

现代教育学论丛

丛书主编◎郑信军 彭小明

000000000000000000000000000001



知识图谱理论 在教育与心理研究中的 应用

*Application of the Theory of Mapping
Knowledge Domain in Educational and
Psychological Research*

郭文斌 著

绘制海量信息的知识图谱是现代教育研究者必备的素养

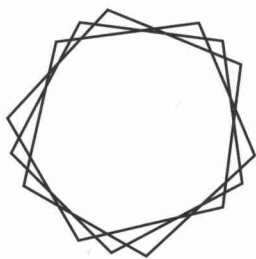
绘制海量信息的知识图谱是现代教育研究者必备的素养，它主要通过三个方面表现出来：一是掌握知识图谱的相关原理；二是能够正确操作绘制知识图谱的各种工具；三是能够对绘制出的知识图谱进行量质结合的科学解读并将其予以展示。



ZHEJIANG UNIVERSITY PRESS
浙江大学出版社

现代教育学论丛

丛书主编◎郑信军 彭小明



知识图谱理论 在教育与心理研究中的 应用

*Application of the Theory of Mapping
Knowledge Domain in Educational and
Psychological Research*



郭文斌 著



ZHEJIANG UNIVERSITY PRESS

浙江大学出版社

图书在版编目(CIP)数据

知识图谱理论在教育与心理研究中的应用 / 郭文斌
著. —杭州:浙江大学出版社,2015.5
ISBN 978-7-308-14673-9

I. ①知… II. ①郭… III. ①知识信息处理—研究
IV. ①TP391

中国版本图书馆 CIP 数据核字(2015)第 097308 号

知识图谱理论在教育与心理研究中的应用

郭文斌 著

丛书策划	吴伟伟 weiweiwu@zju.edu.cn
责任编辑	
封面设计	续设计
出版发行	浙江大学出版社 (杭州市天目山路 148 号 邮政编码 310007) (网址: http://www.zjupress.com)
排 版	浙江时代出版服务有限公司
印 刷	杭州日报报业集团盛元印务有限公司
开 本	710mm×1000mm 1/16
印 张	10.75
字 数	171 千
版 次	2015 年 5 月第 1 版 2015 年 5 月第 1 次印刷
书 号	ISBN 978-7-308-14673-9
定 价	35.00 元

版权所有 翻印必究 印装差错 负责调换

浙江大学出版社发行部联系方式 (0571)88925591; <http://zjdxcs.tmall.com>

前 言

笔者自 1994 年参加工作以来,一直从事科研和教学工作。随着全球数字化时代的到来,海量电子文献不断涌现。如何在这样的时代背景下,对海量信息进行更为高效、客观的处理,快速选择出自己需要领域的有效信息,成为长期困扰笔者的难题。凭借 2009 年笔者承担学校研究生精品课程“教育研究方法”的契机,笔者进行了大量文献的筛选和查找,终于接触到了科学计量方法,通过科学网联系到了陈超美教授和崔雷教授,向他们请教了知识图谱方面的事项,得到了他们的指点。在他们的鼓励和指点下,笔者开始尝试写作发表了几篇以知识图谱为主题的文章,获得了很好的反响,应邀到陕西师范大学、西北师范大学、衡水学院、新疆教育学院等高校进行讲学和示范。为了更好地对知识图谱方法进行梳理,为今后更好的驾驭它,也为了给授课的学生带来更多的关于知识图谱方面的内容,笔者斗胆尝试着撰写了本书。

本书作为《教育研究方法》(科学出版社 2012 年版)的姊妹篇,是对其第九章第四节中的“内容分析法新走向”内容的进一步延展,因此在撰写过程中,保持了《教育研究方法》撰写的几个原则:

