

Construction and Application of the Institution's
Knowledge Map

机构知识图谱的构建 及其应用

魏瑞斌◎著



科学出版社

教育部人文社科基金项目“学术机构知识图谱构建及其应用研究”

(11YJC870024) 成果之一

安徽财经大学“企业信息管理与数据挖掘”学科特区专项资助

Construction and Application of the Institution's
Knowledge Map

机构知识图谱的构建 及其应用

科学出版社

北京

图书在版编目(CIP)数据

机构知识图谱的构建及其应用/魏瑞斌著. —北京: 科学出版社, 2015. 4

ISBN 978-7-03-044135-5

I. ①机… II. ①魏… III. ①引文分析 IV. ①G353.1

中国版本图书馆 CIP 数据核字 (2015) 第 080396 号

责任编辑: 付 艳 杨 静 宋开金/ 责任校对: 李 影

责任印制: 张 倩/ 封面设计: 楠竹文化

编辑部电话: 010-64033934

E-mail: fuyan@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>

三河市骏杰印刷有限公司印刷

科学出版社发行 各地新华书店经销

*

2015 年 5 月第 一 版 开本: 720 × 1000 1/16

2015 年 5 月第一次印刷 印张: 21 1/4

字数: 416 000

定价: 88.00 元

(如有印装质量问题, 我社负责调换)

序

一

近年来,知识图谱(mapping knowledge domains)已成为图书情报学研究的重要内容。从众多的研究成果中可以发现,知识图谱是一个跨学科研究的新兴领域,涉及图书情报学、计算机科学、科学学、数学、物理学等多个学科。

知识图谱作为一个重要的研究主题出现,得益于很多方面的发展。首先,随着科学技术的迅猛发展,知识在数量和质量两个方面都在快速增长,为知识图谱的研究提供了丰富的“原材料”。知识在精细化的同时,也在不断交叉融合。不同学科知识之间的关联性比以往更加紧密,知识图谱的理论和技術为知识可视化提供了强有力的支撑。其次,知识的“用户”对知识的渴求更加强烈,对知识的获取更加注重个性化。而个性化知识需求的满足依赖于更高的效率、更好的知识传递方式。知识图谱为人们从众多的知识当中获取自己所需要的那部分知识提供了便利。最后,知识组织技术与方法、计算机技术与工具等成为“知识”和“用户需求”之间的桥梁。通过 Histcite 这样的工具,可以非常方便地了解文献之间的引用关系。这种引用关系其实是一个知识不断传承、不断创新过程的直观体现。通过 CiteSpace 等类似的工具,学者可以比较快捷地了解到一个领域、一个学科出现的重要文献,通过了解文献之间的关联,进而得知科学家之间的联系及各学科之间的联系与发展图景。

学术机构是生产科学知识的基本组织单元,大到一所高校,小到一个由几个人组成的研究部门,每个学术机构在其发展过程中,均以研究者为中心,形成一系列的研究成果。这些研究成果体现出学术机构研究的内容和研究的轨迹。其题名、关键词、发文时间等属性从不同角度显示了学术机构的“知识印迹”。

魏瑞斌博士在《机构知识图谱的构建及其应用》中,首先对知识图谱的发展历程、知识图谱研究的理论研究与应用现状进行了系统梳理,在此基础上总结了机构知识图谱构建的方法和工具,并提出了一个基于 QICP 模型的机构知识图谱构建的流程。然后,分别以引文数据库、学术搜索引擎和项目数据库为数据源,用实证的方式尝试了机构知识图谱的构建。最后探讨了机构知识图谱在科研创新团队遴选、机构合作演化等方面的应用。该书研究综述详细,引用资料丰富,如有关 Web of Science 的数据、国内研究论文、研究项目等论述对 2014 年的新资料均有所引用。该书内容陈述清晰,评论得当,较为系统地对机构知识图谱进行了一系列相关研究,丰富了知识图谱的研究,对图书情报学、科学学、人才学等的深入研究具有重要参考价值和启发作用。

是为序。

叶继元

南京大学信息管理学院

2014 年 10 月 3 日

【序 二】

千百年来，人类的学习以记诵方式为主，听觉器官发挥着很大的作用。随着信息技术的飞速进步，可视化应用越来越普及，今后的学习将以可视化方式为主，视觉器官将发挥前所未有的作用。由于视觉器官在单位时间内的信息接收能力大大强于听觉器官，可视化方式成为主流学习方式后，人类的学习效率将大大提高，有可能带来一场认知革命。

