



BOSHI WENKU

〔管理学〕

知识图谱构建的 理论与实践

ZHISHI TUPU GOUJIAN DE
LILUN YU SHIJIAN

秦长江 著

知识产权出版社



BOSHI WENKU
〔管理学〕

知识图谱构建的 理论与实践

ZHISHI TUPU GOUJIAN DE
LILUN YU SHIJIAN

秦长江 著

知识产权出版社

内容提要

本书系统总结了知识图谱的理论基础与技术方法,明确了知识图谱与知识地图、信息可视化、知识管理之间的关系。采用科学计量学中的共现分析方法,利用聚类分析、因子分析、多维尺度分析以及社会网络分析等技术手段,对中国农史学科不同时期的研究热点和主要研究领域、学术群体和学科结构、与相邻学科关系等方面进行全方位的学科知识图谱构建。将实证分析知识图谱方法运用到我国人文学科发展历史研究中的适用性和科学性,在研究方法和研究对象方面有创新性。

责任编辑:王金之

责任校对:韩秀天

封面设计:SUN 工作室

责任出版:卢运霞

图书在版编目(CIP)数据

知识图谱构建的理论与实践 / 秦长江著. —北京:知识产权出版社, 2010. 11

ISBN 978 - 7 - 5130 - 0252 - 3

I. ①知… II. ①秦… III. ①农业史 - 中国 - 图谱 IV. ①F329 - 64

中国版本图书馆 CIP 数据核字 (2010) 第 214726 号

知识图谱构建的理论与实践

秦长江 著

出版发行: 知识产权出版社

社 址: 北京市海淀区马甸南村 1 号

邮 编: 100088

网 址: <http://www.ipph.cn>

邮 箱: bjb@cnipr.com

发行电话: 010 - 82000860 转 8101/8102

传 真: 010 - 82005070/82000893

责编电话: 010 - 82000860 转 8112

责编邮箱: wangjinzhi@cnipr.com

印 刷: 北京中献拓方科技发展有限公司

经 销: 新华书店及相关销售网点

开 本: 880mm × 1230mm 1/32

印 张: 7.125

版 次: 2010 年 11 月第 1 版

印 次: 2011 年 3 月第 2 次印刷

字 数: 196 千字

定 价: 50.00 元

ISBN 978 - 7 - 5130 - 0252 - 3/F · 379 (10330)

出版权专有 侵权必究

如有印装质量问题,本社负责调换。

摘 要

计量历史研究是指运用数理统计学的理论和方法来描述、分析、解释历史现象。它作为一种研究方法被引进历史学研究之后，特别是当代计算机信息技术的应用和普及，使之迅速进入众多史学研究领域，成为当代史学研究方法的重要时代特征。科技史作为跨越科学学和历史学两大领域的交叉学科，定量化研究也发展很快，作为定量化研究科技史的研究领域——科学计量学由此产生并发展壮大，它作为科技史和科学学重要的、特殊的分支，现已成为研究科技史的重要工具和手段。

知识图谱这一科学计量学领域的新技术，把科学计量学与应用数学、图形学、信息及计算机科学诸学科交叉结合，将相关学科知识领域的核心结构、学科前沿和新增长点以可视化的图像直观地表达出来，形成全新的知识图谱，以揭示学科领域的动态发展规律。它所应用的技术包括图论、共现分析、网络分析等。迄今为止，知识图谱已在发达国家实际应用并取得较好效果，我国也有了初步的研究和探索。

大量的文献调研表明，知识图谱这一先进的技术方法都是应用在我国自然科学、管理科学学科发展历史的研究中，且研究手段很落后。还没有在人文学科，特别是文史学科中进行研究的报道，这不能不说是我国人文学科研究的一大损失和缺憾。自然科学、技术与人文科学、社会科学既有共性的一面，亦有个性的一面。由于人文社会科学研究及其文献具有自然科学所没有的特点，对人文社会科学学科发展历史的研究，其研究方法更具复杂性和特殊性，其方法与自然科学的方法也有很大的不同。这个问题也是人文社科专家和管理部門一直关注的。