第一,为了保持内容的前沿性和准确性,在文献选取上,大量引用了最近 5 年在权威刊物发表的论文和有代表性的著作。

第二,为了便于读者阅读,在相应章节增加了延伸阅读,通过延伸阅读,将一些有关的概念、材料、操作示范等内容呈现给读者,供他们拓展

阅读使用。

第三,为了促进读者更好地掌握每个章节介绍的内容,在每个章节都提供了应用举例,供读者进一步将理论和应用结合起来进行体会领悟。

第四,为了使本书内容更加实用,切合读者的需求,结合笔者近几年对授课学生的实际教学情况和自身使用情况对每个章节撰写的内容进行了取舍。

本书对每个章节的总体框架是:先介绍每个章节的主要内容,再以具体例子来巩固已有的内容。为了更好地使学习者深入学习每个章节的相关内容,在每章具体内容部分还不时插入延伸阅读材料,供有需要的学习者根据自身学习需求灵活阅读使用。

本书共以六章内容对知识图谱理论在教育与心理研究中的应用进行了初步的论述和示范展示:第一章,通过知识图谱的概念及发展历程、知识图谱的原理及特征、知识图谱在教育与心理研究中应用的意义三个方面对知识图谱进行了概述;第二章,知识图谱的基本方法,主要介绍了引文分析法、共被引分析法、多元统计分析法、词频分析法、社会网络分析法五种具体方法的含义、使用步骤、评价和展望;第三章,重点介绍 CiteSpace、BICOMB 和 SPSS、HistCite 三种知识图谱应用软件的操作原理、运行环境以及操作步骤;第四章,文献信息的查询和保存,较为详细地介绍了两种常用绘制知识图谱文献库 Web of Science 文献库和 CNKI 文献库中文献的查阅和保存技巧;第五章,绘制知识图谱的操作,通过对 CiteSpace、BICOMB 和 SPSS 结合绘制知识图谱的详细操作步骤以及操作示意图展示,来详细呈现其使用的具体过程;第六章,知识图谱论文的呈现,通过知识图谱论文的构成以及知识图谱论文示例具体说明如何撰写知识图谱论文。

在本书的撰写过程中,美国德雷赛尔大学信息科学与技术学院陈超美教授、中国医科大学医学信息学系崔雷教授、上海交通大学图书馆熊海强教授提供了很多有用的资料;我的硕士生导师彭德华教授、博士生导师方俊明教授都在我撰写书稿期间给予了我热情的鼓励、殷切的期望和可行性的建议,方俊明教授还特意为本书作序,尤其令我感动;我校教师教育学院的各位领导和同事给予了我精神和物质方面的大力支持;浙

江大学出版社的吴伟伟编辑为本书的顺利出版对书稿的校对付出了辛勤的劳动；我远方父母的报喜不报忧的处理方式，爱人和女儿悉心帮助我打理生活中的点点滴滴，都是为了支持我全身心的投入书稿的撰写。在此一并表示衷心的感谢！

本书虽然在撰写中借鉴和参考了大量的文献资料，但由于本人能力有限，加之时间较为仓促，书中错误和疏漏在所难免，恳请大家批评指正。

郭文斌

2015年3月于温州

序

人类在长期的实践过程中,不仅积累了丰富的知识,还通过概念、判断、推理、假说、预见等思维形式形成的系统的学科知识。古人云“学成于集,新故相资而新其故”(《周易外传·系辞上》),我们的祖先很早就认识到有效的学习和探究既是得益于对前人知识的继承,也得益于对以往知识的推陈出新。

科学革命理念的倡导者库恩(T. S. Kuhn)认为,在不同的历史时期,科学有不同的发展模式,研究者需要采用收敛式的思维或教条性的思维来维持知识的传承和发展的稳定,更需要采用发散性思维或批判性思维来打破过时的范式而另辟新径。在中国的成语字典中,“按图索骥”原本是用来讥讽那些食古不化、拘泥成法,只会按照图样去寻找好马的人。但是,随着全球信息化、数字化、网络化时代的到来,海量电子文献不断涌现,采用计算机信息处理的方法来编制知识图谱则成为现代科学研究中的一种值得提倡的理论和方法。这是因为它可以帮助我们浩如烟海的文献中,清晰地看到现代学科知识体系是如何相互交叉、相互渗透和相互启迪,又能帮助我们探究某一学科内在的规律和发展趋势。