为了适应这样的进程，知识组织方式也必将走向可视化之路。图书情报研究人员在知识可视化征程中将发挥非常重要的作用。

西方一些学者很喜欢编出若干哲理性很强的话，然后煞有介事地说，这是孔夫子说的，或这是中国人说的。例如，他们常说“一图胜万言”是中国谚语。而 *Natural History* 杂志 2000 年第 7 期的卷首语说：1921 年，美国一位广告主管 Frederic R. Barnard 想出了一句广告词——“一图胜万言”。为了引人注目，他忽悠说这是中国谚语。直到今天，仍有许多人以为这是中国谚语。不过，从某种意义上说，把这句话安在中国人身上也是恰当的，因为中国古代的地图制备能力曾经是十分强大的，郑和之所以能成功地七下西洋，地图作出了重要贡献。科学网博主、旅美华人学者李兆良先生用大量事实论证说：《万国坤輿全图》是中国人编绘的，中国人比哥伦布更早地“发现”了美洲大陆。

不管“一图胜万言”是谁说的吧，反正，图像的吸引力和感染力比文字强，这是公认的事实。因此，人们首先在传播领域大量利用图像，为的是打动人、“征服”人。后来，随着技术进步，可视化技术的成本下降，可视化的应用领域越来越广，学术研究也成为可视化大显身手的舞台，主要表现为，知识图谱的研究与应用受到热烈欢迎。在可视化手段的帮助下，人们的知识发现、知识表达与知识传播能力与日俱增。

魏瑞斌博士所研究的机构知识图谱，是知识图谱研究广阔天地中的一小方园地。现代社会与传统社会相比，组织化程度大大提高，团队、机构的重要性更加突出，因此，研究机构知识图谱是很有意义的。

魏博士的《机构知识图谱的构建及其应用》对知识图谱的研究与应用现状进行了较完整的梳理，介绍了构建机构知识图谱的方法、工具、流程，还有多个应用实例。对机构知识图谱感兴趣者，通过阅读该书可以不断加深对机构知识图谱的认识，并且可以仿照书中的例子做一些机构知识图谱分析的初步尝试。能力增强后，还可以独立探索机构知识图谱方面的新理念、新技术与新应用，开辟新天地。

现在，全社会对图形信息的需求是无穷无尽的，而社会提供高质量图形信息的能力是相当有限的，可以说出现了“供求危机”。危机出现的时刻，便是孕育创新性技术与应用的最佳时机。我们希望，有更多的学者投身于可视化研究领域，作出无愧于时代的贡献。

武夷山

中国科学技术信息研究所

2014年10月28日

目 录

序一 (叶继元)

序二 (武夷山)

第 1 章 知识图谱概述	1
第一节 知识图谱的发展历程	3
第二节 知识图谱的内涵	14
第三节 知识图谱与搜索引擎	28
第 2 章 知识图谱研究与应用现状	34
第一节 知识图谱研究现状分析	34
第二节 知识图谱的应用	58
第 3 章 构建机构知识图谱的方法、工具、数据源和流程	94
第一节 构建机构知识图谱的方法	94
第二节 构建机构知识图谱的工具及数据源	116
第三节 机构知识图谱构建流程	124
第 4 章 基于引文数据源的机构知识图谱构建	137
第一节 基于来源文献属性与被引文献属性的机构知识图谱	139
第二节 基于 Web of Science 数据的机构知识图谱	144
第三节 基于 CSSCI 数据的机构知识图谱比较	162
第四节 安徽财经大学机构知识图谱构建	206
第 5 章 基于学术搜索引擎的机构知识图谱构建	237
第一节 学术搜索引擎及其应用	237

第二节	数据采集与整理	239
第三节	三个研究机构数据的比较与分析	241
第四节	作者引文网络、合作网络与合作路径	246
第 6 章	基于项目数据源的机构项目图谱构建	253
第一节	国内相关研究文献评述	254
第二节	数据源及数据分析概念模型	259
第三节	机构项目知识图谱的实证研究	262
第 7 章	机构知识图谱的应用研究	289
第一节	基于两类合作网络的科研创新团队成员遴选研究	289
第二节	科研合作网络在创新团队成员遴选中的应用研究	297
第三节	基于期刊数据的机构合作演变研究	304
第四节	学术期刊发文主题演变的实证研究	316
后记		330

