



Multilingual Online Academic
Information Mining and Retrieval

多语言网络学术信息

挖掘与检索

吴 丹◎著



科学出版社

Multilingual Online Academic
Information Mining and Retrieval

多语言网络学术信息 挖掘与检索

吴 丹◎著

科学出版社

图书在版编目(CIP)数据

多语言网络学术信息挖掘与检索/吴丹著. —北京: 科学出版社, 2014. 7

ISBN 978-7-03-041020-7

I. ①多… II. ①吴… III. ①互联网络-情报检索 IV. ①G354.4 ②TP393.4

中国版本图书馆 CIP 数据核字 (2014) 第 125059 号

责任编辑: 邹 聪 王茜艳 / 责任校对: 赵桂芬

责任印制: 赵德静 / 封面设计: 无极书装

编辑部电话: 010-64035853

E-mail: houjunlin@mail.sciencep.com

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>

北京源海印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2014 年 7 月第 一 版 开本: 720×1000 1/16

2014 年 7 月第一次印刷 印张: 15 1/2

字数: 285 000

定价: 78.00 元

(如有印装质量问题, 我社负责调换)



序言

PREFACE

初识吴丹博士，是在 2009 年岁末的广州。

是年 12 月 19~21 日，中国科学技术协会第 207 次青年科学家论坛在中山大学举行，主题是“数字环境下情报服务的方法创新与支撑技术”。来自北京大学、武汉大学、中国科学技术信息研究所等近 30 所高校和科研院所的 60 多位青年才俊与会。我和马费成老师有幸受邀参会并担任顾问。开幕式后，研讨会移师清远继续进行。我参加了一个大组的讨论。青年学者们依次上台报告自己的研究进展和学术见解，然后由参会教授进行点评。年轻学者们各具纵深和特色的学术报告，显示了他们最新的研究成果和科学发现，尤其给我留下深刻印象的几位报告人中，就有武汉大学的吴丹博士和北京大学的王继民博士，此外，黄如花教授的精彩点评也令人难忘。那时我不认识吴丹博士，只是由于我先前在中国社会科学院语言研究所工作时曾经从事过机器翻译和自然语言处理研究，对她关于跨语言信息检索的报告印象很深，感到很有思想。现场有老师给我介绍了她的情况，说她确实很优秀。散会后我和她简单聊了几句，鼓励她继续深化研究，有条件时适当延伸研究，如题录翻译研究与实践，努力为国家经济社会发展中一些紧迫的实际问题的解决做出贡献，她表示了认同。在那以后的几年间，又断断续续有过几次见面的机会。2011 年，她的专著《跨语言信息检索：翻译优化理论与技术》出版，她寄来了一本，我抽空阅读了这本凝聚着她心血的著作，看到了她的学术成长，为她感到由衷的高兴。但是由于我没有她的联系方式（她邮寄书的地址看不清楚），加之工作头绪杂乱，较长一段时间中连一句道谢的话都无法表达，还是后来辗转打听到了她的手机号码，得以发短信致谢。说来也巧，后来吴丹博士入选武汉大学首批“珞珈青年学者”，再后来入选中央组织部等部委首批“青年拔尖人才支持计划”，我都有幸参与评审推荐工作，比较详细、系统地阅读了她的全部材料，进一步了解了她在跨语言检索研究方面的学术见解和研究进展。

令人欣慰的是，吴丹博士近年来的研究，没有采取“打一枪换一个地方”的做法，始终围绕着多语言信息检索及用户行为展开，她发表了一系列较高水

平的中英文论文，一步一个脚印地不断深入，逐渐成长为这个领域的一位学有专攻的优秀青年学者。

在数字化、网络化不断发展的今天，人们的日常生活、工作和学习受到互联网的极大影响，而且这种影响还会继续下去。通过互联网，人们获取信息来解答问题或是安排行程，人们建立社交网络来与家人和朋友保持联系，人们在网上传博文、做网页来表达自己的想法和才能。所有这些都渗透到人们日常生活的方方面面，在改变个体行为的同时也改变了人类社会。

学者作为一个特殊的群体，他们的研究行为也在随着互联网的影响而发生改变。许多科学研究从最初想法到最终成果通常要经过思想启发与问题探索，概念形成与数据采集，结果分析与意义解释，以及文章写作与成果发表等阶段。在互联网被广泛使用之前，许多学者都基于传统的科研平台来完成自己的学术研究，利用的学术信息主要是靠传统的、纸质的媒介记载。