在科学研究中,“理解”和“表达”原本也是矛盾的对立的统一。人们一方面可以借助于知识图谱的编制来呈现和凸显学科发展的轨迹,另一方面也可以通过对知识图谱的分析来预测学科的发展方向。作为一种伴随计算机技术发展起来的文献研究的理论和方法本身也是多学科相互影响的产物,它受到地图学、认知地图、学科地图学的启示。例如,伴

随现代信息技术,如地理信息系统(GIS),遥感(RS)和全球定位系统(GPS)的发展,现代地图学已成为技术性很强,理论体系相对成熟的学科;从格式塔心理学家托尔曼在“迷宫实验”基础上最早提出的认知地图,到发生认识论的发展心理学家的认知结构理论,再发展到受现代认知神经学科影响的“意念地图”(mental map)、“概念地图”(concept map)和“学科地图”(map of disciplines)同样也经历了不同的发展阶段。相比之下,尝试采用知识图谱的理论和方法对我国教育,包括普通教育和特殊教育问题的文献研究还是刚刚开始,因此,联系教育实践,比较系统地介绍知识图谱的理论和方法的著作更显得弥足珍贵。

郭文斌是我博导生涯中所指导 23 名博士研究生中的关门弟子,也是我认为最努力、最勤奋好学的研究生之一。据我所知,近些年来,郭文斌一直致力于知识图谱的研究,早在攻读博士学位期间,就发表了数篇用知识图谱的理论和方法来研究心理和特殊教育学领域问题的论文,也曾被邀到一些师范院校进行讲学和示范。因此,当我看到这本书稿时,虽不感到意外,但还是为他的潜心研究和孜孜不倦而感动,颇有“士别三日,当刮目相待”之感。

本书作为《教育研究方法》一书的姊妹篇,用六章的篇幅,系统地阐述了知识图谱理论在教育与心理研究中应用的原理与方法,介绍了分析和绘制知识图谱的相关软件,呈现知识图谱研究的操作过程和研究示范。正如他在前言中所指出的,全书既保持了《教育研究方法》撰写的主要原则,也凸显了知识图谱研究的新意和多学科协调研究的特点。诚然,由于我国教育研究的信息化、数据化还处于发展的初级阶段,还有许多著作和论文没有列入电子文献,这使运用知识图谱来定量化地处理文献,研究宏观和中观的教育问题还受到某些限制。但是,随着我国教育研究信息化、数据化程度的提高,我相信,知识图谱,作为一种现代学科的研究原理和方法,将会越来越展示出研究的效能和诱人的前景。

阅后有感,姑且为序。

华东师范大学终身教授 方俊明

2014年12月6日

(28) 知识图谱的概述 章正深

(38) 知识图谱的原理及特征 章正深

(39) 知识图谱在教育与心理研究中的应用 章正深

(41) 知识图谱的基本方法 章正深

(41) 知识图谱应用软件介绍 章正深

(41) 知识信息的查询和保存 章正深

(41) 知识信息的查询和保存 章正深

(41) 知识信息的查询和保存 章正深

(41) 知识信息的查询和保存 章正深

(41) 知识信息的查询和保存 章正深

(41) 知识信息的查询和保存 章正深

目 录

第一章 知识图谱概述 (1)

 第一节 知识图谱的概念及发展历程 (2)

 第二节 知识图谱的原理及特征 (6)

 第三节 知识图谱在教育与心理研究中的应用的意义 (11)

第二章 知识图谱的基本方法 (13)

 第一节 引文分析法 (13)

 第二节 共被引分析法 (24)

 第三节 多元统计分析法 (30)

 第四节 词频分析法 (42)

 第五节 社会网络分析法 (47)

第三章 知识图谱应用软件介绍 (56)

 第一节 CiteSpace 介绍 (56)

 第二节 BICOMB 介绍 (62)

 第三节 HistCite 介绍 (65)

第四章 文献信息的查询和保存 (71)

 第一节 Web of Science 文献的查询和保存 (71)

 第二节 CNKI 文献的查询和保存 (77)