知识图谱概述

图画作为人类最早期的叙事方式，在人类文明出现以前，便跨越了语言、文字的障碍，成为史前文明时代最重要的交流方式。历史总是惊人地相似，几千年之后的今天，“读图时代”再次到来^①。2011年4月12日，百度首席执行官（Chief Executive Officer, CEO）李彦宏在百度联盟2011峰会上预言，中国互联网未来几年的主要发展机遇，将集中在“中间页”“读图时代”“应用为王”等三大领域。他还表示，全新的读图时代已经来临，主营图片内容的网站将获长足发展^②。

截至2011年12月，中国学术期刊网络出版总库收录全文文献总量3398万余篇；其中超过63%的文献中有规范的图片，平均每篇文献有图片4张左右。中国优秀硕士学位论文全文数据库收录优秀硕士学位论文134万余篇，其中超过80%的优秀硕士论文中有图片，平均每篇论文中有图片40张左右。中国博士学位论文全文数据库收录博士学位论文17万余篇，其中超过80%的博士论文中有图片，平均每篇论文中有图片50张左右^③。这些数据从一个侧面反映出，学术图片在科学研究领域是研究人员表达和传播知识内容的一种重要方式。

中国知识基础设施（China National Knowledge Infrastructure, CNKI）学术图片知识库是我国第一个学术类图

① <http://tech.qq.com/a/20120525/000230.htm> [2014-03-05]

② <http://tech.sina.com.cn/i/2011-04-12/09485394006.shtml> [2014-03-07]

③ <http://image.cnki.net/ProductStandard.aspx#> [2014-04-05]

片的知识库产品。它采用同方知网自主研发的智能挖掘技术，从各类学术文献中提取出千万量级的图形、图像等内容，加以规范化编辑整理，提供相似图表的检索、对比和分析功能等知识发现功能，将大大提高学术文献知识传播和应用效果。它以 CNKI 中国知识资源总库文献中提取的图片为分类元数据，把所有的图片划分为五个大类——形态图、谱线图、曲线图、系统图、分析图，二级类目下又细分为 254 种类型^①。该学术图片知识库现有图片 2 600 万余张，预计 2013 年之后每年新增图片 150 万余张^②。

笔者利用 CNKI 学术图片知识库查询发现，图书情报与数字图书馆、互联网技术、计算机软件及应用、新闻与传媒的学术图片数量分别为 56 129 张、411 305 张、2 708 384 张和 51 557 张（截至 2014 年 10 月 27 日）。从数据可以发现，不同学科之间学术图片的数量有较大的差异。例如，计算机软件及应用专题的学术图片数量达到了百万级，互联网技术的学术图片数量是十万级，图书情报与数字图书馆、新闻与传媒是万级。这一方面与学术期刊的数量有关，期刊数量越多，论文中出现学术图片的几率越大；另一方面也体现了学术图片在表达知识方面的研究习惯。另外，从表 1-1 看，每个专题的学术图片数量是一个稳定上升但有一定波动的状态。2013 年学术图片占总体数量的比例差别较大，可能与数据库中图片更新数据不及时有一定关系。

由于学术图片在表达和传播知识方面的优势，同时随着各种条件的成熟，知识图谱目前已经成为一个快速发展的跨学科研究领域。如果以 2003 年加利福尼亚的学术研究会作为这个领域研究开始的一个时间节点，它已经走过了 11 个年头。尽管学术界目前对知识图谱的概念还存在不同的认识，但已经形成了一个基本共识：知识图谱是通过不同类型的图形和图像来揭示知识单元之间的关系及知识发展的过程。

表 1-1 2005 ~ 2013 年四个专题学术图片数量所占比例 单位:%

时间	图书情报与数字图书馆	互联网技术	计算机软件及应用	新闻与传媒
2005	4	7	5	3
2006	5	8	8	4
2007	10	10	10	7
2008	11	12	13	10
2009	12	12	13	9
2010	13	8	8	9

