● 廖胜姣[1,2]，肖仙桃[1]
（1. 中国科学院 国家科学图书馆兰州分馆，甘肃 兰州 730000；2. 中国科学院 研究生院，北京 100080）

# 科学知识图谱应用研究概述

摘 要：本文从应用的角度阐述了科学知识图谱的研究与发展现状，并结合国内发展概况，总结了科学知识图谱的国内外进展差异，并对其进行了展望。

关键词：科学知识图谱；知识地图；文献计量方法

Abstracts: This article expatiates the research and development status of scientific knowledge map from the application perspective, and combining with the domestic development situation, sums up the differences between China and foreign countries in the development of scientific knowledge map. The article also depicts the prospect for the future.

Keywords: scientific knowledge map; knowledge map; bibliometrical method

知识图谱是可视化显示知识资源及其关联的一种图形，可以绘制、挖掘、分析和显示知识间的相互关系，在组织内创造知识共享的环境，从而最终达到促进知识交流和研究深入的目的。从20世纪50年代至今，科学知识图谱的研究已经有几十年的历史。科学知识图谱出现之前，科学计量学家们一直努力在寻找一种同传统方法相比，具有更大的客观性、科学性、数据的有效性和高效率的新方法来研究科学学科的结构与进展。科学知识图谱出现之后，其相关的理论与应用研究不断涌现。本文试图从应用的角度对科学知识图谱的研究与发展状况进行一个系统的梳理，具体从应用领域、研究机构与网站以及绘图软件方面着手。

## 1 应用研究现状

从20世纪50年代开始兴起的各种文献计量方法为科学知识图谱的出现奠定了坚实的理论基础，是科学知识图谱理论与方法的“根”。如今，知识图谱已经成为计量学领域的一个新兴分支，活跃在各个领域的研究中。笔者将从应用领域、研究机构和软件方面阐述科学知识图谱的应用研究状况。

### 1.1 应用领域方面

科学知识图谱的应用领域很广，从科研到教学到社会问题的解决等，无不渗透。

1.1.1 应用于科研领域 笔者认为，知识图谱最早是在科研领域活跃起来的。在知识图谱中，学科前沿之间的交互关系是以空间的形式展现出来的。研究发现，科学引文与被引文之间往往有着学科内容上的联系。通过引文聚类分析，特别是从引文间的网状关系进行研究，能够探明有关学科之间的亲缘关系和结构，划定某学科的作者集体，分析推测学科间的交叉、渗透和衍生趋势，还能对某一学科的产生背景、发展概貌、突破性成就、相互渗透和今后发展方向进行分析，从而揭示科学的动态结构和某些发展规律。这里仅列举近些年知识图谱的一些应用研究情况。White, McCain, Garfield, Boyack, Huang等对知识图谱的用途进行了不断的扩充，得出知识图谱的主要应用有：文献、专利的结构分析；学科动态、社会网络、领域发展分析等；Shiffrin等认为，涉及到展开的学科间科学区域的知识图谱旨在绘制图形、挖掘、分析、分类、导航以及显现知识，等等[1]。

将知识图谱方法应用于构建学科知识图谱的研究人员也有一些：F. Janssens[2]等将沃德方法和K—值算法，用文本挖掘和文献计量方法分析了选中的五种期刊，得出科学计量学的学科结构图谱，并分析了两者结果的特点，认为将两种方法结合起来分析会得到更好的结果；K. W. McCain[3]等用PFNet方法和卡分类方法产生了软件工程领域作者地图；E. F. Reid[1]等绘制了恐怖主义研究领域的知识图谱，用引文分析、文献计量、社会网络分析方法对科学产出进行了基本的分析，对大量文献集进行内容地图分析，用共引分析来分析成对的研究人员间的联系，用领域可视化技术，如内容地图分析方法，座模型和共引分析方法来研究1965—2003年间文献和作者引用数据；E. Sanz—Casado[4]等用文献计量方法，基于朊病毒在1973—