但互联网逐渐改变了传统的学术环境。数字图书馆的兴起不但将学术信息数字化，而且使学者们利用学术信息的手段大为便捷。学者们利用现代计算机存储技术、信息检索技术、网络传输技术和多媒体技术，不但可以自己直接查询和获取学术信息，而且所获得的信息的数量和时效性远远超过从前。这为学者研究中的思想启发与问题探索，以及文章写作与成果发表等阶段提供了强有力的工具和平台。

同样的变化也发生在学者研究中的概念形成与数据采集及结果分析和意义解释阶段。网络通信的广泛应用使学者间的协同合作可以在互联网上虚拟完成，社会网络的蓬勃发展让学者同时担任数据和信息的生产者与消费者。这些为学者们建立研究团队，保持合作关系，以及与研究成果使用者间更深入的良性互动都起到了促进作用。

所有这些都表明整个科学研究体系结构在发生变化。基于互联网的学术研究与合作平台是发展的方向。在这方面国外发展很快，为此迫切需要国内的学者们在这方面加大研究力度。吴丹博士的这本对网络学术信息研究的专著很及时地弥补了国内在这方面研究的不足。

最近学术界和 IT 界对“大数据”的讨论更清晰地表明了基于互联网的学术研究与合作平台的重要性，从而使更多的学者更加深切地体会到多语言技术和网上学术信息在现在和将来的学术研究中的角色和作用。

由于多年关注多语言的教育和科学研究，切实地体会到多语言技术在当前基于互联网的学术研究与合作平台上的重要意义。对于中国的学者来说，汉语是母语，是中国学者间学术交流的自然选择，因而中文网络学术信息有其存在的基础和意义。但是在一个开放的社会里，我们也应该实事求是地认识其他语言尤其是英语在世界学术交流中的实际影响和地位。同时还要看到，世界上许

多高水平的学术信息，包括传统的和网络上的，都是使用的英语。因此，英语是钥匙，是中国学者打开世界学术信息宝库的重要工具。在这样的环境里，多语言技术就成为了桥梁，成为把世界各地其他语言学术信息引进中国的桥梁，也成为把中国学者的学术成就传播到世界各地的桥梁。

吴丹博士的新专著《多语言网络学术信息挖掘与检索》，正是她在这一大背景下进行认真思考和研究继续与深入。我关注她在多语言技术方面的研究有好几年了，很高兴看到她的这本新著即将与读者见面。该书把多语言技术与网络学术信息科学地加以结合，并在互联网的学术研究与合作平台的大环境下，讲述了她对多语言网络学术信息用户需求与行为的调研和分析，与我们分享了她对多语言网络学术信息标签规范及挖掘的研究发现，观察她对多语言网络学术信息翻译资源与测评体系的构建，使我们能够了解她确立的多语言网络学术信息检索系统的实现路径，以及她对多语言网络学术信息检索实验的操作。所有这些，无不显示出一个年轻学者的研究潜力和学术精进，反映出她多年来对多语言网络学术信息的全面而深入的思考和研究，以及这些思考和研究所带来的可喜成果。相信吴丹博士在已取得的一系列成功的基础上，将进一步深入并适当拓展研究，不断有新的思考、新的发现及新的研究成果。

无论是在理论上的探索与发现，还是在应用上的启示与借鉴，抑或是对相关前沿研究领域的认知与追踪，愿所有对多语言处理技术和网络学术信息感兴趣的读者和学人都能从该书中获益。

在《多语言网络学术信息挖掘与检索》即将出版之际，写下一点感想，既是致贺，也是期望。

是为序。

黄长著^①

二零一四年春于首都北京

^① 黄长著，中国社会科学院学部委员、研究员，国家社科基金图书馆·情报与文献学科规划评审组组长，中国社会科学情报学会理事长，中国社会科学院翻译系列正高级职称评审委员会主任。

前言

PREFACE

对网络学术信息进行准确的表示与描述,从而无缝地获取各种学术信息资源,是科研工作者和信息服务机构亟待解决的一个问题。目前,网络学术信息资源的特点表现为:数量庞大,数字图书馆等网络资源的建设使得大量文本资源转化成更宜传播的数字资源;来源丰富,学术信息不再仅存在于结构化的数据库中,更多的学者或普通用户通过网页、论坛、博客、百科、问答系统等多种方式贡献或获取有价值的学术信息;语种多样,随着世界范围学术交流活动的日益频繁,越来越多的科研工作者需要了解和获取多语种学术资源;信息表示由受控标记发展为自由标记,用户不再仅是网络信息资源的消费者,同时也是创造者、描述者、组织者;无缝智能获取,网络学术信息的发展趋势是用户超越时空限制、跨越语言障碍而无缝智能地获取学术信息,形成信息高度共享。