① <http://image.cnki.net/PicTypeDef.aspx#> [2014-04-05]
② <http://image.cnki.net/ProductStandard.aspx#> [2014-04-05]

续表

时间	图书情报与数字图书馆	互联网技术	计算机软件及应用	新闻与传媒
2011	19	17	16	18
2012	15	11	12	14
2013	11	4	6	10

知识图谱的出现得益于很多方面的因素。第一，数据、信息和知识的数量在不断增加，为知识图谱的研究提供了丰富的“原材料”。第二，相关理论与技术的发展为研究人员加工“原材料”，使其以不同类型的图形、图像等图片形式呈现提供了可能。尤其是数据可视化、信息可视化和知识可视化理论和技术在近年来发展非常迅速，产生了许多可以让研究人员应用的各类可视化工具。第三，由于知识的急剧增加，研究者从现有知识当中查找自己所需要的知识的难度也相应增加，效率在下降，因此产生了从大量知识当中快速获取自己所需知识的需求。第四，传统互联网和移动互联网技术的不断发展，催生了许多个性化的应用，为用户需要的满足提供了不同的平台。第五，谷歌、百度、搜狗等搜索引擎企业和中国知网、万方数据等数据库服务发现了通过可视化技术提升信息服务水平的商机。

第一节 知识图谱的发展历程

一、一次学术研讨会

在《科学知识图谱：方法与应用》一书的导言中，大连理工大学刘则渊教授描述了他于2004年4月10日读完题为“科学家拟绘制科学门类图”的消息之后的那个不眠之夜的感触^①。他特别提到了2003年5月9日和10日在加利福尼亚大学欧文分校的美国国家科学院贝克曼激光研究中心举办的主题为“知识图谱测绘”（Mapping Knowledge Domains）的大型学术研讨会。来自计算机、信息与认识科学、数学、地理学等领域的专家、学者，从不同角度介绍了有关知识图谱的研究成果，共发表20多篇学术论文。从Ron Wild提供的图片（图1-1），我们可以看到美国科学情报研究所名誉所长加菲尔德（E. Garfield）等熟悉的面孔。

^① 刘则渊，陈悦，侯海燕，等. 科学知识图谱：方法与应用. 北京：人民出版社，2008：1-3

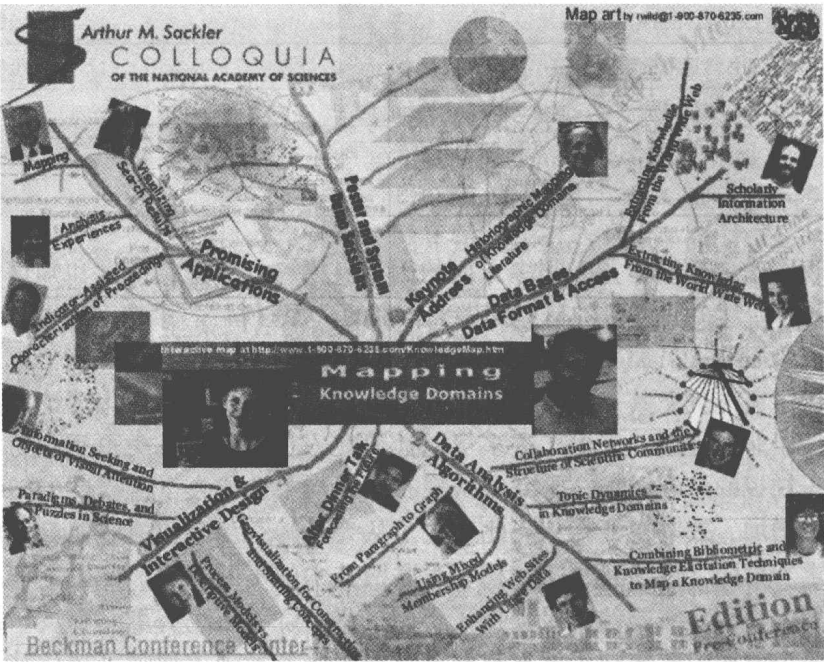


图 1-1 “知识图谱测绘”大型学术研讨会部分与会作者

资料来源: <http://www.1-900-870-6235.com/KnowledgeMap.htm> [2014-02-33]