2002年间的相关文献，用知识图谱的形式研究了该领域的现状以及发展趋势。

将共词应用于绘制各个领域的概念图的研究也有很多，如 de Looze和 Lemarie用于植物生物学领域，Bhattacharya和 Basu用于浓缩物质物理学领域，Peters和 van Raan用于化学工程领域，Ding Chowdhury和 Foo用于信息检索领域，Onyancha和 Ocholla用于医学领域。引文分析方法应用于知识图谱的绘制中的实例，可以以美国科学情报研究所（ISI）名誉所长加菲尔德（E Garfield）为首的科学团体创建了一系列关于知识域资料数据库为例。Garfield认为“引文数据的使用在书写科学的历史”，由此利用他们开发的 HistCite软件包，通过 ISI光盘引文索引（SCI SSCI或 AHCI）形成某一学科发展的历时的图谱[5]。

1.1.2 应用于教育领域 将图形和文字结合起来进行教学有比较久远的历史，特别是互联网和多媒体技术出现之后，这种教学方式更是得到推广。有研究表明，通过概念图等形式可以获得比传统教学更好的效果。这方面的研究有：R H Hall[6]研究了知识图谱在教学中的作用，通过实验，证明知识图谱有助于提高学生的学习效率；J Bidarra等[7]将知识图谱用于网络环境下的教学中；H E Her等[8]将知识图谱应用于教学中，让学生绘制自己的知识图谱，以了解其对内容的理解程度和解决问题的能力。所以，知识图谱可真正实现教与学的连接，可对教学有比较好的反馈。

1.1.3 应用于社会问题的解决方面 从笔者掌握的资料来看，将知识图谱用于解决社会问题，是知识图谱应用的一个拓展。该应用在 21世纪初才开始兴起，N Haritash和 B M Gupta[9]将知识图谱应用于政治中，用于政府的决策制定。他们通过绘制印度议会的两个机构的 S&T问题图谱，可以了解哪些是大家都关心的问题，了解民向，还可以在一个具体问题上了解大家的看法，便于政府的决策制定。R E Horn在“Knowledge Mapping for Complex Social Messes”[10]中将知识图谱应用于解决现实存在的问题，分析、认识复杂的社会信息间的关系，帮助决策者快速做出决策。该作者认为，知识图谱可以应用于很多方面：基于一个争论的焦点可以绘制一个知识图谱，清楚地将各方的理解放上去，有助于直观的认识事物，展示各方的相关细节，便于对比分析；可以显示逻辑和视觉结构，有助于从细节上了解主题；可以将不同的观点集合在一起，便于增加对话题的正确评价；可以是彩色的、一体化的有用的隐喻和图像，压缩了价值和看法，使得参与者可更详细地看到别人的观点，并增进相互沟通，更快地达成一致，使得参与者们跨越地理限制一起工作。

1.2 知识图谱的研究网站、机构方面

如今，国内外已经有专门的知识图谱研究机构，如 CWTS，致力于科学知识图谱的研究。

1）http://www.cwts.nl/ed/projects/home.html[11]。在 CWTS的网站上有专门的 Mapping板块。

该网站上刊载了 CWTS工程的作者、完成情况等内容。该工程的主要责任人是 E C M Noyons和 R K Buter，他们从 20世纪 90年代末至今，已经对文献计量方法绘制知识图谱进行了一系列的研究，如 1998年通过对一个领域进行多层次绘图[12]，首先产生一个领域的整体图，然后对强关联的主题聚类进行多维尺度分析，绘制低一级的图谱，产生每个区域的详细的结构图。文章使用的主要是共现方法，但是该方法绘制的图谱经常滞后于真实的发展。因为那些词基本都是清理过的、统一的和明确的。通过他们的受控特征，当编索引的人同意他们的领域相关性时，他们只能被输入数据库，所以该文引用了一种新工具 Nprool。另外他们还尝试开放了绘制科学图谱时使用的共词库，即以“开源”的形式，将自己绘图时的词库（词库中的词是在机选的条件下进行了人工筛选）公布出来，让读者和相关专家根据自己的认识添加或删除某些词，对词库进行修正、补充，从而使绘制的图谱具有可拓展性、动态性，同时也解决图谱的可读性问题，提高图谱的效率。这样构建的图谱具有极强的动态更新能力，也具有非常好的可读性。2001年 CWTS针对图谱没有发挥它的最大效用，写了一篇改进文献计量图谱功能的文章“Improving the Functionality of Interactive Bibliometric Science Maps”[13]。为了改善这个问题，文章结合了自顶向下和自底向上的过程，试图通过标出用户知道的点、熟悉的元素，让用户在一个熟悉的环境内理解图中的其他含义。从而让用户充分地认识图谱表达的含义，发挥它的最大效用。该网站上公布该机构目前正在研究 R&D的新一代交互图谱。