据统计,目前网上学术信息按语言分布的情况是:英语大约占70%,法语、德语分别占6%,而汉语仅占网上学术信息总量的3%左右。这个数据说明,仅对汉语学术信息的单一表达处理,网络发挥的学术效益只占网络效益的3%左右。要提高学术信息获取效率,实现无缝获取,必须进行多语言网络学术信息的挖掘与检索创新研究。

从国内外的研究情况可以看出,采用协同标注对多语言学术信息资源进行描述、分类、组织、检索的过程中存在着信息描述精确度不高、标签组织混乱、同一领域不同用户的表示结果差异较大等缺陷。因此,我们需要对网络学术信息的多语言标签进行规范控制。同时,还需充分考虑对已有的网络学术信息的标签进行挖掘和有效利用,形成高质量的领域翻译资源。信息表示的目的是为了信息获取,利用良好的多语言学术信息标签去提高多语言学术信息检索的效率,这种模式对于深度开发网络信息资源是十分重要的。

本书对多语言网络学术信息资源进行了以下四个方面的理论与实践研究。

(1) 多语言网络学术信息用户需求与行为研究。通过对两类网络学术信息资源——数字图书馆和新型网络学术信息资源,展开多语言用户需求与行为调查。结果显示:对于数字图书馆而言,用户对多语言学术信息有较强的需求,

不同类型用户对数字图书馆提供的多语言学术信息服务有不同程度的要求。对于新型网络学术信息资源而言,大学生普遍认为新型网络学术信息资源比较重要,且不同国家的用户对网络信息资源使用的偏好也不同,这些新型网络学术信息资源的质量有待提高。

(2) 多语言网络学术信息标签规范与挖掘研究。通过对能够提供学术信息的新型网站,包括图书标注系统、学术论文标注系统、学术博客等的社会标签(social bookmarking/social tagging)与主题词表的比较实验发现,尽管标签与词表的重合度并不高,且中文网站与英文网站的表示方法也存在异同,但是从新型网络学术资源网站中抽取出来的社会标签对多语言网络学术信息的检索具有一定的帮助作用,可以用来形成翻译资源。

(3) 多语言网络学术信息翻译与测评技术研究。为了支持多语言网络学术信息检索,一方面,我们以图书情报领域为例,构建了两个图书情报领域的语义词典:一是由网络学术信息资源作者提供或自动抽取术语形成的专业词典,二是由网络学术信息资源用户提供的社会标签所形成的标注词典。另一方面,我们构建了一套图书情报领域多语言学术信息检索测评体系。

(4) 多语言网络学术信息检索系统实现与实验研究。我们开发了一个多语言网络学术信息检索系统——Multilingual Information in LIS: Knowledge Translation, Evaluation and Access (简称 Milk-Tea),并进一步进行了跨语言信息检索实验。实验结果显示, Milk-Tea 能够支持真实用户的检索,并且我们所构建的两个专业领域翻译词典(专业词典和标注词典)能够在多语言学术信息检索中发挥作用。

全书共分为六章。

第一章网络学术信息概述,介绍网络学术交流方式,网络学术信息的特点、组织和获取。

第二章多语言网络学术信息用户需求与行为调查,进行了用户对数字图书馆和新型网络学术资源的需求和行为两项调查分析。

第三章多语言网络学术信息标签规范与挖掘,抽取了图书标注系统、学术论文标注系统、学术博客三类网络学术资源上的中英文标签,将其与主题词进行比较,并探讨了挖掘这些标签作为跨语言学术信息检索的翻译资源的可能。

第四章多语言网络学术信息翻译资源与测评体系构建,利用所抽取的标签构建了中英文标注词典,同时还构建了中英文专业词典,以及一套图书情报领域跨语言学术信息检索的检索主题、语料库、标准答案集等。

第五章多语言网络学术信息检索系统实现,构建了一个图书情报领域的中英文跨语言学术信息检索系统 Milk-Tea,该系统内嵌了三种翻译资源(标注词典、专业词典、通用词典)。

第六章多语言网络学术信息检索实验,在 Milk-Tea 上进行了自动信息检索和用户信息检索两项实验,主要是对比三种翻译资源组合模式的检索效果,同时也进行了 Milk-Tea 的性能测试,以及用户跨语言学术信息检索行为分析。