第五章 绘制知识图谱的操作	(83)
第一节 CiteSpace 绘制知识图谱的操作	(83)
第二节 BICOMB 和 SPSS 结合的操作	(88)
第六章 知识图谱论文呈现	(108)
第一节 知识图谱论文的构成	(108)
第二节 知识图谱论文示例	(114)
附录	(142)
索引	(155)
跋	(159)

第一章 知识图谱概述

过去,人们对一个学科研究进行资料综述的时候,更多的是基于个体主观经验对于资料的加工^{①②},较少采用科学计量的方法对资料进行综合分析。在今天浩如烟海的文献中,如果没有科学的计量分析方法,仅凭个人的主观经验判断,难免会产生错误或者不当的归类和总结。^③随着互联网技术的普及和数字化时代的到来,采用数据挖掘和信息可视化技术,对已有信息进行整理,产生新的知识的科学计量学逐渐发展并且成熟起来。^④知识图谱作为当前国际科学计量学领域热门的方法之一,在近年越来越受到研究者的重视和青睐,2003年美国科学院组织的“Mapping Knowledge Domains”讨论会,预示着世界科学计量学中知识图谱和可视化研究的春天已经到来。要认识和把握知识图谱的准确含义,需要对其概念、发展历程、应用原理以及意义进行全面的了解。

① 郭文斌:《马斯洛人际关系心理学思想初探》,《渭南师范学院学报》2006年第1期,第82—85页(《人大复印资料·心理学》2006年第5期,第74—77页。全文转载)。

② 陈秋珠、郭文斌:《当前我国心理学中国化进程中存在的问题》,《渭南师范学院学报》2002年第1期,第80—82页(《人大复印资料·心理学》2002年第5期,第2—4页。全文转载)。

③ 郭文斌、陈秋珠:《特殊教育研究热点知识图谱》,《华东师范大学学报》(教育科学版)2012年第3期,第49—54页。

④ 郭文斌、方俊明、陈秋珠:《基于关键词共词分析的我国自闭症热点研究》,《西北师大学报》(社会科学版)2012年第1期,第128—132页。

第一节 知识图谱的概念及发展历程

一、知识图谱的概念

了解地图、知识地图以及图谱的概念有助于准确把握知识图谱的概念。

(一)地图与知识地图

地图(map)指以二维或三维空间形式显示地形和人类活动及相关特征的地理学概念。地图能够科学地反映出自然和社会经济现象的分布特征及其相互关系。在电子和数字时代,地图已经由传统的纸质地图,演变成了数字地图和电子地图。但不论其形式如何演变,依然不变的是地图的主要特征^①:第一,由特殊的数学法则产生的可量测性。特殊的数学法则包含地图投影、地图比例尺和地图定向三个方面。第二,由使用地图语言表示事物所产生的直观性。地图语言包括地图符号和地图注记两部分。第三,由实施制图综合产生的一览性。第四,必须遵循一定的数学法则。地图是绘制在平面上的,必须准确地反映它与客观实体在位置、属性等要素之间的关系。第五,必须经过科学概括。缩小了的地图不可能容纳地面所有的现象。第六,具有完整的符号系统。

知识地图(Knowledge Map)也称为知识分布图或知识映射图。最初源于美国捷运公司绘制的充满知识资源的美国地图。此后,知识地图表示带有索引号或用其他方式表示层次关系的表格和文件,以及用来表示信息资源与各部门或人员之间关系的信息资源管理表和信息资源分布图。^② 知识地图描述了一个组织在知识转化周期过程中的知识资源具体分布及变化