2）http://www.successcouncil/knowledge/[14]。这是 C Zins建立的人类知识图，其中对人类知识进行归类，总共十大类，包括知识基础、超自然物、物质和能量、空间和地球、非人类生物体、肉体和智力、社会、思想和艺术、技术和历史。每个分支下又有很多小的二级、三级分支。该网站主要的特色是图文并茂，其中的知识主要是以传统的主题目录方式组织在一起，辅以图片。该网站上的东西很全，其中的学术论坛是专业人员间的论坛，主要是反馈意见的平台。

C Zins有大量绘制图谱方面的研究经验，这从他发表的相关论文量中就可以看出来，从其网站上可以看到 1999年以来他至少发表了 15篇知识图谱方面的论文。

3）http://www.macrovu.com/[15]。该网站上简要介绍了知识图谱的几种应用，并提供深入学习的链接，是一个知识图谱相关知识的培训网站。

4）http://web.hku.hk/~jwilam/PCEd_FT_2003_IT/mappingware.html[16]。网上商业与共享的知识图（尤其概念图）软件极多，且大多能支持中文，例如 Inspiration（国外学界极流行的知识图软件）或 MindMapper（脑图创始人 Buzan所开创公司的产品），十分好用，但是都要收费。这个网站上提供了很多免费且支持中文，而又能作教与学用途的知识图软件的链接。还介绍了一些用于商业的收费绘制知识图的软件链接，是笔者认为比较全面的教学软件的网站。

5）http://cluster.cis.drexel.edu/~cchen/citespace/[17]。这个网站是陈超美的个人网站，上面提供了陈超美自己设计的绘制知识图谱的免费软件及其下载链接，还提供了陈超美个人取得的有关科学知识图谱方面的成果。可以发现，陈超美在知识图谱方面的研究还是具有一定的深度。

**1.3** 知识图谱的软件工具的增加方面

1）最初的一些软件简述。Garfield利用他们开发的 HistCite软件包，通过 ISI光盘引文索引形成某一学科发展的历时的图谱。HisCite系统是一个很好的引文历史分析工具，当在 WoS上显示出一个有标记的列表时，对每一个源文件都生成包括所有被引文献的专家文件，这些引文收集被存储成由 HisCite处理生成的 ASCII文件，用以产生历时代和其他类型表格，以及显示出在本收集之内和之外被引用最多的文献的编年图表[18]。Small等人首先开发了基于共引理论的单机系统 SCI—Map来描绘科学文献间的结构；通过连续时间内共引聚类图的历时比较，反映科学结构的变化；从不同学科间的共引关系中寻找某一学科到另一学科的可通路径，从而描述知识结构；基于 ISI数据将共引聚类用于科学研究前沿分析[19]。

2）SPSS。SPSS（Statistical Package for Social Science）是由美国 SPSS公司自 20世纪 80年代初开发的大型社会科学统计软件包，是目前世界上流行的三大统计分析软件之一，具有完整的数据输入、编辑、统计分析、报表、图形制作等功能，除了适合于社会科学之外，还适用于自然科学各领域的统计分析。近年来，SPSS为各领域的科研工作者所广泛使用。

SPSS内嵌的相关距离分析、因子分析（主成分分析）、多维尺度分析和聚类分析功能是进行科学知识图谱绘制常用的多元统计分析工具。

3）Thomson Data Analyzer。Thomson Data Analyzer是 Thomson科技集团基于 VantagePoint技术开发的一种数据挖掘软件，可用于跟踪竞争对手，俯瞰整个技术背景，发现新的趋势，从不同角度考察某一主题等。Thomson Data Analyzer对于分析科学文献数据具有强大优势，不仅可载入 ISI中多种数据库的数据，还能对大量数据进行清理合并等，其强大的数据处理功能是其他软件所不能比拟的[19]。

