

980020

文献情报计量理论和方法

Theory and Method of information Measurement

孙建军 编著

564



南京大学出版社

文献情报计量理论和方法

孙建军 编著

南京大学出版社

1994·南京

(苏)新登字第 011 号

文献情报计量理论和方法

Wenxian Qingbao Jiliang Lilun He Fangfa

孙建军 编著

南京大学出版社出版

(南京大学校内 邮政编码:210008)

句容印刷厂印刷

*

开本 850×1168 1/32 印张 9.125 字数 240 千

1994 年 8 月第 1 版 1994 年 8 月第 1 次印刷

印数 1—1000

ISBN 7—305—02501—1/G·411

定价: 7.80 元

前 言

文献情报计量学是当今科学定量化趋势中发展起来的新学科。在许多领域,诸如科学学、未来学、情报学、图书馆学、信息论以及经济管理方面都有广泛的应用,因此受到情报工作者的广泛注意,愈来愈多的情报工作者要求学习和掌握文献情报计量的基本理论和方法,并解决实际工作中的问题。

本书力求用深入浅出的方法系统地论述文献情报计量理论和方法。全书共9章,大致可以分成三大部分。第一部分介绍了文献情报计量的发展历史、特点及其数学基础;第二部分研究了文献规律,包括文献情报随时间演变规律、文献集中与分散规律、文献作者分布规律、词频分布规律;第三部分论述了文献情报计量规律的理论基础及计量方法。

本书在编著和出版过程中得到了南京大学文献情报学系和南京大学出版社的支持。书中引用了情报工作者的一些研究成果,邹志仁先生审阅了书稿,出版社徐雁副编审为本书的出版付出了劳动,在此谨表谢意。

由于作者水平有限,书中如有疏漏之处,敬请广大读者指正。

孙建军

1993年9月20日

目 录

第一章 绪论

- 第一节 情报学的理论研究..... (1)
- 第二节 情报学研究方法..... (7)
- 第三节 文献计量学的发展历史 (11)
- 第四节 文献计量学的特点和发展趋势 (14)

第二章 文献情报计量的数学知识

- 第一节 集合论 (19)
- 第二节 矩阵及其运算 (24)
- 第三节 概率论基本知识 (32)
- 第四节 时间序列数据的获取和计算 (46)
- 第五节 回归分析简介 (62)

第三章 文献情报随时间演变规律的分析与预测

- 第一节 文献情报随时间演变的规律 (72)
- 第二节 文献情报演变规律的机理分析 (87)
- 第三节 文献出版的定量预测 (94)

第四章 文献集中与分散理论

- 第一节 文献集中与分散规律..... (108)
- 第二节 布拉德福定律的基本内容..... (111)
- 第三节 布拉德福定律的发展..... (117)
- 第四节 布拉德福定律的应用..... (130)

第五章 文献作者分布理论

- 第一节 科学生产率与洛特卡定律····· (136)
- 第二节 普赖斯定律····· (142)
- 第三节 洛特卡定律的验证····· (146)

第六章 词频分布理论

- 第一节 词频分布理论的产生····· (149)
- 第二节 齐普夫定律的发展····· (156)
- 第三节 词频分布理论的应用····· (159)

第七章 引文分析研究

- 第一节 概述····· (162)
- 第二节 引文分析的测度指标····· (166)
- 第三节 引文分析的主要工具····· (171)
- 第四节 科学引文的聚类分析····· (175)
- 第五节 引文分析法的应用····· (190)

第八章 文献计量学的理论基础

- 第一节 普赖斯模型····· (197)
- 第二节 伯利亚分布的单缸模型····· (202)
- 第三节 布-洛-齐分布系····· (205)
- 第四节 文献交流过程中的两种效应····· (219)

第九章 文献情报计量方法

- 第一节 概述····· (226)
- 第二节 运筹学与情报计量研究····· (234)
- 第三节 信息论方法····· (249)
- 第四节 应用举例····· (250)

CONTENTS

Chapter 1. Introduction

- 1.1 Theoretical Research of Information Science (1)
- 1.2 Research Methods of Information Science (7)
- 1.3 History of Development of Bibliometrics (11)
- 1.4 Characteristics and Trends of Bibliometrics (14)

Chapter 2. Mathematics Basis of Documents and Information Measurement

- 2.1 Set Theory (19)
- 2.2 Matrix and Matrix Operations (24)
- 2.3 Elementary Knowledge of Probability (32)
- 2.4 Acquisition and Calculus of Time Series Data ... (46)
- 2.5 Brief Description of Regression Analysis (62)

Chapter 3. Time Series Analysis and Predication for the Evolution Rule of Information

- 3.1 Time Series Evolution Rule of Documents and Information (72)
- 3.2 Mechanism Analysis for the Evolution Rule of Documents and Information (87)
- 3.3 Quantitative Predication of Published Literature (94)

