP）数据

语义图书馆知识检索的方法与应用 / 楼雯著. —上海：上海科学技术文献出版社，2019
ISBN 978-7-5439-7846-1

Ⅰ.① 语… Ⅱ.①楼… Ⅲ.①智能检索系统 Ⅳ.① G254.929.1

中国版本图书馆 CIP 数据核字 (2019) 第 049560 号

责任编辑：徐 静
封面设计：袁 力

语义图书馆知识检索的方法与应用
YUYI TUSHUGUAN ZHISHI JIANSUO DE FANGFA YU YINGYONG
楼 雯 著
出版发行：上海科学技术文献出版社
地 址：上海市长乐路 746 号
邮政编码：200040
经 销：全国新华书店
印 刷：常熟市人民印刷有限公司
开 本：787×1092 1/16
印 张：11.25
字 数：273 000
版 次：2019 年 5 月第 1 版 2019 年 5 月第 1 次印刷
书 号：ISBN 978-7-5439-7846-1
定 价：68.00 元
http://www.sstlp.com

出版说明

情报学是一门研究世上一切信息、情报的交叉学科，它会运用到语言学、传播学、数学、计算机科学等学科的知识、方法和工具。但情报学提供给学界、业界的，是方法，是手段。本书是为未来情报服务工作提供知识组织的一种方法、一种手段。计量学是情报学的特色，作者在多年的计量学应用研究实践中发现，将计量学与信息组织的方法和内容进行融合，得到的新方法和新技术可以利用到知识组织与知识管理中，最终利用数据可视化的方法和技术，对知识进行可视化展示，因此，将这几个研究方向联系到一起，做到以知识显性化为目标，希望能为在知识海洋中寻找方向的人们提供帮助。

本书是作者从事科研工作至今的总结，其研究成果从局部到整体、从理论到实践的各个层面论述了知识组织的形式——资源本体的概念、关键技术、实现目标、系统地位和构建过程，利用资源本体可以无领域界限的实现半自动资源语义化。一方面，在阐述资源本体构建的过程中，强调了机器自动识别的过程，但需要人工干预信息计量的部分工作，做到半自动；一方面，资源本体中由于融入了信息计量学的内容，挖掘出领域外的隐性知识，做到无领域界限；另一方面，从研究整体上看是一套图书馆资源语义化的实现过程规范，加之关键技术和实现目标的描述，可以作为图书馆资源语义化实践的指导，为建设图书馆资源成为语义网内容的一部分做贡献。

本书共分为 7 章，其中第 1 章对图书馆和知识检索的相关概念及理论进行介绍，第 2 章是分析认知科学视角下的图书馆知识检索规律与影响，第 3 章是语义图书馆内涵与外延的界定，第 4 章介绍了语义图书馆知识检索模型，第 5 章展示了语义图书馆知识检索系统的设计过程，第 6 章是以武汉大学图书馆和美国俄勒冈州波特兰市的 Powell's City of Books 书城为例，展示了语义图书馆知识检索系统是如何实现和运作的。

本书受国家社科基金青年项目“学者驱动的学术资源语义共享模式及其应用研究”(17CTQ025)资助。

目 录

绪 论

快节奏的生活、高速发展的社会、新兴的科学技术、多元的学科融合，这些是我们身处全球知识时代的显著特征，带给我们科学工作者巨大挑战的同时，也处处充满着机遇，在这样的时代背景、技术背景和学科背景三个方面的共同促进下，本书将研究目标锁定在图书馆知识关联，以期达到提升图书馆知识服务和促进语义化方法和技术的目的。

大数据作为“第三次浪潮的华彩乐章”已经成为当代社会人人知晓且不得不提及的词汇，正如著名未来学家阿尔文·托夫勒(Alvin Toffler)早在20世纪80年代的预测[1]，大数据在21世纪初卷起的研究浪潮可称为大数据时代。不像人们用了一万多年的时间从农业时代走向工业时代进行了第一次浪潮到第二次浪潮的变革，信息化的第三次变革是悄无声息但又爆发式发生的，既是由于瞬息万变的社会，也是由于高速发展的信息技术，正如马云所说，在好不容易实现全民个人电脑时代时，移动互联网的概念猛然充斥了人们生活，在人们还没完全接受移动互联网时代时，大数据时代已经来临了[2]。

从国家层面上，如今，数据无孔不入，已经像水土资源、劳动力资源、资本资源等融入人们的现实生活，从2008年开始，依靠IBM、谷歌、微软等互联网信息技术行业巨头的技术，全球各国展开了一次大数据强国发展战略制订的追逐赛，美国、法国、英国政府牵头建立开放政府数据研究所、平台和数据库等；德国、韩国、日本政府倚重国家高科技能力着重建设智能数据平台，如智慧城市、智能工厂等；美国、加拿大、英国、欧盟确立了支持大数据驱动的科学研究、教育体系、国家第一和第二生产力、国家安全的战略。我国政府长期重视科学技术发展力，尤其是党的十八届五中全会公报提出要实施“国家大数据战略”[3]的这一举措更是顺时顺势，战略从政府、立法、市场、安全四个方面指导全民进入大数据时代，我国的大数据发展战略要从构建平台、突破技术、普及民众的这些方面入手。

从行业层面上，大数据应用前景是广泛存在于人们生活中的，比如医疗行业搭建大数据平台，从而更高效地掌握医院运作情况、患者和药品出入情况；生物信息行业利用大数据技术可以对基因组进行大规模数据分析，加快生物信息学科学研究的进程；金融行业利用大数据技术的帮助可以进行精准的市场定位、决策支持、风险监管等。这样也可以看出，大数据时代最大的特征之一就是信息技术的前沿性，可以说，没有技术支持不可能实现大数据，技术工作人员包括科研工作人员则是大数据时代的主力军，我们图书情报学研究人员作为信

---

〔1〕杨敏. 大数据向人类认知方式提出新挑战[N]. 中国社会科学报，2013-06-28.