上述研究的意义在于有助于解决网络学术信息表示的规范化问题,有助于形成高质量的领域翻译资源,有助于提高多语言学术信息获取的效果,有助于不同语言学术信息的交流与共享。

(1) 有助于解决网络学术信息表示的规范化问题。在 Web 2.0 环境下,由大众参与的信息描述与组织方式存在自由、凌乱等问题。本书针对多语言网络学术信息,对用户的协同标记、作者的规范表示,以及从学术信息本身抽取的领域术语进行聚合研究,并形成多语言对照语义词典。这将为多语言网络学术信息的规范表示提供指导。

(2) 有助于形成高质量的领域翻译资源。当前大多数翻译资源是从大规模平行语料库中获得的,多为通用翻译资源,存在大量噪声,并缺乏对专业词汇翻译的覆盖。本书以图书情报领域为例,利用学术信息的规范性,结合用户、作者对信息的表示,从多语言网络学术资源中提取翻译信息。这样形成的领域翻译资源将具有很高的质量和对专业领域词汇准确翻译的覆盖,也使信息表示得到应用,同时还为基础资源的构建开拓了新的思路。

(3) 有助于提高多语言信息获取的效果。对多语言网络学术信息表示进行聚合研究所形成的翻译资源与一般翻译资源相比,具有更准确、更专指的优点。这一翻译资源应用在多语言学术信息检索中,能够辅助消除查询翻译歧义性,优化检索结果,获得使用通用翻译资源无法达到的效果。

(4) 有助于不同语言学术信息的交流与共享。本书产生的对不同语言的学术信息标签进行挖掘,并应用于跨语言学术信息检索的模式,将有助于世界范围内的不同语言学术信息之间的广泛交流与共享,有助于不同语言的学者更好地理解其他语种学术成果。

本书得到了国家社会科学基金和中国博士后科学基金的资助,在此深表感谢。

吴 丹

二零一四年二月于武汉

目录

CONTENTS

序言	i
前言	v
第一章 网络学术信息概述	1
第一节 网络学术交流方式	2
一、正式的网络学术信息交流方式	2
二、半正式的网络学术信息交流方式	3
三、非正式的网络学术信息交流方式	4
第二节 网络学术信息的特点	6
第三节 网络学术信息的组织	7
一、数据库	8
二、学术搜索引擎	8
三、数字图书馆	8
四、用户标注	9
第四节 未来网络学术信息的无缝获取——cyberscholarship	9
第五节 多语言网络学术信息研究述评	10
本章参考文献	13
第二章 多语言网络学术信息用户需求与行为调查	16
第一节 网络学术信息的用户研究	16
一、用户使用信息资源的类型	17
二、不同国家用户的信息行为	18
三、大学生的信息行为	18
四、团队的信息行为	19
第二节 数字图书馆用户的多语言信息需求与期望调查	20

一、调查目的	20
二、问卷设计	21
三、调查对象	22
四、问卷信度和效度测试	23
五、问卷结果统计分析	24
第三节 用户与新型网络学术信息资源的交互行为调查	34
一、调查目的	34
二、问卷设计	35
三、调查对象	36
四、问卷结果统计分析	37
第四节 用户对多语言网络学术信息的需求与行为分析	45
一、用户对多语言网络学术信息的需求	45
二、用户对多语言网络学术信息的行为	47
第五节 本章小结	48
本章参考文献	49
 第三章 多语言网络学术信息标签规范与挖掘	53
第一节 Web2.0 环境下的网络学术信息资源	53
一、图书标注系统	53
二、学术论文标注系统	55
三、学术博客	57
第二节 社会标签——网络学术信息表示的新方法	58
一、社会标签在网络学术信息表示中的作用	59
二、社会标注与主题标引的比较	61
第三节 中英文学术资源网站的社会标签规范性测评实验 ——以图书情报领域为例	63
一、图书标注的规范性测评	63
二、学术论文标注的规范性测评	76
三、学术博客标注的规范性测评	81
第四节 网络学术信息表示的规范控制方法	85
一、图书标注规范性控制	86
二、学术论文标注规范性控制	87
三、学术博客标注规范性控制	89
第五节 本章小结	90
本章参考文献	90