该会议的专门网站 (<http://vw.indiana.edu/sackler03/>) 介绍了会议的相关信息。用户可以从网站免费下载加菲尔德的主题演讲和四个专题——数据库、数据格式和获取 (Data Bases, Data Format & Access)、数据分析算法 (Data Analysis Algorithms)、可视化和交互设计 (Visualization & Interaction Design)、应用前景 (Promising Applications) 的 PPT (表 1-2) 和视频资料。美国科学院院刊 (PNAS) 在 2004 年第 4 期上以专题形式刊出了该次会议的 20 篇论文 (表 1-3)。美国印第安纳大学 (Indiana University) 的谢夫林 (R. M. Shiffrin) 和伯尔纳 (K. Borner)^①, 介绍了知识图谱产生的背景及其现实意义, 并简要对该次会议的部分成果进行了评价, 最后提出了知识图谱面临的机遇和挑战。

陈悦和刘则渊^②在国内较早地介绍了这次学术会议的主要内容, 并根据知识图谱研究的性质和特征, 将 “Mapping Knowledge Domains” 译为 “科学知识图谱绘制”。从此之后, 国内研究人员开始陆续把 “知识图谱” 作为一个专业术语在研究成果中使用。

① Shiffrin R M , Borner K. Mapping knowledge domains. *Proceedings of the National Academy of Sciences*, 2004, 101 (4): 5183-5185

② 陈悦, 刘则渊. 悄然兴起的科学知识图谱. *科学学研究*, 2005, 23 (2): 149-154

表 1-2 “知识图谱”学术研讨会部分研究成果列表

报告主题	报告题目（作者及其机构）
Keynote Address	A historiograph of mapping knowledge domains（Eugene Garfield, ISI）
Data Bases, Data Format & Access	Extracting knowledge from the World Wide Web（Monika Henzinger & Steve Lawrence, Google）
	Mapping subsets of scholarly information（Paul Ginsparg, Cornell University）
Data Analysis Al- gorithms	From paragraph to graph（Thomas Landauer & Darrell Laham, University of Colorado）
	The structure of scientific collaboration networks（Mark Newman, University of Michigan）
	Using mixed membership models for mapping knowledge domains（Elena Erosheva, Uni- versity of Washington; Stephen Fienberg, Carnegie Mellon University; John Lafferty, Carnegie Mellon University）
	Topic dynamics in knowledge domains（Tom Griffiths, Stanford University; Mark Steyvers, University of California）
	Enhancing web sites with usage data（Jonathan Aizen, Daniel Huttenlocher; Jon Klein- berg & Antal Novak, Cornell University）
	Combining bibliometric and knowledge elicitation techniques to map a Knowledge domain （Kate McCain, Drexel University）
Visualization & In- teraction Design	Paradigms, debates, and puzzles in science: a visual exploration（Chaomei Chen, Drexel University）
	The simultaneous evolution of article and author networks in PNAS（Katy Börner, Jeegar Maru & Robert Goldstone, Indiana University）
Promising Applica- tions	Visualizing search results（Susan Dumais, Microsoft Corporation）
	Analysis experiences using information visualization（Beth Hetzler & Alan Turner, Pacific Northwest, National Laboratory）
	An indicator-based characterization of the Proceedings of the National Academy of Sciences （Kevin Boyack, Sandia National Laboratories）
	Mapping: from science papers to technology patents and on to company financial perform- ance（Francis Narin, President of CHI Research, Inc.）

资料来源：http: //vw. indiana. edu/sackler03/ [2014-02-09]

表 1-3 PNAS 2004 年第 4 期刊出的知识图谱研究成果

序号	论文
1	Shiffrin R M, Borner K. Mapping knowledge domains. <i>PNAS</i> , 2004, 101: 5183-5185
2	Henzinger M, Lawrence S. Extracting knowledge from the World Wide Web. <i>PNAS</i> , 2004, 101: 5186-5191
3	Boyack K W. Mapping knowledge domains: characterizing <i>PNAS</i> . <i>PNAS</i> , 2004, 101: 5192-5199
4	Newman M E J. Coauthorship networks and patterns of scientific collaboration. <i>PNAS</i> , 2004, 101: 5200-5205