〔2〕马云. 淘宝十周年晚会演讲[R/OL]. [2016-02-07]. http://www.gkstk.com/article/1368366177.html.

〔3〕何哲. 五中全会，大数据战略上升为国家战略[EB/OL]. [2016-02-07]. http://politics.people.com.cn/n/2015/1108/c1001-27790239.html.

息海洋中颇具指导性的一分子，更是应该以应用甚至引领大数据技术为己任，而我们学科最佳的实验基础就是图书馆（广义上的图书馆），一方面是我们对图书馆的了解和熟悉，最重要的是图书馆已不再仅仅是存储图书的场所，而是蕴含着巨大信息和文献的知识宝库，打个可能不恰当的比方，如果我们图书情报学人不能首先组织好自己所熟悉的知识领地，如何分析其他行业的信息，为其他行业提供信息支撑，如何证明我们站在时代的前沿。由此带来了数字图书馆、关联数据和语义网应用于图书馆等一系列的研究，本书正是基于此时代背景，试图利用语义网技术组织图书馆资源，从而提供特定知识给用户进行服务。

“碎片化”的原意为完整的事物被破成零散的状态。“知识碎片化”的意思是人类获得的知识不系统、不完整，这是网络时代带给人类社会学习史上的第二次大挑战。传播学理论中，信息载体决定了接收信息的方式，纸质或实体介质信息的凋零，线上学习成为主流，使得人类的认知方式和学习方式发生了重大的改变。人们从微信、知乎、微博接收信息和知识，并快速传递分享，这样的信息和知识容易被篡改、扭曲，亦因为快速传播的便利而简化了逻辑过程。长此以往，使得人们不再善于严谨的逻辑思考、全方位的分析问题，这样的现实，无论是笔者还是专家，我们每个人都很清楚，虽不是身陷其中而不自知，但却因为大环境所趋无法逃离。但事物本质的两面性特征，让我们的思考也可从两面入手，知识碎片化也可以看成是知识在广度方向上的传播，借助物理学的理论，分子的运动是杂乱的，杂乱的分子运动是稳定的。那么把知识碎片当成是分子的话，知识碎片在人类社会中的快速传播和分享是杂乱的，也可以看成是稳定的传播状态。同样分子在运动时会发生碰撞、分裂和聚合等活动，知识碎片也可以发生创新、分化和融合，并非皆是坏事。

人类对知识的探讨永远离不开科学技术的支持，正如上文的分析，知识碎片化现象的一部分成因也是由技术的快速发展带来的，笔者将之称为技术对知识的离散，辩证的来说，技术的快速发展导致了知识碎片化，知识碎片更需要卓越的技术对其整理和组织。从广义的技术来说，我们电子信息业的龙头和基础产业是软件产业和数据库产业，电信业则更需要瞄准网络化、互动可视化的重要方向，传播出版产业又需要多媒体、高速度、大容量、个性化的技术支持，从行业产业的这些技术来看，都已经离散在各个不同的需求和不同的技术层面上了。而从狭义的技术本身而言，又因为载体、使用方法、面向对象的不同，而有更多的细分。这样的离散，可以说成是技术碎片化。举一个最简单的塑料袋的制造技术，工厂和设计师要考虑到的技术包括了原材料的选择、塑料袋最后出品时的形状色泽、模具的设计、出品是否符合国际国家标准等等，单独来看，每一个过程都可以视为技术本身的离散，如果技术都已经碎片化，可想而知技术所带动的知识如何不成为碎片。

正如上文提到的，人类已然认识到知识碎片化的现象，如何解决这一现实问题呢？还是得依靠技术，笔者称之为技术对知识的融合，这种融合可以从技术的两个维度上分析。首先可以利用信息技术、大数据技术反向推导知识碎片的来源，进而找到一条信息传播渠道，沿着信息传播渠道反向追踪，对沿路和源头信息进行适当管理，这是技术对知识的一种融合方式；另一种我们需要介入建构主义的理论，建构主义认为知识不能从教师、课本本身习得而来，而是应从学习者自身通过既存事实的学习消化吸收习得而来[1]，这也是导致知识碎片

[1] SMITH A K. Beyond modularity: a developmental perspective on cognitive science[M]. Boston: MIT Press, 1992: 142.

化的其中一个原因，那么反向来思考，这些经过消化吸收的知识何尝不是一种对知识的再组织再创造呢，这是技术对知识的另一种融合方式，这种技术则是人类的学习技术。

本书的研究其实就是技术对知识的离散与融合的呈现，本书利用语义网技术首先将图书馆书目数据打散成一个个单独的知识单元，这是对知识的离散，再将书目数据之间联系起来，完成知识关联，这是对知识的再融合。