第四章 多语言网络学术信息翻译资源与测评体系构建	93
第一节 多语言网络学术信息检索的关键技术	93
一、跨语言信息检索技术	93
二、机器翻译技术	97
三、跨语言信息检索评价体系	98
第二节 翻译资源及其构建方法	100
一、双语词典	100
二、双语语料库	102
三、机器翻译系统	103
四、多语叙词表	104
五、多语本体	105
第三节 图书情报领域多语言学术信息的翻译资源构建	107
一、专业术语翻译的特点	107
二、图书情报领域中英文专业词典构建	108
三、图书情报领域中英文标注词典构建	112
四、词典构建的特点	113
第四节 图书情报领域多语言学术信息的检索测评体系构建	114
一、测试文档集建立	114
二、检索主题设计	117
三、标准答案集构建	122
四、该测评体系的特点	126
第五节 本章小结	126
本章参考文献	127
第五章 多语言网络学术信息检索系统实现	130
第一节 系统实现的流程与环境	130
一、系统实现的流程	131
二、系统实现的环境	132
第二节 系统相关资源与算法	132
一、系统语料	132
二、翻译资源	132
三、查询翻译	134
四、检索模块	134
五、检索结果翻译	136

第三节	系统主要功能及界面	137
一、	用户管理	137
二、	图书情报领域多语言学术信息检索测评	140
三、	图书情报领域跨语言学术信息检索	140
第四节	本章小结	145
	本章参考文献	145
第六章	多语言网络学术信息检索实验	146
第一节	自动检索实验设计	146
一、	实验目的	146
二、	评价指标	147
第二节	自动检索实验结果分析	148
一、	短查询实验结果	148
二、	长查询实验结果	151
第三节	用户检索实验设计	153
一、	实验目的	153
二、	检索主题选择	154
三、	系统界面及参数设置	154
四、	实验用户确定	155
五、	实验步骤	157
第四节	用户检索实验结果分析——系统检索性能	157
一、	平均 NDCG 值	158
二、	最佳 NDCG 值	160
三、	用户实验的 NDCG 值与自动实验的比较	163
第五节	用户检索实验结果分析——用户评价	166
一、	用户背景	167
二、	用户对每次检索的评价	168
三、	用户对 Milk-Tea 的综合评价	170
第六节	用户检索实验结果分析——用户行为	171
一、	针对每个检索主题的用户行为分析	171
二、	针对每个查询式的用户行为分析	174
第七节	实验结论	177
一、	自动检索实验结论	177
二、	用户检索实验结论	178
	本章参考文献	178

附 录	180
附录 1 数字图书馆用户的多语言信息需求与期望 调查问卷 (中文版)	180
附录 2 数字图书馆用户的多语言信息需求与期望 调查问卷 (英文版)	185
附录 3 用户与新型网络学术信息资源的交互行为 调查问卷 (中文版)	191
附录 4 用户与新型网络学术信息资源的交互行为 调查问卷 (英文版)	194
附录 5 用于“中英文学术资源网站的社会标签规范性测评实验” 的图书情报领域中英文检索词	198
附录 6 50 个图书情报领域检索主题的部分字段 (中文)	199
附录 7 50 个图书情报领域检索主题的部分字段 (英文)	207
附录 8 用于用户检索实验的 10 个图书情报领域检索主题的 全部字段 (中文)	218
附录 9 Milk-Tea 用户检索实验调查问卷	223
后 记	229

第一章

网络学术信息概述

随着网络社会化进程的加快,信息技术与互联网深入渗透到社会生活的各个领域。2014年1月,中国互联网络信息中心(CNNIC)发布的《第33次中国互联网络发展状况统计报告》^[1]显示,截至2013年12月底,我国网民规模达到6.18亿,其中手机网民规模达5亿,成为中国网民的重要组成部分,互联网普及率攀升至45.8%。即时通信、搜索引擎、网络新闻、博客、微博、社交网站、电子商务类应用、网络娱乐类应用是互联网主要的应用模式,电子邮件、论坛/BBS等传统互联网应用的使用率继续走低。网络用户的增加和网络活动的丰富对中国网民在学习、生活、工作、娱乐等方面提出了更多的要求。

学术交流是人类科学活动的重要组成部分,是推动科学发展的重要手段。有效的学术交流能使科研人员快速、高质量地交流学术信息,能促进科学研究活动的开展,从而促进知识的增长。计算机技术、网络技术对学术领域的影响,导致了学术交流模式的深刻变化,网络学术信息资源应运而生。网络学术信息资源是网络信息资源中进入到学科领域,并具有学术价值的那部分资源。越来越多的科学研究人员热衷于通过网络来交流学术信息,网络学术信息资源正逐步成为学术信息交流中的主流信息资源。