续表

序号	论文
5	Dennis S. An unsupervised method for the extraction of propositional information from text. <i>PNAS</i> , 2004, 101: 5206-5213
6	Landauer T K, Laham D, Derr M. From paragraph to graph: latent semantic analysis for information visualization. <i>PNAS</i> , 2004, 101: 5214-5219
7	Erosheva E, Fienberg S, Lafferty J. Mixed-membership models of scientific publications. <i>PNAS</i> , 2004, 101: 5220-5227
8	Griffiths T L, Steyvers M. Finding scientific topics. <i>PNAS</i> , 2004, 101: 5228-5235
9	Ginsparg P, Houle P, Joachims T, et al. Mapping subsets of scholarly information. <i>PNAS</i> , 2004, 101: 5236-5240
10	Wilkinson D M, Huberman B A. A method for finding communities of related genes. <i>PNAS</i> , 2004, 101: 5241-5248
11	Hopcroft J, Khan O, Kulis B, et al. Tracking evolving communities in large linked networks. <i>PNAS</i> , 2004, 101: 5249-5253
12	Aizen J, Huttenlocher D, Kleinberg J, et al. Traffic-based feedback on the web. <i>PNAS</i> , 2004, 101: 5254-5260
13	Menczer F. Evolution of document networks. <i>PNAS</i> , 2004, 101: 5261-5265
14	Borner K, Maru J T, Goldstone R L. The simultaneous evolution of author and paper networks. <i>PNAS</i> , 2004, 101: 5266-5273
15	Skupin A. The world of geography: visualizing a knowledge domain with cartographic means. <i>PNAS</i> , 2004, 101: 5274-5278
16	MacEachren A M, Gahegan M, Pike W. Visualization for constructing and sharing geo-scientific concepts. <i>PNAS</i> , 2004, 101: 5279-5286
17	Mane K K, Borner K. Mapping topics and topic bursts in <i>PNAS</i> . <i>PNAS</i> , 2004, 101: 5287-5290
18	Morris S A, Yen G G. Crossmaps: visualization of overlapping relationships in collections of journal papers. <i>PNAS</i> , 2004, 101: 5291-5296
19	White H D, Lin X, Buzydlowski J W, et al. User-controlled mapping of significant literatures. <i>PNAS</i> , 2004, 101: 5297-5302
20	Chen C M. Progressive knowledge domain visualization. <i>PNAS</i> , 2004, 101: 5303-5310

资料来源: <http://www.pnas.org/search?fulltext=Mapping+knowledge+domains&submit=yes&x=11&y=18> [2013-10-07]

从这次会议研讨的内容及其研究成果可以看出,知识图谱是一个涉及计算机科学、数学、图书情报学等学科知识的跨学科研究领域。它不仅引起了高校和科研机构研究人员的关注,同时也得到了谷歌、微软等企业的研究人员的青睐。这些会议的议题也大体勾画出了知识图谱的八个研究领域。

(1) 知识图谱研究的目标是:揭示科学结构的演化规律,刻画科学领域的相关对象的影响力、重要性和科研资助;探索大数据规律等 [goals (e.g. science structure, vitality and changes over time; maps of impact, importance and fun-

ding; exploration and search of large databases)]。

(2) 实体图的绘制 (人名、词语、主题、论文和科研机构) [entities to be mapped (e. g. names, terms, topics, articles, science departments)]。

(3) 数据的表征 (相似空间、2-D/3-D 图、表格、矢量图) [representations of the data (e. g. similarity spaces, 2-D/3-D maps, tables, vectors)]。

(4) 数据源 (data sources)。

(5) 数据表征的产生和压缩算法 (algorithms for production and compression of the data representations)。

(6) 数据产生后表征的算法和方法 (algorithms and methods for mining the representations once they are produced)。

(7) 搜索数据表征的方法 (methods for searching the representations)。

(8) 与数据 (在线数据) 相适应的可视化和展示方法 [methods for (online) adaptive visualization and display]。

从会议 PPT 上截取的一些知识图谱的图片中可以看出, 知识图谱表现形式非常丰富和多样化。

二、一个研究团队

从 2005 年算起, 国内知识图谱的研究已经走过了 9 个年头。从杨思洛和韩雪珍^①的研究看, 国内已经形成了以刘则渊、汤建民、赵蓉英、宗乾进、马费成和卢章平为中心的 6 个研究团队。刘则渊团队是国内知识图谱研究的早期探索者, 也是目前该领域中规模和影响力较大的科研团队。