本书尝试将我国农业史学科作为研究对象，全方位地构建该学科的知识图谱，通过对构建农业史学科知识图谱的实证研究，探讨知识图谱这一先进的科学计量学研究方法运用到我国人文学科发展历史研究中的适用性和科学性。通过实证分析，探讨该方法在我国人文学科发展历史的研究中是否适用？科学性如何？有多大的误差？存在什么问题？产生的原因是什么？与以前知识图谱方法所选用的自然科学、管理科学学科的研究对象相比，有什么不同？等等。因此，本书选题的研究和实证分析，有以下较为重要的学术意义：

（1）将知识图谱的研究范围扩展到我国人文学科发展历史的研究中，对我国科学计量学和知识图谱的深入研究将有较为重要的借鉴意义和作用。

（2）对我国的人文社会科学研究也有一定的学术价值和实际应用意义。知识图谱在一定程度上可以使学科发展历史的研究更加“精确”和“客观”。因为即使是资深的本学科专家，也不可避免地会受到主观判断的影响。知识图谱在一定意义上可以避免这种主观性造成的偏差，从而为展现我国人文社会科学学科的历史演进，揭示人文社会科学的发展态势和规律提供客观和有意义的参考。

（3）学科领域的发展研究也是科技史研究的一个重要方面。农业史（以下简称“农史”）学科覆盖范围较广，包括农业生产、农村经济、农民生活等方面的历史内容，是一个横跨多种学科的专题研究领域，内容涉及农业科学、经济、历史等多个相关学科。本书构建的中国农史学科知识图谱以独特的视角较为全面地、形象地展现了我国农业史学科的发展。因此，本书的研究对我国农业史学科的发展也具有一定的现实意义。

（4）对我国科技史的研究也有一定的促进作用。知识图谱在一定程度上可以弥补科学计量学、科技史研究人员在专业素质上的欠缺。知识图谱可以帮助研究人员对相关专业的历史有大致地认识，使这些科技史研究人员较快地熟悉相关专业。

本书主要研究方法是：采用科学计量学、文献计量学中的共现分析法（主要包括：关键词共现分析法和作者、期刊共引分析法），分别利用聚类分析、因子分析、多维尺度分析以及社会网络分析等技术手段，对我国农史学科不同时期的研究热点和主要研究领域、学术群体和学科结构、与相邻学科关系等方面进行全方位的农史学科知识图谱构建。

具体来说，本书有以下主要研究内容：

（1）知识图谱的理论基础与技术方法。介绍了科学计量学的概念和发展；阐述了知识图谱的概念；论述了知识图谱的理论基础；分析了构建知识图谱的关键技术，重点对国内鲜有了解的几种软件进行了评价。

（2）构建中国农史学科知识图谱的学术意义和设计思想。基于知识图谱方法的先进性和知识图谱在我国研究的现状和不足，从这两方面论述了本研究的迫切性；又从四个方面说明了本书所研究主题的学术意义和学术价值；结合农史学科的特性和共词分析、共引分析的特点和国内外研究的状况，阐述了本研究的设计思想。

（3）基于关键词共现分析的我国农史学科研究热点和主要研究领域知识图谱。建立中国农史学科论文关键词数据库，分别对1980~1995年和1996~2008年两个时期的农史文献进行共词分析，利用聚类分析以及社会网络分析方法和软件，绘制两个时期的战略坐标图、类团关系图，构建农史学科研究热点和主要研究领域演化发展知识图谱，分析不同时期我国农史学科的研究热点以及研究领域之间具有怎样的交叉、渗透、转移。

（4）基于作者共引分析的我国农史学科学术群体和学科结构知识图谱。建立中国农史学科引文数据库，选定共引作者的具体范围，采用作者共引分析方法，分别利用因子分析、聚类分析、多维尺度分析以及社会网络分析方法和软件，构建农史学科学术群体和学科结构知识图谱，分析农史学科最主要的研究分支以及每个分支代表性的学者群体；同时对采用的两种技术方法所构建的知识图谱

的差异和效果进行对比分析。

(5) 基于期刊共引分析的我国农史学科与相邻学科关系知识图谱。选定共引期刊的具体范围,利用社会网络分析方法和软件,构建农史学科与相邻学科关系知识图谱,分析农史学科与相邻学科的关系,考察我国农史学科的影响力和渗透力,对促进学科的发展及与其他学科的相互交融提供有意义的参考。

本书的创新点主要体现在以下三个方面:

(1) 研究对象。本书把研究对象选定为文史特征很强的学科——农史学科,把知识图谱方法应用到我国人文学科发展历史的研究中,从实证角度证明知识图谱这一科学计量学方法在人文学科发展历史研究中的有效性,是科技史研究方法的一次有意义的创新尝试。因为传统、成熟的历史学科,其自身的学科特点和文献特征很强,文献与引文中古籍、灰色文献种类繁多、数量巨大,文献半衰期较长,与很活跃的自然科学、管理学科领域的研究相比,数据来源的选取以及分析有很大难度。同时尝试将关键词、作者、期刊、论文一起作为研究对象,对国内一具体学科领域进行全方位的知识图谱构建,难度更大。因此,本书对国内科学计量学和知识图谱的研究,是一次极有益的探索。

(2) 研究方法。本书尝试采用社会网络分析方法和软件 Pajek,与传统的多维尺度分析、聚类分析、因子分析相结合进行对比研究,对由这两种不同方法获得的知识图谱进行了对比分析,从实证的角度说明这两种方法的各自特点,为以后的知识图谱分析和研究提供了借鉴;同时根据共词分析和共引分析的不同特点,尝试将这两种方法结合,进行知识图谱构建;还首次尝试采用社会网络分析软件 Pajek 绘制类团关系图。

(3) 系统总结了知识图谱的理论与技术方法。明确了知识图谱与知识地图、信息可视化、知识管理之间的关系,重点对国内鲜有了解的几种软件进行评价,为国内开展知识图谱研究提供理论和实践参考。

ABSTRACT

This approach, which uses the mathematical statistical theory and methods to describe, analyze, interpret the phenomenon of history, is called cliometrics. As a research method, after it is introduced to the study of history, especially with the application and popularization of modern computer information technology, making it quickly access to many areas of historical research, it becomes an important feature of time in historical research methods. History of Science and Technology as a interdisciplinary subject in Scienology and History, at the same time, with the speed development of its quantitative studies, scientometrics borned as a new research field. Now, it has become an important research tool as a branch of History of Science and Technology and Scienology.

Knowledge Domains map, the new technology on the field of Scientometrics, combines Scientometrics with applied mathematics、graphics、information science、computer science subjects and so on, and expresses the core structure, the latest development and new research direction in the related knowledge field by visual image. This brand-new Knowledge Domains Map can reveal the dynamic development law. It adopts such technologies as graph theory, network analysis and co-occurrence analysis. So far, Knowledge Domains Map has been applied in many developed countries and made good achievements. It has got a preliminary study and exploration in our country.

A large number of documents show that the latest technological methods are used in Natural Sciences and Management Sciences, and not



applied in The Humanities, especially domestic Literature and History subject. Therefore, it is a great loss and defects in the study of The Humanities in our country. Natural Sciences, Technical Sciences and The Humanities, Social Sciences have the general and specific character. Because of The Humanities and Social Sciences and their literatures do not have the characteristics of the Natural Sciences, the study methods on the development history of The Humanities and Social Sciences are more complexity and specificity, its methods are quite different from the Natural Sciences. The experts of The Humanities and Social Sciences have been concerning about the issues.

Therefore, This paper tries to introduce the study methods of knowledge Domains map to the study of the development history of The Humanities, it has much more important academic value. This paper attempts to build a comprehensive knowledge Domains map of Chinese Agricultural History subject, and to examine the applicability and scientific of the advanced research methods used in the study of the development history of The Humanities by empirical study. Through empirical analysis, it discusses whether the methods in the study of the development history of The Humanities are applicable and Science? How much error? What's the problems? What's the reason? What's the difference from the former chosen as Natural Sciences of research object? And so on. Therefore, the research and empirical analysis of this paper will be more important academic significance:

(1) The study of knowledge Domains map expands to the development history of The Humanities, which will have an important influence and reference on the intensive study of Chinese Scientometrics and knowledge Domains map.

(2) It also has a certain academic value and practical application significance to the study of The Humanities and Social Sciences in China.



To some extent, knowledge Domains map can make the study of the development history of The Humanities become more “scientific”, “precision” and “objective”. Because even Senior Experts in this subject, they are inevitably affected by their own subjectivity. In a certain sense, knowledge Domains map can avoid bias caused by subjectivity so as to show the development history of The Humanities and Social Sciences in China, and to reveal the historical evolution and development trend of The Humanities and Social Sciences, and to provide objective and meaningful reference.