一个学科是否有其本身的核心研究内容是判断其是否独立于学科之林的重要指标，图书情报学在改革开放后的迅速发展，使得学科的核心内容也有了长足的进步，但与物理学、化学、英语这些学科的核心内容显而易见不同，图书情报学的核心内容常常不为人所熟知，尤其是情报学的学科名称更是让不知情者认为与特工间谍有关，因此图书情报学家们从未间断过对图书情报学的核心内容的阐述以及对学科的界定，特别是核心内容需要顺时而变。美国图书馆学会(ALA)曾强调图书情报学(library and information science)为研究利用各个信息技术对信息进行储存、加工、管理、利用的学科，iSchools 则认为需要强调信息(研究对象)、技术(研究方法)、人(研究主体)三者的统一[1]，那么可以简单地说图书情报学是信息工作者利用技术研究信息的学科，这里说的信息不仅仅是图书馆书目，而是世界上各式各样的信息，这里说的技术也不仅仅是排架分类技术，而是包括了统计学、计算机科学、社会学等学科引入的技术，因此提出了图书情报学是一个应用性很强的跨学科研究领域[2]。与图书情报学联系较多较为密切的学科有管理科学与工程、传播学、计算机科学、文学或语言学，它们与图书情报学有着很多的共性，跨学科研究的热潮让各个学科之间相互渗透，学科的跨界和过界研究并不特别，使得图书情报学的核心内容在模糊与清晰的来去中循环，学科界限减弱。那么在这种情况下图书情报学是否还应该引入其他学科的方法和技术，与其他学科整合？答案是肯定的，原因是跨学科研究更是大势所趋，更能发挥各学科的优势，产生更高的效应。据统计，诺贝尔自然科学奖颁布的 100 余年中，跨学科研究获奖的比率高达 52%[3]，这也印证了跨学科研究的优势。跨学科研究指的是从研究对象、研究角度、研究方法的多样性进行的科学研究，跨学科研究带来科学界全新的创造性解决科学问题的方法和手段，是当今重要的研究方向和关注重点[4]。我国政府对其的重视体现在对跨学科研究项目的资助和支持上，国家哲学社会科学研究五年规划中多次阐明要加强基础研究、新兴交叉学科、跨学科综合研究[5]，各教育机构也纷纷设立交叉学科二级博士点、硕士点等。因此跨学科研究并不是阻碍图书情报学学科界定的条件，而是挑战，只要我们认清图书情报学的核心本质甚至明确坚定学科的研究对象，就不会在学科之林中迷失自我。

本书的研究对象是图书馆信息资源、知识资源，涉及图书馆学知识，研究方法涉及情报学方法和计算机科学技术。另外引入用户认知，是信息行为研究的一部分，脱离心理学和认知科学研究信息行为是不可取也不全面的，因此本书综合了图书情报学、计算机科学、心理

---

〔1〕叶继元. 图书情报学(LIS)核心内容及其人才培养[J]. 中国图书馆学报，2010(6)：13-19.

〔2〕全国哲学社会科学办公室. 国家哲学社会科学"十一五"研究状况与"十二五"发展趋势[M]. 北京：社会科学文献出版社，2011：1638.

〔3〕陈其荣. 诺贝尔自然科学奖与跨学科研究[J]. 上海大学学报(社会科学版)，2009(5)：48-62.

〔4〕全国哲学社会科学办公室. 跨学科研究系列调查报告选登之十一跨学科研究：辨析及跨学科研究项目的界定—评审—管理[R/OL]. [2016-02-07]. http://www.npopss-cn.gov.cn/GB/220182/227704/15319170.html.

〔5〕全国哲学社会科学办公室. 跨学科研究系列调查报告选登之一跨学科研究：理论与实践的发展[R/OL]. [2016-02-07]. http://www.npopss-cn.gov.cn/GB/220182/227704/15318717.html.

学及细分的认知科学的知识进行跨学科研究。

本书内容期望能够提高图书馆提供高效知识服务的能力。知识管理和知识经济的兴起，引发了知识社会化和社会知识化的趋势。计算机和网络技术的广泛应用，克服了时间、地域、机构之间的隔离，使科研人员、信息资源、科学仪器设备和计算工具等紧密联系在一起，营造了 e-science 协同科研环境和 e-learning 的协同学习环境。在新的环境下，用户的信息需求更加多样化、知识化，要求也更高了。面对浩如烟海的各类信息，用户并不需要图书馆只提供一个文献线索这种简单的服务，或者提供大量相关性不强的信息列表。当前，由于缺乏对信息资源的深入的知识组织和规范控制，虽然身处“信息海洋”，却面临“信息泛滥，知识匮乏”的困境，图书馆等传统信息机构的信息服务工作面临着巨大挑战。

正是这种难题与挑战，才使得快速存取、有序组织、深度挖掘和有效利用数字化、网络化信息资源，成为图书馆适应知识经济时代的必然要求，基于语义的文本挖掘、信息组织和信息检索等新课题应运而生，越来越多的研究使得该领域在近几年成为图书情报领域的研究热点并得到长足发展。但大多研究都停留在模型、框架和体系的设计等层面上，很少在技术层面或微观层面解决图书馆个性化服务或书目检索技术等问题。

本书内容期望能够促进广泛多元的语义实践活动。语义网从提出就引起了全球的广泛关注，学者、企业都进行了众多的研究和构建尝试，哪怕只是简单地构建了某个本体，也是进行了语义化活动[1][2]。人们在语义化活动中积累了大量丰富和宝贵的经验，需要加以总结、提炼为语义化理论。没有系统、完善的语义化理论指导的语义化实践活动，是一种低层次的重复劳动，无法实现突破、创新和发展。如果语义化理论研究滞后于语义化的实践活动，则会制约实践活动的发展，所以要对语义化进行突破和创新，寻找语义化的理论基础和理论来源，为语义化实践活动提供理论支撑，以更快地实现语义网。

本书内容期望能够丰富语义化理论的形成。语义化理论研究分散，不成体系，是当前语义网技术研究的普遍特点，这致使语义化实践活动受到了来自各方面的质疑，才使得在蒂姆·伯纳斯·李(Tim Berners Lee)提出语义网后而昙花一现。理论的缺乏也使得实证和应用难以有效拓展。语义化理论的自身发展和完善，已经成为语义网研究中一个不可回避的关键问题。源自各个方面、各种类型和各种层次的语义化活动的日渐丰富，而至今只有学者研究了语义网的理论基础[3][4]，并没有学者将语义化理论形成体系，更没有形成针对图书馆资源的语义化理论体系，所以迫切需要从理论提升的高度对其加以概括、总结、凝练和上升，形成完善的语义化理论。只有实现了由语义化实践活动到语义化理论这一质的飞跃，才能真正完成全球人民对语义网的渴望和语义网的最终实现。

本书内容期望能够从多维角度提升语义化技术。语义网的提出至今已过去十几年，全世界憧憬着语义网环境下的生活，众多学者将身边的信息资源发布成语义信息，使之成为语义网的一部分。信息资源语义化的形式有很多种，凡是将人们掌握的知识通过先进技术转化成机器能够理解的语言，都可认为信息被语义化了，所以发布语义信息的途径不仅仅是构

---

〔1〕邱均平，楼雯. 基于 CSSCI 的情报学资源本体构建研究[J]. 情报资料工作，2013(3)：57－63.