在绘制知识图谱过程中，该软件常用于绘制基础的图谱，常用的功能有：数据清理、列表功能、矩阵功能、图功能。其中，Thomson Data Analyzer的生成矩阵功能可对各种字段进行矩阵分析，可产生共现聚类、自动相关矩阵、交叉关联矩阵和因子矩阵，其中使用的相关系数是 Pearson相关、余弦或最大比例相关。该功能减少了传统手工统计频次的工作量，节省了时间；Thomson Data Analyzer的生成图功能可依据导入的数据产生交叉关联图、自动相关图和因子图。

4）Bibexcel。Bibexcel是由瑞典科学计量学家 Persson开发的一个计量软件[20]。目前该软件为仅用于科学研究的免费软件。Bibexcel可帮助用户分析文献计量数据以及任何以相似模式存储的数据。其思想是：为进一步处理产生能导入到 Excel的数据文件或其他获得标记数据记录的项目。Bibexcel的功能包括：文献计量学分析、引文分析、共引分析、引文耦合分析，聚类分析，科学知识图谱的绘制等，它可以和 Pajek，Netdraw结合起来使用。它可以使用的数据包括 ISI的 SCI，SSCI，A&HCI的记录，也可对其他类型数据进行分析。

5）WordsmithTools。OxfordWordsmithTools是牛津大学开发的商业性词频分析软件。其主要功能包括 Wordlist和 Concord Tool两种。Wordlist Tool可以将一个文本中的所有单词按使用频次进行排序，而 Concord Tool可以帮助我们找到与任意一个单词搭配的词组。该软件被牛津大学的语言教师和学生用于辞典编辑工作，研究语言模式的学者们也借助它对世界上多种语言模式进行比较研究。

6）Pajek。Pajek是一个基于 Windows的用于将大型网络可视化的社会网络分析软件。在斯洛文尼亚语中，Pajek的意思是蜘蛛。Pajek的设计是基于图论、网络分析以及可视化软件等发展而来的。其主要功能是将一个大型网络分解为一些小型的子网络，并展示这些子网络的关系。

7）CiteSpace。CiteSpace是陈超美个人网站上提供的分析和可视化科学文献的一个免费的 Java应用程序。CiteSpace扩展了它的范围，包括了其他额外的数据来源，例如 NSF奖总结。它面向的用户主要是科研人员、医学界、科学政策研究者和医学图书馆员，将信息可视化方法、文献计量方法和数据挖掘算法集成起来，是一个在引

文数据中提取模式的交互式的可视化工具，其绘制图谱、建立节点之间关联的依据是“共引”与“引文”。

还有很多其他的绘图工具，这里就不一一列举。

## 2 国内知识图谱应用研究现状

总体来讲，无论是企业还是科研领域，我国对知识图谱的关注滞后于国外。

相比国外知识图谱的研究状况，我国起步稍晚，但是也取得了一些成绩。从20世纪90年代至今，我国的专业人员也开始了科学知识图谱的研究，并有专门的研究机构（如大连理工大学的科学学与科学技术管理研究所）一直在关注科学知识图谱的研究和发展。不过，追溯起来可以发现，“知识图谱”、“知识地图”这些术语在我国学术界出现也就是这几年的事情，之前的研究并不系统，而且大多是对绘制方法进行研究。有关知识图谱方面的零零散散的研究成果也有很多，如中国科学院的耿海英[19]毕业论文最后的实证部分，是用共引分析方法等绘制了情报学作者间知识图谱，并和White的结果进行对照，分析异同。涉及到具体构建知识图谱系统的专家有我国社会科学院的李思经老师，他在知识图谱方面研究比较深入，也有了一些成果，他的学生康永兴在2006年的毕业论文中构建了学科知识图谱系统，是将知识图谱系统应用于科学的一个探索[21]。大连理工大学科学学与科学技术管理研究所的刘则渊老师等是纯科学知识图谱绘制方面研究的专家，该研究所的一系列研究人员是我国系统研究科学知识图谱的领头人。中国科学院国家科学图书馆刚刚建成了一个基于SCI和ESI数据库绘制各领域科学图谱的系统。2008年5月17日在中国科学院国家科学图书馆举办了一次科学地图展览等。以上实例说明，越来越多的人开始关注知识图谱的研究和应用。不过，有关知识图谱本身的系统研究几乎没有。