^① 祝国瑞:《地图学》,武汉大学出版社2004年版,第1—5页。

^② 乐飞红、陈锐:《企业知识管理实现流程中知识地图的几个问题》,《图书情报知识》2000年第3期,第15—17页。

情况^①,有助于组织成员把握本系统内部知识配置,为其进行知识寻求、创造提供准确的可用信息。知识地图产生的步骤如图 1-1 所示。

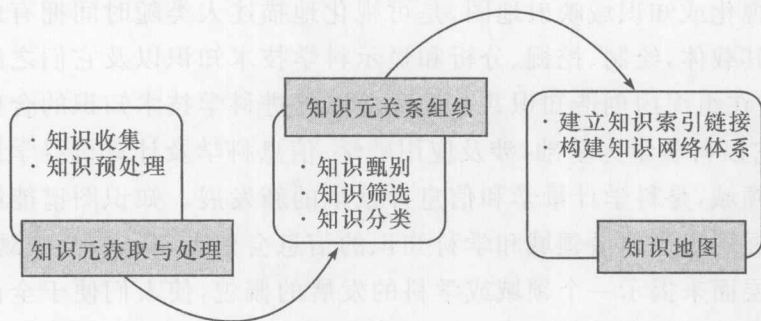


图 1-1 知识地图的绘制步骤

从图 1-1 可以看出,知识地图的绘制包括三个步骤:首先是通过知识收集和知识预处理获得知识元;其次,通过对知识的甄别、筛选和分类^②,将无序的知识元间信息进行重组,以构建知识之间的相互关系;最后,在各个知识之间建立索引,通过索引来链接和标识它们之间的位置和关系,以此来呈现知识地图。绘制知识地图的目的就是对组织知识资源总体分布情况进行可视化描述,包括组织知识资源的存在情况及其载体,以及资源之间可能存在的联系。实质上就是利用现代信息技术制作的组织知识资源的总目录和各知识条目之间关系的综合体以及组织专家的导航系统,是利用构造地图的方法将各类知识资源中的知识关联起来,使之成为一个网络^③,提高了知识的利用率。

(二) 图谱与知识图谱

图谱(atlas)指系统地编辑起来的、根据实物描绘或摄制的图,是研

^① Alavi M, Leidner D E. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quarterly*, 2001:107-136.

^② 梁勇、章成志、王昊:《基于 CSSCI 的期刊知识地图的构建》,《现代图书情报技术》2008 年第 2 期,第 58—63 页。

^③ 秦长江、侯汉清:《知识图谱——信息管理与知识管理的新领域》,《大学图书馆学报》2009 年第 1 期,第 30—37 页。

究某一学科所用的资料。^① 后来,泛指按类编制的图集。

知识图谱(Mapping Knowledge Domain)也被称为科学知识图谱、知识域可视化或知识域映射地图,是可视化地描述人类随时间拥有的知识资源及其载体,绘制、挖掘、分析和显示科学技术知识以及它们之间的相互联系,在组织内创造知识共享的环境以促进科学技术知识的合作和深入。^② 它以科学学为基础,涉及应用数学、信息科学及计算机科学诸学科交叉的领域,是科学计量学和信息计量学的新发展。知识图谱能够用直观图像展现出最前沿领域和学科知识的信息会聚点,从宏观、中观、微观等不同层面来揭示一个领域或学科的发展的概貌,使人们便于全面审视一个学科的结构和研究热点、重点等信息^③,生成新的知识。借助于知识图谱,人们可以将知识和信息中引人注目的最前沿领域或学科制高点,以可视化的图像直观地展现出来,挖掘、分析和显示知识及其联系,进而判定学科前沿的历史演进路径。

值得注意的是,虽然可以将知识地图作为知识图谱的一种形式,但知识图谱比知识地图更能揭示知识之间的联系及知识的进化规律。知识图谱与知识地图的区别在于前者一般不提供方便使用者快速获取与知识资源关联的相关信息。

二、知识图谱的发展历程

(一)知识图谱的产生

人类经历过了五次信息革命:语言的使用;文字的创造;印刷术的发明;电报、电话、广播、电视的发明和普及;计算机技术及现代通信技术的普及与应用。现在,正在经历第六次信息革命:云计算与物联网的发展与应用。在互联网和数字化时代没有到来之前,学者们为了解一个学科