(3) The research of the development of discipline is an important aspect in the study of History of Science and Technology, Agricultural History is a research field which covers many subjects relatively, including the historical contents of agricultural production, countryside economy, peasant life, and so on. It refers to many related subjects, such as agricultural science, history, economies, ect. The Knowledge Domains map of Chinese Agricultural History subject all-roundly, vividly shows the development history of Chinese Agricultural History subject with unique Angle of view, meanwhile it also has certain practical significance to the development of Chinese Agricultural History subject.

(4) It also has certain stimulative effect to the study of Chinese History of Science and Technology. To some extent, knowledge Domains map can compensate for professional quality that researchers on Scientometrics and History of Science and Technology are lack of. It can help researchers to the relevant professional development history have roughly understanding, and be master of relevant profession rapidly.

The main research methods in this thesis are: keywords co-occurrence analysis, co-citation analysis for author and journal in scientometrics and bibliometric, and the traditional methods such as cluster analysis, factor analysis, multidimensional scaling analysis and social network



analysis on the main research direction and main field, academic schools and structure and the relations with other related subjects in the different periods of Chinese Agricultural History subject to build the all-round Knowledge Map.

Specifically, the main research contents are as follows:

(1) The theoretical basis and technological methods of Knowledge Domains Map. This paper introduces the definition and development of Scientometrics; it elaborates the definition of Knowledge Domains Map; it discusses the theory basis of Knowledge Domains Map; it analyzes the main technology of building Knowledge Domains Map; the introduction and comparison of several softwares that are stressed on abroad but known little in our country.

(2) The academic significance and design ideas to build knowledge Domains maps of Chinese Agricultural History subject. Based on the advancing front method of knowledge Domains map and its present research level and deficiency, the paper dissertates the necessity and urgency of the study from two aspects; it illustrates the significance of academic research and academic values from four facets; at the same time, it sets forth the design ideas of this study combining the characteristics of Chinese Agricultural History subject with the features of co-word and co-cited analysis.

(3) The hotspots of Chinese agricultural history subject and its' knowledge Domains map of the main research expertise which base on keywords co-occurrence analysis. Set up the co-word database of Chinese agricultural history. In this paper, we carry out co-word analysis of Chinese agricultural history literature for 1980 ~ 1995 and 1996 ~ 2008 separately, and draw coordinates figures and the relationship of clusters via traditional cluster analysis and advanced social network analysis methods and software to build the hotspots of Chinese agricultural history subject



and its' evolutional knowledge Domains map of the main research expertise, and to analyze the hotspots of Chinese agricultural history subject and the main research expertise how to intersect、infiltration and transfer in different period.

(4) The academic schools and knowledge Domains map of the subject structure which base on author co-citation analysis. Build Chinese agricultural history disciplines Citation Database in order to choose its' specific author scope and number. In this paper, we mainly use author co-citation analysis to build academic schools and knowledge Domains map of the subject structure , and to analyze the most important branch of the subject and the most influential scholars and groups of each branch via the traditional factor analysis, Cluster analysis, multidimensional scaling analysis, as well as advanced social network analysis methods and software. At the same time, analyze the causes and effects on the differences of knowledge Domains map depending on the two techniques.

(5) The knowledge Domains map of Chinese agricultural history subject and its neighboring disciplines based on journal co-citation analysis. Choose its' specific journal scope and number. In this paper, we mainly use advanced social network analysis methods and software to build the knowledge Domains map of Chinese agricultural history subject and its neighboring disciplines, and to analyze the relationship between Chinese agricultural history and its neighboring disciplines, and to study its' influence and penetration, so that it can provide meaningful reference to promote the development and circuminsession of its own and other disciplines.

Finally, the paper gives a conclusion about the study, and expounds problems which must be dealt with urgently in the future.

The main contribution of this paper is as follows:

(1) Research object. It is the first time to try to put keywords, au-

thors, journals, thesis studied together as research object to build a comprehensive knowledge Domains map of Chinese agricultural history subject. This is the first attempt to build a full range of knowledge Domains map to domestic one specific subject areas. The research object is selected as the very traditional、sophisticated literature and history subject--agricultural history, therefore it is more difficult. Because it is quite difficult for the very traditional、sophisticated literature and history subject to choose suitable data to analyze, its literature and citation own strong features, including a wide range of gray literatures and large quantities of ancient literature, having long half-life, the difficulty of its data source compares with very active natural and management sciences which often are used as scientometrics and bibliometrics research object. Therefore, the study to the development of domestic scientometrics and knowledge Domains map is a useful exploration. This paper uses the methods of knowledge Domains map on the study of development history of The Humanities in our country, and it demonstrates the effectiveness in the study of development history of domestic The Humanities by knowledge Domains map--the method of Sientometrics, it is also a meaningful、innovative attempt in the study method of History of Science and Technology.