Chapter 4. Theory of Documentary Centralization and Dispersion

4.1 Rule of Documentary Centralization and Dispersion	(108)
4.2 Outlines of Bradford's Law	(111)
4.3 Development of Bradford's Law	(117)
4.4 Application of Bradford's Law	(130)

Chapter 5. Theory of Author Distribution

5.1 Scientific Productivity and Lotka's Law	(136)
5.2 Price's Law	(142)
5.3 Verification of Lotka's Law	(146)

Chapter 6. Theory of Words Frequency Distribution

6.1 Emerge of Words Frequency Theory	(149)
6.2 Developments of Zipf's Law	(156)
6.3 Applications of Words Frequency Theory	(159)

Chapter 7. Research of Citation Analysis

7.1 Brief Description	(162)
7.2 Measuring Target of Citation Analysis	(166)
7.3 Basic Means of Citation Analysis	(171)
7.4 Clustering Analysis of Scientific Citation	(175)
7.5 Application of Citation Analysis Methods	(190)

Chapter 8. Theoretical Basics of Bibliometrics

8.1 Price Model	(197)
8.2 Single Vat Model of Polya Distribution	(202)
8.3 Bradford—Lotka—Zipf Distribution	(205)

8.4 Two Effects in the Process of Document Communication (219)

Chapter 9. Measuring Methods of Documents and Information

9.1 Brief Description (226)

9.2 Operational Research and Measuring Research of Information (234)

9.3 Methods of Information Theory (249)

9.4 Cases for Application (250)

References

第一章 绪 论

第一节 情报学的理论研究

随着人类社会的发展、信息社会的迫切需要,情报交流日益广泛和复杂,情报工作日益专业化,并且情报已经成为社会发展不可缺少的一种因素。情报业的兴起和发展,从事情报业人员的不断增多,以及情报业在国民经济总产值中所占比例的日益增多,不仅表明了情报对社会发展的重要性,而且也充分说明情报是现代人类社会的一种普遍需要,是普遍存在着的一种社会现象。

一、情报学的研究内容

情报学自从 50 年代诞生以来得到了迅速发展,得到了社会的广泛承认。由于学科历史短,而且研究的角度不同,所以对情报学的理解有不少差异,给情报学提出了很多有不同特色和原则区别的定义,产生了以布鲁克斯为代表的知识学派,以萨瑞塞维克、米哈依洛夫、贝尔金等为代表的科学交流学派。

布鲁克斯(*B. C. Brooks*)认为,波普的三个世界理论能够成为情报学的理论基础。情报学要研究客观知识的分析、组织和利用,研究主观知识和客观知识之间的相互作用,描述和解释它们,以有助于对知识进行组织,从而更有效地加以利用。1967 年,布鲁克斯曾为情报与知识的关系建立了一个基本方程,并且表明知识是通过情报的获得而增加的,而所获得情报量的大小又取决于情报接收者原有的知识。因此,可以概括地说,布鲁克斯认为情报学的任务是探索和组织客观世界。

美国情报学家萨瑞塞维克(*T. Saracevic*)认为,情报学不同于文献学、图书馆学等学科,它有着自己独特的研究领域,是专门研

究人类通讯现象和通讯系统特征的科学。他把情报学分为两大分支，理论情报学与应用情报学。理论情报学是研究情报的特性、情报用户及利用、情报系统的情报过程及评价、情报系统环境的各类问题；应用情报学研究知识交流的技术问题、情报的表达、情报系统以及情报产业的发展等。

前苏联情报学家米哈依洛夫(А. И. Михайлов)认为，情报学是属于社会科学的范畴，是研究科学情报的构成及共同特性，以及研究科学交流全过程的规律性，研究内容包括情报学的理论、科学交流、文献加工和情报检索、科学情报工作的规律、理论、方法、组织和历史、情报用户需求的研究、情报活动的效率、计算机在情报活动中的应用、情报学的语言问题和心理学问题等。

英国情报学家贝尔金(N. J. Belkin)认为情报概念应从三个方面来考虑。一是定义方面应注重情报学所要解决的问题。二是应该观察到用户在情报方面的行为方面具有以下特点：不同的用户会对同一组数据做出不同的响应；同一用户在不同的时间对同一数据也会做出不同的响应；用两种不同的语言表述同一文献内容，会对情报接受产生不同的影响；具有不同知识状态的情报生产者会给人们对文献信任程度产生不同的影响。三是情报研究的方法论必须与情报概念的如何使用以及使用情报后的效果有关。贝尔金通过对定义、行为和方法论三个方面的探讨研究，提出了应从以下八点来分析情报研究：