〔2〕邱均平，楼雯，余凡，等. 基于资源本体的馆藏资源语义化研究[J]. 图书馆论坛，2013(6)：1－7.

〔3〕姚绍文，邵剑飞，余江. 语义 Web 的技术基础与理论基础[C]//中国科学技术协会. 第六届全国计算机应用联合学术会议论文集，2002：7.

〔4〕邱均平，余凡. 基于计量分析的馆藏资源语义化理论研究[J]. 中国图书馆学报，2012(4)：71－78.

建成本体或关联数据。但本体和关联数据是目前学者们首肯的语义化方式，近年来，世界著名机构如 BBC[1]、路透社[2]、维基百科[3]、美国国会图书馆[4]、中国国家科技图书文献中心[5]等，纷纷将其资源语义化，在互联网上发布和提供查询服务。2007 年，W3C 设计了全民关联数据的计划，最大限度地接近了语义网。

信息资源语义化已经成为知识交流和知识共享的必经之路，图书馆作为蕴含巨大信息资源和知识的集合，图书馆资源的语义化在世界一些地区已经成为语义网建设之路的重要组成部分，在另一部分地区也即将成为重点研究的对象。语义网何以经过十几年还未能实现，不仅仅是浩瀚的信息海洋造成的，也是因为语义化过程中会遇到的种种逻辑难题和技术难题。语义网的实现是一个层层推进的过程，首先将一部分易于语义化的现有资源语义化，可以带动语义化，而图书馆就是现成的实验对象。在中国，图书馆的数字新形象虽已被人们接受，但多数用户仍然以单纯的书目信息检索为目的来使用图书馆，图书馆资源的语义化进程将在用户层面遇到很大挑战。同时在技术层面，图书馆资源语义化是否有标准可依，是否有关键技术可循，都值得思考。

本书在总结前人研究的理论、方法和应用的基础上，以图书馆的馆藏书目资源为研究对象，综合运用情报学、计算机科学、图书馆学、认知科学等多学科的知识、方法和工具，明确语义图书馆的概念，提出了一套图书馆资源语义知识检索方案，并为各类图书馆提供示范性的馆藏资源语义化解决方案和语义检索系统，全文贯穿以用户需求为目标、以用户认知为导向，为图书馆用户提供知识服务和未来图书馆的发展进行了良好的探索，提供了有益的帮助。本书的主要内容包括以下 6 个方面：

第一方面是分析认知科学视角下用户的图书馆知识检索行为和用户认知与图书馆知识检索的关联。本书将利用认知行为的方法分析图书馆用户认知的方式和过程、图书馆用户认知的影响因素以及图书馆用户知识需求的规律，对比传统的以系统为导向的检索行为，从认知心理学的角度分析以用户认知为导向的检索过程中，检索设计对用户认知的影响，以及用户认知对检索设计的反作用。这一部分内容将在第 2 章中阐述。

第二方面是定义语义图书馆的概念和认知科学视角下的内涵。图书馆资源是本书的研究对象，语义图书馆是本书的研究目标，因此，本书将利用多伦多大学开发的 iSquare 协议从认知科学的角度审视语义图书馆的意义和内涵，从而对语义图书馆的概念进行界定，提出语义图书馆的特点和作用，厘清电子图书馆、虚拟图书馆、数字图书馆、移动图书馆、智能图书馆与语义图书馆的界线。这一部分内容将在第 3 章中阐述。

第三方面是设计语义图书馆信息资源的构建方式和服务方式。信息资源的组织和构建是信息检索的基础和依靠，因此，本书将首先对语义图书馆的存在形式进行界定，以便确定语义图书馆信息资源构建的方向，进而利用系统动力学方法分析了语义图书馆信息资源配置的关键因素，本书将阐述语义图书馆的信息资源组织方式，主要是与传统图书馆和数字图

---

〔1〕杨爱武. 基于关联数据的图书馆创新服务研究[J]. 图书与情报，2012(3)：85－88.

〔2〕新浪科技. 路透社发布 Calais 网络服务开放式 API[EB/OL]. [2013－04－29]. http://tech.sina.com.cn/i/2008－01－31/14382008679.shtml.

〔3〕张海粟，马大明，邓智龙. 基于维基百科的语义知识库及其构建方法研究[J]. 计算机应用研究，2011(8)：2807－2811.

〔4〕夏翠娟，刘炜，赵亮，等. 关联数据发布技术及其实现——以 Drupal 为例[J]. 中国图书馆学报，2012(2)：49－57.

〔5〕乔晓东，白海燕，梁冰. NSTL 的关联数据构建与应用场景设想[J]. 数字图书馆论坛，2012(2)：54－60.