随着云时代的来临,大数据吸引了越来越多的关注。在大数据环境下,全球数字学术交流系统成为现实。数字图书馆、学术网站、学术博客、网络学术社区、在线互动知识构建等多种在线学术交流形式,丰富了学术交流的内容,

产生了大量的科学数据。而科研的继承性需要借鉴和使用已有的数据,由此带来了网络学术信息的组织、获取、使用等问题,受到越来越多的科研管理机构和信息服务机构的重视。

第一节 网络学术交流方式

网络学术发轫于 20 世纪 90 年代兴起的网络文学,21 世纪以后发展迅猛。它不仅改变了传统学术交流的方式,也改变了传统学术交流的生态。其作用表现为知识创造方式、研究方式的变化,信息获取方式的变化,研究成果的发布,评价、传播方式的变化,以及出版物形态与出版模式的变化等。网络学术突破了传统学术的局限性,拓展了学术信息交流的空间,丰富了学术研究的内容和形式,对学术思想和学术研究的重要性日益突出,越来越成为现代信息用户学术信息交流的重要选择。

网络环境下,学术信息交流向立体化、多层面发展,学术交流的“正式”与“非正式”界限逐步淡化,呈现出多种交流方式并存的局面。

一、正式的网络学术信息交流方式

正式的学术信息交流是指借助于公开发表的文献进行的学术信息交流过程。在纸质载体时代,公开发表的文献主要是指印刷型图书、学术期刊、科技报告等。网络环境下的正式交流过程可以被看做是纸质载体时代的正式交流在互联网上的延伸,主要分为网络出版物、网络上的数字化文献及开放存取期刊(open access journal, OAJ) 三种。

(一) 网络出版物

网络出版又称互联网出版,是指具有合法出版资格的出版机构,以互联网为载体和流通渠道,出版并销售数字出版物的行为。与传统出版相比,网络出版具有快速、便捷、低定价、低成本、无需仓储、无需运输等优势。在资源利用上,它不需要纸张、不需要油墨等,是一种纯粹的环保、绿色产品。这些优点给网络出版的发展开拓了更广阔的空间。网络出版物主要有互联网图书、互联网报纸、互联网杂志、互联网音像出版物等。

(二) 网络上的数字化文献

在印刷出版时代,出版商建立了以学术专著和学术期刊为基础的科学信息

交流系统,网络环境下,许多出版商将传统的印刷版文献数字化处理,主要表现为:电子期刊、电子图书、电子版工具书、标准、专利等。相比印刷版文献,这类经过数字化处理并放到互联网上的正式出版物更易于获取和使用。

(三) 开放存取期刊

开放存取期刊在 20 世纪 90 年代末兴起,它是互联网上的在线出版物,免费提供给用户使用,用户只需支付上网的费用,而不必支付其他费用。瑞典隆德大学的开放存取期刊名录(directory of open access journals, DOAJ)对开放存取期刊所下的定义是:采用集资的模式,不向作者或其机构收取使用费的期刊。开放存取期刊具有投稿方便、出版快捷、出版费用低廉、便于传送或刊载大量的数据信息、检索方便、具广泛的读者群和显示度等优势。开放存取期刊逐渐得到传统文摘索引服务商的认可并成为他们收录的对象,同时也是一种正式的网络学术信息交流方式。

二、半正式的网络学术信息交流方式

网络环境下学术信息的半正式交流方式与正式交流方式的主要区别在于:正式交流方式中的学术信息是经过出版商、文摘索引商等专职信息人员进行评审、编辑、加工整理后才正式出版发行的,而半正式交流方式中的学术信息可能有经过加工处理,但是未正式出版。半正式的网络学术信息交流方式主要有三种,即学术机构网站、学科专业网站及开放存取仓储(open access repository, OAR)。

(一) 学术机构网站

从各种学术团体和科研院所、国际组织和政府部门、行业协会等组织机构的网站上可以获得政府白皮书、学科最新研究信息、学术动态信息等,还可以获得非正式出版物等“灰色文献”。这类文献主要包括:不公开刊载的会议文献,非公开出版的政府文献、学位论文,不公开发行的科技报告、技术档案、工作文件、内部刊物、未刊稿等。在互联网上,“灰色文献”的获取要比在传统方式下更容易些,而且获取费用也相对较低。

(二) 学科专业网站

学科专业网站包括提供各类科学信息的大型学科综合网站,集中了诸如研究动态、专家学者、专业论坛、重要文献、组织机构、学术期刊、统计数据、法律法规、学术会议、专业基本知识等各类信息,如中国经济信息网、中国法