第一，情报研究必须指的是在有目的、有意义的交流环境中的情报；

第二，应该说明情报是在人类社会交流过程中的；

第三，应该说明情报是接受者所希望得到的；

第四，应该说明情报是对接受者的效果；

第五，应该说明情报是和产生者及接受者知识状态之间的关系；

第六，应该说明各种不同形式表示的内容效果不同；

第七,应该说明情报包括每个个体,具有普遍意义;

第八,应该说明提供预测情报效果的手段。

二、情报学体系结构

众所周知,对于任何一个学科的体系结构,不能采取绝对的态度,要想构造一个永恒的、无所不包的体系结构是根本办不到的。情报学是以作为一种普遍存在着的社会现象的情报和整个情报交流活动为研究对象的,专门研究情报工作的理论和方法的学问,是随着现代科学技术的发展而不断变革和发展。因此,任何一个情报学体系结构都是暂时的。然而,尽管体系结构是暂时的,却仍然是必要的。因为零散的知识不能形成为一门科学,所以,研究情报学的理论就不能回避体系结构。

1. 情报学体系结构的回顾

关于情报学体系结构的探讨,并不是一个新的课题。一些情报学家在本世纪六、七十年代发表的论著中,就已经从不同的角度涉及到情报学研究的范围及其分支学科。例如,美国西部预科大学图书馆学校的 *Rees* 和 *Saracenic* 在 1967 年专科图书馆协会的会议上提出:情报科学是从研究情报的特征、活动状态和情报流的各学科吸取原理、方法和技术的—个研究领域。它包括系统分析、情报和通讯的环境问题、情报载体、语言分析、情报组织、人与系统的关系等。另外,情报学理论体系的各家学说数量众多。由于受到时代的局限,他们当时并没有明确提出情报学体系结构这一概念。随着情报学研究领域的扩展,尤其是自然科学的一些交叉学科的研究成果,使情报学产生了许多新的分支学科,而且体系结构也成为多维的多层次结构。但从目前情报学发展的角度来看,就会发现过去的体系结构有严重的缺陷。

第一,现行的情报学体系结构基本都是块状结构,缺乏内在的有机联系,只是外在地把一些原理形式地拼凑在一起,形成了“理论情报学与应用情报学”或“理论情报学、技术情报学和应用情报

学”的体系结构,这些块状结构没有系统的相关性和结构上的过渡性。

第二,从社会制度和意识形态角度来判断可以看出,目前情报学体系结构没有完整、准确和全面地反映出许多情报学的理论和原理。例如,米哈依诺夫的情报学理论体系仍然是建立在文献层次而不是知识层次或情报层次上;布鲁克斯情报学理论体系未给出其情报理论与具体情报工作关系的一般性分析和方法。

第三,现存的体系结构所表现的内容陈旧,不能将新的成果、新的经验总结概括出来纳入体系结构中,这样难以发挥情报的作用。

第四,现行的情报学体系结构元素在内容上前后重复太多,给教学和自学都带来不少问题,这种现象的出现,是由于体系结构不够严谨、合理造成的。

由此可见,现存的体系中存在着许多问题,是不能依赖于仅仅从一些细小的环节和方面加以修改和补充所能解决得了的,而是必须改革现存的体系结构,建立一个紧凑的、功能明确的情报学体系结构,使情报学理论获得新的生机和活力。

2. 建立原则

任何系统包括两个以上,以致许多个相互作用和相互联系的元素。系统内部的这些元素之间有着自己特有的相互结合方式。尽管元素相同,但结合的形式不同,也会出现质的差别。例如,用同样的电子元件按不同的线路图来装配电视机,所装配出电视机的质量显然是不同的。在系统内部各部分之间存在着错综复杂的相互影响,存在着信息流、能量流和物质流的交换,这种关系不仅是多元的,而且往往是多层次的交叉和反馈。因此,我们在建立学科的科学体系结构时,必须要有一致的原则。情报学体系结构的改革必须服从于更好地体现情报学的学科内容,这是总的原则。下面几点当然是不言而喻,必须做到的。

第一,必须充分估计学科不断分化以及由此导致的“知识总

汇”必然走向瓦解的历史事实和发展趋势。

科学的发展总是一方面在走向分化,另一方面又趋于统一。随着不断的分化,统一的形式也要不断地更新。这个规律同样支配着情报学与科学之间的关系变化,适应科学不断分化的趋势。情报学应当向“一体化”的方向发展。我们在处理体系结构问题上,必须考虑到这一发展趋向。所谓“一体化”类似说不能把内容划分成不同的问题。这里主要是说在区分问题时,必须以体现结构的完整性、观点的统一性为前提。