书馆信息资源构成、组织过程的区别和特点，最后提出语义图书馆的重点知识服务方式。这一部分内容将在第 3 章中阐述。

第四方面是设计语义图书馆知识检索模型。在总结语义图书馆的理论基础和分析语义图书馆知识检索整体环境的基础上，依据理论基础中的相关知识和整体环境包括的各因素，本书将设计一套语义图书馆知识检索模型，并将对模型的内容和各层运作方式进行详细阐述，在设计时，每一层的运作方式都将结合第 2 章中用户认知和知识检索的相互影响的规律进行设计。模型将作为语义图书馆知识检索构建的通用模型，为图书馆知识服务提供参考。这一部分内容将在第 4 章中阐述。

第五方面是设计语义图书馆知识检索系统。为了验证本书设计的面向三种结构数据的语义图书馆知识关联模型、方法和技术体系的可用性和适用性，本书设计了语义图书馆知识检索系统，利用系统工程方法、系统开发工具、本体构建工具、可视化展示等方法，构建语义知识库，为用户提供知识检索服务，从而实现图书馆的知识服务。

第六方面是以武汉大学图书馆和 Powell's City of Books 为实验用户，实现语义图书馆知识检索系统。依托典型的高校图书馆——武汉大学图书馆，以及一种新型图书馆——Powell's City of Books 的馆藏资源，将本书设计的语义图书馆知识检索系统应用于两所图书馆，提供知识检索服务，测试并验证系统的可用性。系统的设计和实现两个部分内容将集中体现在第 5、6 章。

第1章 图书馆资源语义化概述

1.1 语义网与图书馆资源语义化

图书馆资源语义化的研究对象是图书馆资源，实现途径是语义化。图书馆资源是相对固态的，语义化理论和方法是随着技术革新而不断变化。确定了研究对象，那么研究语义化理论和方法才是重中之重，因此经过对大量文献的阅读和研习，本书将从整体到局部，从理论到技术来分析语义化的理论与方法的研究现状，包括了语义网理论、语义化理论的应用、语义网的关键技术及细分到数字图书馆的关键技术、语义化标准研究五个方面。

1. 语义网理论研究

早在20世纪70年代末，俄罗斯学者索科洛娃(Stokolova)发表了一系列论文，较为全面地介绍了基于语义理论的信息检索，讲述了相关概念、情报语言[1]、句法工具[2]、情报语言对信息检索的作用[3]。而后，琼斯(Jones)在1981年分析了现代语义学中的语义理论在信息检索中的作用[4]，尽管离语义网的提出还有近20年的时间，但他们两个是最早从语言学的角度分析了信息检索理论研究，在之后只有杨志峰[5]等人跟随他们的脚步从语义基础分析信息检索的可行性。在语义网提出的10年内，不少学者开始总结语义网的理论基础[6][7]，概括的结果基本是：语义网的理论基础以W3C制订的标准为核心[8][9]，由资源描述框架[10]、XML[11]、本体[12]组成，还有一些学者分析了语义链接网络的理论基

[1] STOKOLOVA N A. Elements of A Semantic Theory of Information-Retrieval. 1. Concepts of Relevance And Information Language[J]. Information Processing & Management, 1977, 13(4): 227-234.

[2] STOKOLOVA N A. Elements of A Semantic Theory of Information-Retrieval. 3. Paradigmatic Relations[J]. International Classification, 1977, 4(1): 11-19.

[3] STOKOLOVA N A. Elements of A Semantic Theory of Information-Retrieval. 2. Syntactic Tools And Semantic Power of Information Languages[J]. International Classification, 1976, 3(2): 75-81.

[4] JONES K P. Semantic Theory—Towards A Modern Semantics[J]. Journal of Documentation, 1981, 37(4): 225-226.

[5] 杨志峰，王斌，李素建. 信息检索相关性理论的语义基础分析[J]. 计算机科学，2004(3): 1-4.

[6] GRANITZER M, LINDSTAEDT S. Semantic Web: Theory And Applications[J]. Journal of Universal Computer Science, 2011, 17(7): 981-982.

[7] POOLE D, SMYTH C, SHARMA R. Semantic Science: Ontologies, Data And Probabilistic Theories[C]// Proceedings of 6th International Semantic Web Conference/2nd Asian Semantic Web Conference (ISWC 2007/ASWC 2007). Berlin: Springer-Verlag Berlin, 2008: 26-40.

[8] 周静怡，黄国彬. 2007—2008年国外语义网研究与应用进展[J]. 图书馆建设，2009(1): 19-23.

[9] 朱成兵. 语义网理论研究[J]. 赤峰学院学报(自然科学版)，2010(4): 18-20.

[10] 姚绍文，邵剑飞，余江. 语义Web的技术基础与理论基础[C]//中国科学技术协会. 第六届全国计算机应用联合学术会议论文集，2002: 7.

[11] 罗庆云，赵巾帼. 语义化Web的理论基础与技术基础[J]. 甘肃联合大学学报(自然科学版)，2007(5): 75-79.

[12] 杨倩. 面向语义Web的本体理论和工程方法研究[D]. 天津：天津大学，2012.

础[1]、模型[2][3]、应用[4]、趋势[5]，但明显发现依据 W3C 制订的这些技术标准作为理论基础是不全面的。另外，还有学者将语义化过程中使用的知识进行总结，比如潜在语义分析[6][7]、语义标注[8]，强调他们是语义化的重要理论，这是很有益的探索，但同样这些都是技术方法提升为理论的探索。

2. 各学科理论在语义化的应用

从上面的分析中，可以发现有关语义化理论自身的研究和语义化理论的应用研究不够全面或是不够多，但有关来自其他学科的理论在语义化的应用研究却不少，说明语义化过程是一个多学科交叉融合的过程。应用得较多的是数学，数学作为基础学科，众多应用科学都需要借助它的知识才能完成逻辑层面的分析，有学者利用模糊理论[9][10]、概念格理论[11]、图论[12]、贝叶斯理论[13]等完成语义分词、语义相似度计算、语义聚类等语义化过程。语言学的理论也是应用的理论，有学者分别利用认知理论[14]、命题逻辑[15]、范畴理论[16]、修辞理论[17]等设计语义检索模型中的初始部分。另外，信息科学的理论也是重要的语义化应用理论，比如 IF 理论[18]、HNC 理论[19][20]。总结发现，无论是数学、语言学的理论，还是信息科学的理论，它们在语义化过程中扮演的角色都是语义化最初语义处理阶段的任务，可以说它们是奠基的理论，也可以说是非核心理论。

---

[1] YE L, CHEN J L. Automatic Composition of Semantic Web Services—A Theorem Proof Approach [C]// Proceedings of 1st Asian Semantic Web Conference. Berlin: Springer-Verlag Berlin, 2006: 481 - 487.