(2) Research methods. According to the different features between co-word analysis and co-citation analysis, it is the first time to combine these two approaches to build knowledge Domains map in our country. There is the only one in our country to combine advanced social network analysis method and software Pajek with the traditional multidimensional scaling analysis techniques, cluster analysis techniques, factor analysis techniques to compare and analyze. Through comparing and analyzing the differences of knowledge Domains maps via adopting the two different methods, it is proved the respective characteristics of the two methods from the empirical point of view and meaningful reference for the future



analysis and research. It is also the first time to adopt social network analysis software--Pajek, to draw the cluster relationship diagram.

(3) The theory and technical methods of knowledge Domains map are summarized systematically for the first time in our country; the paper knows the relationship between knowledge Domains map and information visualization、knowledge management、knowledge map; the introduction and comparison of several softwares that are stressed on abroad but known little in our country. So the paper provides a theoretical and practical reference for domestic knowledge Domains map research.

目 录

第一章 绪论	(1)
一、选题的背景和意义	(1)
二、本书的研究方法和主要内容	(6)
三、本书的结构和创新点	(7)
第二章 知识图谱的理论基础和技术方法	(10)
一、科学计量学的理论和发展概况	(10)
二、知识图谱的概念	(14)
三、知识图谱的理论基础	(18)
四、知识图谱的技术方法	(40)
第三章 构建中国农史学科知识图谱的学术意义和设计思想	(61)
一、本研究的迫切性	(61)
二、本研究的学术价值	(77)
三、本研究的设计思想	(84)
第四章 中国农史学科研究热点和主要研究领域知识图谱	(92)
一、实验设计	(92)
二、实验结果和分析	(110)
三、小结	(127)
第五章 中国农史学科学术群体和学科结构知识图谱	(129)
一、实验设计	(129)
二、实验结果和分析	(150)
三、小结	(164)
第六章 中国农史学科与相邻学科关系知识图谱	(166)
一、实验设计	(166)



二、实验结果和分析·····	(171)
三、小结·····	(174)
第七章 结语·····	(175)
附录·····	(178)
附录1 《中国农史》等三种期刊高被引的前100篇文献···	(178)
附录2 58位作者原始矩阵(部分)·····	(184)
附录3 86个关键词原始矩阵(部分)·····	(186)
附录4 113个关键词共词原始矩阵(部分)·····	(188)
附录5 113个高频关键词处理过的共词矩阵(部分)·····	(190)
参考文献·····	(193)
后记·····	(206)

图目录

图 2-1	文献耦合与文献共引对比	(27)
图 2-2	多维尺度图	(44)
图 2-3	科研合作网络图	(49)
图 2-4	期刊网络图	(50)
图 2-5	类团关系图	(52)
图 2-6	12 个类团战略坐标图	(54)
图 3-1	信息可视化参考模型	(62)
图 4-1	1980 ~ 1995 年聚类图	(113)
图 4-2	1996 ~ 2008 年聚类图	(114)
图 4-3	1980 ~ 1995 年 12 个类团的 Pajek 数据格式	(118)
图 4-4	1996 ~ 2008 年 15 个类团的 Pajek 数据格式	(119)
图 4-5	1980 ~ 1995 年 12 个类团关系图	(120)
图 4-6	1996 ~ 2008 年 15 个类团关系图	(121)
图 4-7	1980 ~ 1995 年 12 个类团战略坐标图	(123)
图 4-8	1996 ~ 2008 年 15 个类团战略坐标图	(125)
图 5-1	58 位作者的聚类图	(152)
图 5-2	58 位作者的 MDS 图	(157)
图 5-3	58 位作者的原始 Pajek 图	(160)
图 5-4	58 位作者处理后的 Pajek 图 1	(161)
图 5-5	58 位作者处理后的 Pajek 图 2	(162)
图 5-6	58 位作者处理后的 Pajek 图 3	(163)
图 6-1	原始数据生成的 Pajek 网络图	(171)
图 6-2	数据处理后生成的 Pajek 网络图	(172)