## 3 结束语

总体来讲，国内在知识图谱的应用方面缺少理论上的实证分析，主要是将知识图谱作为一个工具，应用于各个领域，而且相对于国外，应用研究还比较薄弱。但是由于知识图谱是科学计量学领域的一个新的活跃分支，其历史还比较短，所以，国内外在知识图谱的应用研究方面，差距并不大。

如今，科学知识图谱已经成为一种理论与方法得到了很多科研人员的肯定，其应用领域也在不断的拓展，已经成为科学计量学领域的一个热点研究方向。我们有理由相信，在不久的将来，我国将会加入到科学知识图谱方向的研究及应用的世界前沿之列，其理论与应用将会得到进一步的发展。□

参考文献


[1] REDDY F, CHEN H. Mapping the contemporary terrorism research domain [J]. Int J Human-computer Studies, 2007, 65.

[2] GLENISSON P, et al. Combining full text and bibliometric information in mapping scientific disciplines [J]. Information Processing and Management, 2005, 41.

[3] MCCAIN K W, VERNER J M, HISLOP G W, et al. The use of bibliometric and knowledge elicitation techniques to map a knowledge domain-software engineering in the 1990's [J]. Scientometrics, 2005, 65 (1).

[4] SANZ-CASADO E, SUAREZ-BALSEIRO C, INBARRENMAESTRO J, et al. Bibliometric mapping of scientific research on prion diseases, 1973—2002 [J]. Information Processing and Management, 2007, 43.

[5] 侯海燕. 基于知识图谱的科学计量学进展研究 [D]. 大连：大连理工大学，2006.

[6] HALL R H. Cognitive and affective outcomes of learning from knowledge maps [J]. Contemporary Educational Psychology, 1996, 21.

[7] BIDARRA J, DIAS A. Ecological strategies and knowledge mapping [M]. [S. l.]: Springer-Verlag Berlin Heidelberg, 2004.

[8] HERL H E, O'NEIL H F, CHUNG G K W K, et al. Reliability and validity of a computer-based knowledge mapping system to measure content understanding [J]. Computer in Human Behavior, 1999, 15.

[9] HARITASHN, GUPTA B M. Mapping of S&T issues in the indian parliament: a scientometric analysis of questions raised in both houses of the parliament [J]. Scientometrics, 2002, 54 (1): 91-102.

[10] HORN R E. Knowledge mapping for complex social messes: a presentation to the "foundations in the knowledge economy" at the David and Lucile Packard Foundation [EB/OL]. http://www.Stanford.edu/~thorn/spchpackard.htm.

[11] http://www.cwts.nl/ed/projects/home.html.

[12] NOYONS E C M, VAN RAAN A F J. Advanced mapping of science and technology [J]. Scientometrics, 1998, 41 (1-2): 61-67.

[13] BUTER R K, NOYONS E M. Improving the functionality of interactive bibliometric science maps [J]. Scientometrics, 2001, 51 (1): 55-68.

[14] http://www.successcoil/knowledge/.

[15] http://www.macrovu.com/.

[16] http://web.hku.hk/~jwilam/PCEd_FT_2003_IT/mappingware.htm.

[17] http://cluster.cis.drexel.edu/~cchen/citespace/.

[18] 陈悦，刘则渊. 悄然兴起的科学知识图谱 [J]. 科学学研究，2005, 23 (2).

[19] 耿海英. 共引分析方法及其应用研究 [D]. 北京：中国科学院，2007.

[20] http://www.umu.se/inforsk/Bibexcel/.

[21] 康永兴. 科研机构知识管理中知识地图的设计与构建 [D]. 北京：中国社会科学院，2006.


作者简介：廖胜姣，女，1983年生，硕士生。
肖仙桃，女，1965年生，硕士生导师，研究馆员。

收稿日期：2008—08—06