[2] 诸葛海. 语义网格的基础理论、模型与方法研究进展[J]. 中国基础科学, 2007(6): 27 - 29.

[3] ZHUGE H, SUN Y C. The Schema Theory For Semantic Link Network[J]. Future Generation Computer Systems-The International Journal of Grid Computing-Theory Methods And Applications, 2010, 26(3): 408 - 420.

[4] JOO J. Adoption of Semantic Web From The Perspective of Technology Innovation: A Grounded Theory Approach [J]. International Journal of Human-Computer Studies, 2011, 69(3): 139 - 154.

[5] SUN Y C, BIE R F, YU X F, et al. Semantic Link Networks: Theory, Applications, And Future Trends[J]. Journal of Internet Technology, 2013, 14(3): 365 - 377.

[6] 盖杰, 王怡, 武港山. 潜在语义分析理论及其应用[J]. 计算机应用研究, 2004(3): 9 - 12, 20.

[7] 李华云. 潜在语义分析的理论研究及应用[J]. 现代情报, 2006(11): 205 - 206.

[8] 宋彦. 视频语义标注方法和理论的研究[D]. 合肥: 中国科学技术大学, 2006.

[9] KHOURY R, KARRAY F, BASIR O. Semantic Context Classification By Means of Fuzzy Set Theory [C] //Proceedings of 2005 IEEE International Conference On Natural Language Processing And Knowledge Engineering. 中国应用技术发展中心, 北京邮电大学, 武汉科技大学, 2005: 6.

[10] 李祯, 杨放春, 苏森. 基于模糊多属性决策理论的语义 Web 服务组合算法[J]. 软件学报, 2009(3): 583 - 596.

[11] 张小红. 基于概念格理论的语义相似度模型研究及验证[J]. 郑州大学学报(工学版), 2011(5): 80 - 83.

[12] 黎英. 基于图论的语义 Web 服务聚类方法[J]. 计算机工程, 2011(22): 51 - 52, 55.

[13] ZHENG X Q, CHEN H J, WU Z H, et al. A Computational Trust Model For Semantic Web Based On Bayesian Decision Theory[C]//Proceedings of 8th Asia-Pacific Web Conference And Workshops (APWEB 2006). Berlin: Springer-Verlag Berlin, 2006: 745 - 750.

[14] 李海军, 侯建军. 认知理论在语义网支持下的现代远程教育中的运用探索[J]. 教育理论与实践, 2008(21): 59 - 61.

[15] PAN Xiaodong, XU Yang. Semantic Theory of Finite Lattice-Valued Propositional Logic [J]. Science China (Information Sciences), 2010, 10: 2022 - 2031.

[16] 颜丽. 基于范畴论的应急预案语义模型研究[D]. 南京: 南京邮电大学, 2011.

[17] HAOUAM K, MARIR F. SEMIR: Semantic indexing and retrieving Web document using Rhetorical Structure Theory[C]//Proceedings of 4th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2003). Berlin: Springer-Verlag Berlin, 2003: 596 - 604.

[18] 鲍泓, 封军康, 刘宏哲. 基于 if 理论的分布式系统语义互操作研究[J]. 计算机科学, 2008(3): 261 - 263, 273.

[19] 吴佐衍, 王宇. 基于 hnc 理论的词语相似度计算[J]. 中文信息学报, 2014(2): 37 - 43, 50.

[20] ZHANG Quan, WU Chen, WEI Xiangfeng. The Analysis of Chinese Sentence Semantic Chunk Share Based On Hnc Theory[C]//Proceedings of 20th Language, Informaton And Informatics International Conference of Asia-Pacific Area. National Natural Science Foundation of China, 2006: 8.

3. 语义网的关键技术

实现语义网的技术是连接图书馆资源语义化与万维语义网的关键，目前的研究多以总结语义网技术和提出新型语义化技术为主。在本体研究的热潮中，相关学者已将语义网的关键技术默认为本体及其相关技术，这一类的研究包括了全面介绍语义网信息组织的技术和方法，并总结出 RDF、OWL 和本体是语义网的核心技术[1][2]，本书一再强调广义的语义化，因此这些总结出的关键技术并不能代表所有的语义网技术。在新型技术的研究上，有学者提出了语义化网络的学习算法[3]、知识的自动分类技术[4]、微格式技术[5]可以作为语义网实现的关键技术，但这些技术的使用环境较为局限，研究也缺乏全面性。另外一方面研究主要集中在语义相似度计算标准上，较少研究的是语义相似度计算标准，其中大部分学者进行的是基于本体的语义相似度计算标准研究，比如设计语义相似度计算标准用于评价基于本体的检索质量[6]、覆盖度[7]、分类关系[8]等。有学者将设计了引文标准评价语义相似度的框架[9]，还有学者将语义相似度的计算结果用于评价主题图的相关性[10]、网络服务匹配质量[11]。当然也有学者领悟到总结归纳语义网关键技术的必要性[12]，但只描述了问题，并未解决问题。

4. 数字图书馆的关键技术

数字图书馆在图书馆资源从数字化到语义化的过程中起着重要作用，数字图书馆相关技术的研究包括了对语义化技术的应用以及微观、中观层面技术的研究。在技术的应用方面，目前学者偏向于利用 RDF[13]、元数据[14]、本体[15]和关联数据[16]进行图书书目的语义化或提出新的知识组织方法，也就是说这些学者将 RDF、元数据、本体和关联数据视为数字图书馆实践中的关键技术。另外，微观层面的技术包括了概念提取[17][18]、概念转换[19]、互操作[20]、