^① 中国社会科学院语言研究所词典编辑室:《现代汉语词典》(修订版),商务印书馆1996年版,第1275页。

^② 刘则渊、陈悦、侯海燕等著:《科学知识图谱:方法与应用》,人民出版社2008年版,第5页。

^③ 郭文斌、陈秋珠:《特殊教育研究热点知识图谱》,《华东师范大学学报》(教育科学版)2012年第3期,第49—54页。

领域发展的整体状况,必须查阅该领域的几乎所有文献,然后经过自己的加工,从大量文献中筛选出相对重要的文献。这样的工作不仅耗费时间,而且也非常困难。不同学者选取文献时,因为站立角度和主观判断的差异,就如盲人摸象中的各个盲人,往往选取的材料有很大出入,结论也难以得到重复验证。

随着知识大爆炸和信息化时代的到来,海量信息时代也随之到来,它就像一头不仅奔跑而且还在不断变化形状的大象。这个时候想通过传统方法来捕捉学科发展的脉动越来越困难。在对多学科领域进行研究时,对文献的动态发展做一个综述性的回顾尤其困难。这个时候,迫切需要具有更大客观性、科学性、高效性的方法来研究科学学科的结构与发展。1955年加菲尔德(E. Garfield)在 *Science* 上发表关于引文索引的文献,奠定了引文分析的基础。这不仅推动了代表学术共同体的多学科数据库——SCI的发展,而且还为研究科学的动态发展状况设计了一系列成熟的概念性关注。引文分析概念成为当今科学计量学、文献计量学、信息计量学、网络计量学的基础。加菲尔德的发明极大地改变了科学计量学家们研究科学共同体的方式。经过多年发展,特别是美国信息研究所(ISI)提供的引文数据库使引文结构的大样本统计分析越来越便利,知识图谱已成为科学共同体结构与发展实证研究的主流方法,广泛用于很多学科领域。

(二)知识图谱的发展

知识图谱的发展经历了三个阶段:

第一阶段,引文分析技术的出现。1999年斯莫尔(H. Small)明确提出借助引文图谱实现科学可视化途径。从普赖斯(D. Price)、加菲尔德到斯莫尔,已确立起日臻完备的引文分析理论与方法,构成科学计量学的基础与主流,在一定意义上形成了科学计量学中一门成熟的分支学科——引文分析学。20世纪90年代以来,科学计量学运用统计分析、引文分析和网络分析的方法,以及计算机图形学、图像处理与可视化技术,在科学知识图谱和知识可视化方面得到了迅猛的发展。

第二阶段,社会网络分析技术阶段。在引文网络研究中,引入复杂网络和社会网络的基本概念与最新成果,把引文分析、复杂网络和社会网络三种理论与方法统一起来,将科学知识图谱理论与方法提高到一个

新的水平。这种变化不仅可以对引文网络知识分布、知识流动、知识演化等特有规律产生深化认识,而且还可以促进探索普遍存在于自然、社会和人文的复杂网络的一般规律,具有重大的学术价值。

第三阶段,可视化知识图谱阶段。1987年,美国国家科学基金会发表《科学计算中的可视化》,标志着科学可视化的诞生。信息可视化(information visualization, InfoVis)最早由罗伯逊(G. Robertson)等在1989年提出,指在计算机、网络通信技术支持下,以认知为目的,对非空间的、非数值型的和高维信息进行交互式视觉表现的理论、方法与技术。计算机可视化信息处理软件,是通过直观的动态图像信息处理的方式,显示出专业领域中出现的交叉学科的复杂现象,从而获得详尽的前沿科学信息分析结果,它不仅有助于科学家在最短的时间里了解和预测前沿科技研究动态,而且还有助于在复杂的科研信息中开辟新的未知领域,提供快速独立科学判断的客观依据。2003年,美国科学院组织的“Mapping Knowledge Domains”讨论会预示着世界科学计量学中知识图谱和可视化研究的春天的到来。我国大连理工大学的刘则渊教授以此次会议为契机,展开了对知识图谱的研究。他于2005年在国内提出知识图谱研究,于2008年出版了《科学知识图谱:方法与应用》一书。此后,知识图谱的应用研究在国内不断涌现,取得了丰硕的成果。