从刘则渊教授在《科学知识图谱: 方法与应用》^②一书的导言中可以了解到, 该研究团队围绕知识图谱领域开展了一系列相关活动。首先, 他们分工协作, 从美国知识图谱学术研讨会的网站下载了会议的 PPT 文稿, 并翻译了 PNAS 2004 年第 4 期的全部论文, 举办了以“科学知识图谱”为主题的学习报告会。其次, 刘则渊教授与德国著名科学计量学家克雷奇默 (H. Kretschmer) 教授一起创办了网络-信息-科学-经济计量实验室 (Weometrics, Informetrics, Scientometrics and Econometrics Lab, WISE LAB)。实验室的建立, 为师生从事知识图谱的研究与国内外学者的学术交流提供了优秀的硬件平台和良好的研究氛围。再次, 实验室团队撰写了一系列研究成果, 并参与了许多国际和国内的学术会议, 其研究成果也得到了国内外同行的认可。最后, 从 2009 年开始, 该实验室与中

① 杨思洛, 韩瑞珍. 知识图谱研究现状及趋势的可视化分析. 情报资料工作, 2012, (4): 22-28

② 刘则渊, 陈悦, 侯海燕, 等. 科学知识图谱: 方法与应用. 北京: 人民出版社, 2008: 1-3

国科学学与科技政策研究会科学学理论与学科建设专业委员会、汤森路透科技集团等联合举办了“全国科学知识图谱与科学计量学方法与应用高级讲习班”。讲习班的举办,向国内这个领域的一些研究人员普及了知识图谱的理论、方法和工具等方面的相关知识。

笔者从知网的学术期刊全文数据库,以题名中包含“知识图谱”为检索条件,共检索到406条记录(截至2014年2月10日)。检索结果按时间排序之后发现,在前20篇期刊论文中,有18篇的作者均来自大连理工大学;按被引频次排序后,前10篇高被引论文中,除排名第5、第8和第10之外,其余7篇的论文作者都来自大连理工大学。在知网博硕士论文数据库,以题名或关键词中包含“知识图谱”为检索条件,共检索到32条记录,有6篇论文的作者来自大连理工大学;其中2010年之前的5篇论文当中,有4篇的作者来自大连理工大学。另外,从国家图书馆和当当网等来源,笔者共检索到国内18本知识图谱研究领域的学术专著。其中,刘则渊教授领衔的学术团队从2008年开始先后出版了12本学术专著,占总体的67%。这些研究成果表明,大连理工大学 WISE 实验室的研究团队是国内知识图谱研究领域起步早、影响力较大和成果较为丰富的一支队伍。

三、一个人和一款软件

美国德雷克塞尔大学信息科学与技术学院的陈超美(Chaomei Chen)在1999年出版的《信息可视化与虚拟环境》(2004年增订第二版改为《信息可视化:跨越地平线》)和在2002年出版的《科学前沿图谱:知识可视化探索》作为信息可视化与科学知识图谱新领域的两部奠基性著作和突破性成就,在国际科学计量学界崭露头角^①。他还创办并主编了信息可视领域目前唯一的国际性期刊 *Information Visualization* (《信息可视化》,期刊主页: <http://ivi.sagepub.com/>)。他首先把一种探索关键路径的 Pathfinder 算法引入网络分析中,开发了 CiteSpace 软件。他创造性地把信息可视化技术结合起来,开创了以知识单元为分析基础的可视化综合性学术与应用领域,把对科学前沿的知识计量和知识管理推进到以知识图谱与可视化为辅助决策手段的新阶段。陈超美博士现在是大理工长江学者讲座教授,他与大连理工大学 WISE 实验室的团队成员合作完成了多篇学术论文,并且是全国科学知识图谱与科学计量学方法与应用高级讲习班主讲人之一。从2006年开始,南京大学、中国科学院及大连理工大学的多位教授和学生到德雷克塞尔大学做访问学者。陈超美博士与中国很多高校建立了长期的

^① 刘则渊,陈悦,侯海燕,等. 科学知识图谱:方法与应用. 北京:人民教育出版社,2008:23-24