---

〔1〕戴维民. 语义网信息组织技术与方法[M]. 上海：学林出版社，2008：46.
〔2〕李青山，陈平. 语义化互联网的关键技术[J]. 计算机科学，2002(6)：86－89.
〔3〕姚绍文. 语义化 Web 的关键技术及其应用研究[D]. 成都：电子科技大学，2002.
〔4〕代印唐. 基于语义网络的知识协作关键技术研究[D]. 上海：复旦大学，2009.
〔5〕厉毅，郑炜. 数字学习网站资源的微格式语义化组织[J]. 中国教育信息化，2012(17)：30－33.
〔6〕陆宝益，李保珍. 基于本体的检索质量的语义相关度评价[J]. 情报杂志，2006(10)：63－65.
〔7〕欧阳柳波. 领域本体覆盖度评价关键技术研究[D]. 长沙：中南大学，2012.
〔8〕吴芳. 基于语义相似度的本体中分类关系的评价研究与实现[D]. 上海：华东师范大学，2010.
〔9〕PAKHOMOV S V S, PEDERSEN T, MCINNES B, et al. Towards a framework for developing semantic relatedness reference standards[J]. Journal of Biomedical Informatics, 2011, 44(2): 251－265.
〔10〕李丽冬. 主题图的语义相关度评价方法研究[D]. 大连：大连理工大学，2008.
〔11〕王玉影. 基于语义相关度评价的 Web 服务匹配技术研究[D]. 大连：大连理工大学，2006.
〔12〕罗庆云，赵巾帼. 语义化 Web 的理论基础与技术基础[J]. 甘肃联合大学学报(自然科学版)，2007(5)：75－79.
〔13〕朱大丽. 图书馆目录数据关联的语义化探析——充溢着背景知识的图书馆目录数据[J]. 图书馆学研究，2012(1)：54－58，95.
〔14〕白海燕，乔晓东. 基于本体和关联数据的书目组织语义化研究[J]. 现代图书情报技术，2010(9)：18－27.
〔15〕欧石燕. 面向关联数据的语义数字图书馆资源描述与组织框架设计与实现[J]. 中国图书馆学报，2012(6)：58－71.
〔16〕王军，卜书庆. 网络环境下知识组织规范的研究与设计[J]. 中国图书馆学报，2012(4)：39－45.
〔17〕王睿佳，刘耀. 面向科技文献的多模态语义关联特征提取与表达体系研究[J]. 大学图书馆学报，2012(5)：71－76.
〔18〕董慧，余传明，姜赢，等. 基于本体的数字图书馆检索模型研究(Ⅱ)——语义信息的提取[J]. 情报学报，2006(4)：451－461.
〔19〕王丽华. 基于语义网的数字图书馆的关键技术[J]. 情报杂志，2004(4)：5－8.
〔20〕刘炜. 基于本体的数字图书馆语义互操作[D]. 上海：复旦大学，2006.

语义互联[1][2]、概念格[3][4]，中观层面的技术包括了语义网格、SOA[5]、本体构建[6]、本体映射[7]、本体进化[8]，可以看到，这些研究重点描述了数字图书馆语义化的某种技术，并没有形成一套完整的流程和技术体系。

1.2 图书馆资源的语义检索

图书馆不再是图书和资源的管理者，而是知识的提供者，这一角色的变换，使得图书馆担起了知识服务的重任，图书情报工作者的任务就是将图书知识组织，尽可能地将知识关联起来，完成知识检索，将检索结果推荐给用户。因此，从理论、技术、应用三个方面看这一部分的研究，主要集中在语义化在知识组织的应用、知识服务的关键技术、语义检索，其中由于信息检索服务是图书馆信息服务的重点内容，语义检索则是这一部分研究的重中之重。

1. 语义化理论在知识组织的应用

语义化理论能够提供给知识组织、图书馆资源组织才是最终目的，是实践上升为理论，再应用于实践的循环过程，也是语义化理论的核心内容。可是，语义化技术应用于知识组织的研究非常多，但将语义化技术提升到语义化理论，再提供应用的研究则不多。比如应用于图书馆资源建设的语义网格理论研究[9]、语义互联研究[10][11]、计量分析的语义化理论[12]、语义信息理论[13]研究，只有这些是对语义化理论应用的大胆尝试。还有更多的探索是应用于信息检索，比如语义检索系统的实现[14]、检索需求的确定[15]、检索模型的设计[16][17]等。另外还有学者针对语义化的本体层的探索，比如设计探测相似度度

---

[1] 牟冬梅. 数字图书馆知识组织语义互联策略及其应用研究[D]. 长春：吉林大学，2009.

[2] 董慧，余传明，徐国虎，等. 基于本体的数字图书馆检索模型研究(Ⅳ)——历史领域知识推理机制[J]. 情报学报，2006(6)：666-678.

[3] 韩毅. 语义网格环境下数字图书馆知识组织策略与应用研究[D]. 长春：吉林大学，2008.

[4] 滕广青. 基于概念格的数字图书馆知识组织研究[D]. 长春：吉林大学，2012.

[5] 刘成山，刘怀亮. 基于语义网的数字图书馆[J]. 情报杂志，2008(1)：49-54.

[6] 董慧，余传明，杨宁，等. 基于本体的数字图书馆检索模型研究(Ⅲ)——历史领域资源本体构建[J]. 情报学报，2006(5)：564-574.

[7] 董慧，杨宁，余传明，等. 基于本体的数字图书馆检索模型研究(Ⅰ)——体系结构解析[J]. 情报学报，2006(3)：269-275.