第二节 知识图谱的原理及特征

一、知识图谱的原理及种类

(一)知识图谱的原理

知识图谱的基本原理是科学文献、科学家、关键词等分析单位的相似性分析及测度。根据不同的方法和技术可以绘制不同类型的科学知识图谱。该方法首先,通过计算机和互联网搜索引擎强大的自动查询功能,在极短的时间里面完成对海量信息的准确查询。其次,通过计算机对已查询到的海量零散信息进行文献计量统计分析,不仅可以通过量化

模型将其以科学的、可视化的形式直观地呈现出来,而且还可以发现它们之间的深层次关系和趋势,为今后在该领域的研究提供更有力的客观数据和科学支持。^① 其主要的绘制流程见图 1-2。

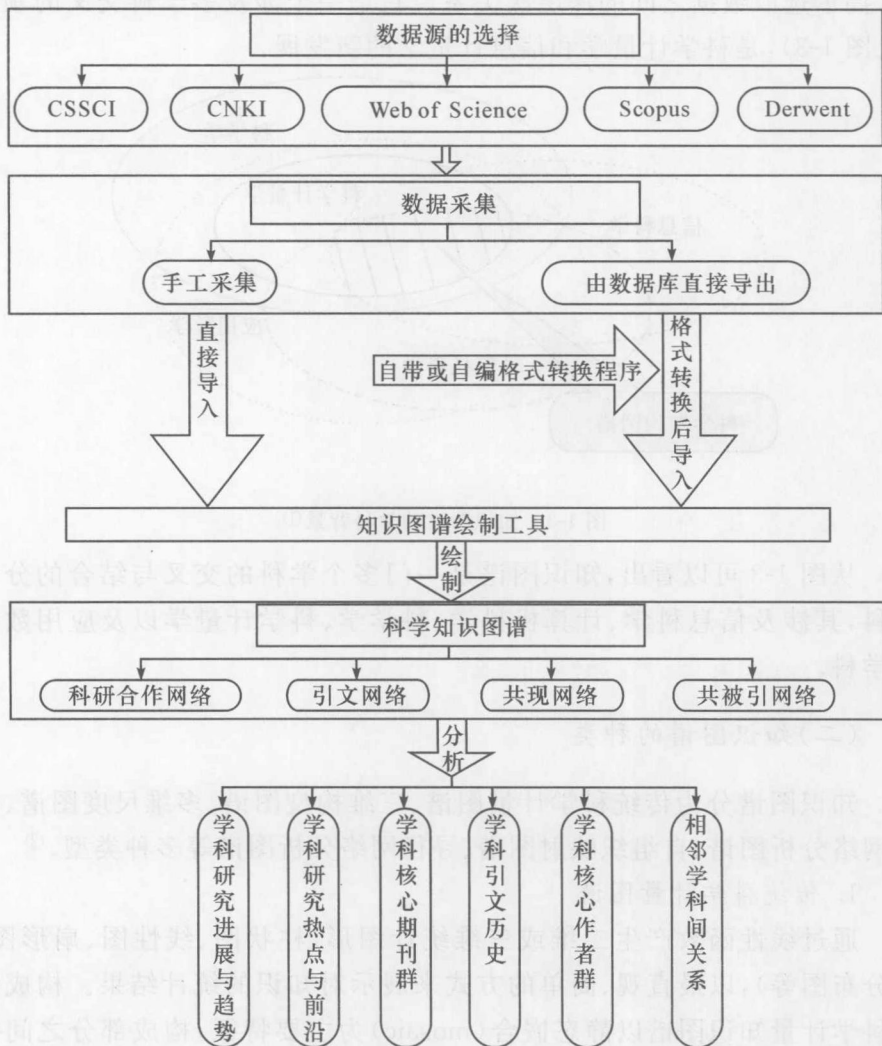


图 1-2 知识图谱的绘制流程^②

① 郭文斌、高峻峰:《国内心理学界记忆研究热点——基于 2003—2013 年 900 余篇的关键词共词分析》,《渭南师范学院学报》2014 年第 7 期,第 37—43 页。

② 胡泽文、孙建军、武夷山:《国内知识图谱应用研究综述》,《图书情报工作》2013 年第 3 期,第 131—137 页。