[8] 贾保先，鲍素贞，杨吉宏. 虚拟数字图书馆语义平台建设关键技术研究[J]. 聊城大学学报(自然科学版)，2009(4)：93-96.

[9] 毕强，牟冬梅. 语义网格环境下数字图书馆知识组织理论、方法及其过程研究[J]. 图书情报工作，2007(8)：6-9，20.

[10] 古新生，陈清. 面向对象的语义关联数据模型理论[J]. 软件学报，1993(5)：24-37.

[11] 沈秀丽，牟冬梅. 数字图书馆知识组织语义互联论纲[J]. 情报科学，2010(3)：379-383，429.

[12] 邱均平，余凡. 基于计量分析的馆藏资源语义化理论研究[J]. 中国图书馆学报，2012(4)：71-78.

[13] 陈明先. 语义情报理论及其发展[J]. 情报理论与实践，1987(6)：9-11.

[14] 余传明. 基于本体的语义信息系统研究[D]. 武汉：武汉大学，2005：91-97.

[15] WANG S F，FENG J K. Identifying And Formulating Information Requirements Based On Semantic Theories of Information[C]//Proceedings of 6th International Conference On Machine Learning And Cybernetics. New York：IEEE，2007：4080-4086.

[16] YUE K，LIU W Y. Semantic Field：A Theoretical Perspective of Modeling Information Retrieval[J]. International Journal On Artificial Intelligence Tools，2009，18(6)：825-851.

[17] MAGALHAES J，RUGER S. An Information-Theoretic Framework For Semantic-Multimedia Retrieval[J]. ACM Transactions On Information Systems，2010，28(4)：19.

量方法[1]、桥本体[2]的构建等。

2. 知识服务的关键技术

语义网和数字图书馆的建设和实现实际上都是为了知识交流和知识共享，因此上文提到的许多研究已经表现出知识组织或个性化服务的内容和关键技术。不仅如此，有文献[3][4][5]总结了作为聚类技术、数据挖掘技术在图书馆建设中的具体方法；有文献[6]认为手工决策技术、基于内容的推荐系统、基于本体的服务系统和智能信息推拉技术是个性化服务的技术支持；也有文献[7]提出基于读者行为的知识服务的关键技术有读者特征提取技术、兴趣模型分析技术和协同推荐技术。这些是对知识服务技术特征的总结和探讨。还有文献[8]利用关联数据将多种数据源的知识关联到一起形成语义扩展，则是对关键技术的应用。

近年来，由于图书馆资源的语义化进程加快，一些学者提出了有建设性的语义化模型和框架，为今后图书馆资源语义化和知识服务提供了参考[9][10][11]。上述研究有的仅设计了模型或进行实验验证，有的则仅描述一部分技术，尚缺乏对图书馆资源语义化过程整套系统的关键技术的归纳总结。

3. 语义检索

早在20世纪80年代对语义检索的讨论就出现在SIGIR会议论文中，但语义检索研究始终受制于语义信息处理发展水平的局限。随着自然语言处理、人工智能的发展，尤其是语义网技术的兴起与发展，语义检索研究自20世纪末以来得以迅速发展[12]。语义信息检索就是要让用户在输入自然语言作为检索词的时候，能出现与该检索词相关的更多词，而不是机械地将与该检索词匹配到的所有信息一举列入检索结果。目前国内外语义检索研究主要集中在以下几个方面。

第一个方面是基于本体的查询技术，查询技术首先涉及查询语言，由W3C推出的RDF、SPARQL等系列查询语言已经可以实现对语义数据的查询，应用广泛。如余传明[13]阐述并比较了三种基于查询语言的检索机制。国外的研究一般集中在利用语言本体（如WordNet）中的同位词、上下位词以及上下文检索技术对所要查询的内容进行语义消歧并进

---

[1] 王凯. 面向医学领域的概念语义本体相似度度量理论与方法研究[J]. 江汉大学学报（自然科学版），2014(2)：37-40.

[2] XU B W, WANG P, LU J J, et al. Theory and Semantic Refinement of Bridge Ontology Based On Multi-Ontologies [C]//Proceedings of 16th IEEE International Conference on Tools With Artificial Intelligence. Los Alamitos: IEEE Computer Soc, 2004: 442-449.

[3] 潘伟. 个性化信息服务的关键技术——聚类分析[J]. 现代情报，2007(10)：212-214.

[4] 李静. 数据挖掘技术在高校图书馆个性化服务中的应用研究[D]. 天津：天津大学，2012.

[5] 赵红霞. 数据挖掘技术和RSS技术在图书馆个性化服务中的应用[D]. 郑州：解放军信息工程大学，2008.

[6] 周庆. 图书馆个性化信息服务的技术支持[J]. 大学图书情报学刊，2008(6)：60-64.

[7] 张炜，洪霞. 基于OPAC读者行为挖掘的个性化服务系统关键技术分析[J]. 图书馆论坛，2010(1)：62-64.

[8] 王思丽，祝忠明. 利用关联数据实现机构知识库的语义扩展研究[J]. 现代图书情报技术，2011(11)：17-23.

[9] 贺德方，曾建勋. 基于语义的馆藏资源深度聚合研究[J]. 中国图书馆学报，2012(4)：79-87.

[10] 邱均平，余凡. 基于计量分析的馆藏资源语义化理论研究[J]. 中国图书馆学报，2012(4)：71-78.

[11] 邱均平，楼雯. 基于共现分析的语义信息检索研究[J]. 中国图书馆学报，2012(6)：89-99.

[12] 黄敏，赖茂生. 语义检索研究综述[J]. 图书情报工作，2008(6)：63-66.

[13] 余传明. 基于本体的语义信息系统研究[D]. 武汉：武汉大学，2005：91-97.