第二,必须满足体系结构的基本特点。体系结构的基本特点包括下列要素:一是具有物质成份,这是形成体系结构的首要的基本的条件,没有它,根本谈不上体系结构问题;二是物质成份间的相互作用,这是形成体系结构的最重要条件。只有物质成份,没有相互作用,物质成份就成了杂乱无章的偶然堆积和彼此孤立、静止地存在着,就无所谓体系结构;三是运动和演化。物质成份的相互作用决定了体系结构的运动和演化形成,从一种体系结构形态转化为另一种体系结构形态,例如从较低级的系统结构发展到较高级的系统结构。

第三,必须充分考虑到以情报和整个情报交流活动为核心的系统性、层次性原则。所谓层次,不仅仅是一个系统中包含的有差别的等级,而是既有差别又有内在联系的等级。没有差别固然无所谓层次,但如没有联系,仍然不成其为层次。所以,情报学的体系结构应是对立的相互包含或相互渗透。只有这样,它才是有层次的、才有真正的结构。

第四,必须执行科学性和实践性相统一的原则。一种情报学的体系,是一种情报学说的理论表达形式。这里,形式决不是无关紧要的东西,而是形式和内容统一起来表现情报学理论的性质。情报学是一门理论和实践很密切的学科,因此,在建立情报学体系结构时必须贯彻理论和实践相统一的原则。

3. 基本体系结构

结构是某一系统中各种元素的相互联系和相互关系的方式。要研究情报学的体系结构,必须先认清构成对象的诸元素,进而研究元素之间如何联结成具有内在联系的结构系统,这样才能揭示情报学的全貌及其发展过程。因此,我们必须首先考虑情报学的基本元素。

一般来讲,情报学的基本元素是由情报学的研究对象所决定的,包括基本概念、规律和原理等。例如,“文献计量学”的基本元素是由布拉德福定律、洛特卡定律、齐普夫定律、文献增长和老化规律、文献的引用规律、自引、同被引等等。情报学的基本元素是十分广泛和丰富的,在其体系结构中占有特定的地位。

情报学的体系结构就是由这些众多的基本元素以自身特定的运动规律和联系方式有机地结合在一起,形成了情报学的基本体系结构。随着在发展过程中与相关学科(往往涉及心理学、社会学、思维科学、计算机科学、人工智能、数学、统计学、经济学、哲学、逻辑学、行为科学、图书馆学、目录学、文化史、科学史以及关于知识的理论等等)的联系产生了一些新的情报学学科。这样就完整地组织起了情报学的体系结构。这些运动变化过程用数学集合论表示为:

设情报学基本元素的集合为 A , 即 $A = \{a_1, a_2, \dots, a_m\}$; B 为对集合 A 中的元素简洁而有效结合的表达, 变成新的集合 B ; $B = \{b_1, b_2, \dots, b_n\}$; C 是集合 B 经过和相关学科 M 结合而得到新的集合 C , 其变化过程如下:

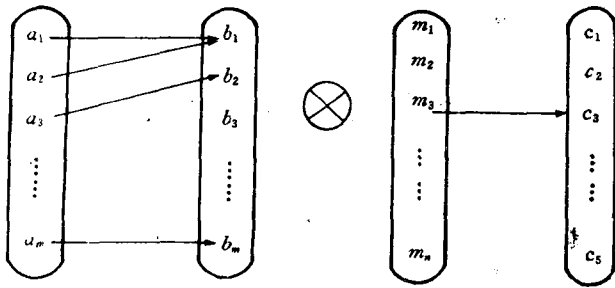


图 1-1

第二节 情报学研究方法

情报学方法论是情报学理论体系的有机组成部分,是情报学研究的重要课题。由于情报学本身学科的特殊性使得情报学从一诞生起就不断地重视方法论研究。

一、情报学方法论研究的意义

任何科学研究活动都离不开研究方法的运用,正如著名生理学家巴甫洛夫指出的:“科学是随着研究法所获得的成就而前进的,研究法每前进一步,我们就更提高一步,随之在我们面前也就开拓了一个充满着种种新鲜事物的、更辽阔的远景,因此,我们头等重要的任务乃是制定研究法。”作为表现多学科、多目标的综合统一的情报学研究也毫不例外。

情报学方法论研究的意义表现在以下三点:

第一,情报学要有所突破、有所发展,它自身的首要内在条件正在于方法论的革新。因而在被证实的科学基础上建立和运用情报学方法论,将研究方法抽象化、理想化纳入科学方法论的体系之中,以作为系统的研究对象,只有这样,情报学方法论才能成为情报学理论研究的基础。

第二,用科学方法论思想作指导,情报学理论的建立和完善不能是主观的臆造,从反映论的原则上看,情报学必须从自身的联系中形成情报学对象自身的结构,情报学理论应该反映客观对象的发展规律,反映人的认识由低级到高级、由简单到复杂的发展规律。建立理论时要遵循逻辑与历史相统一的原则,即根据概念、判断、推理之间的逻辑联系,按照客观对象在发展过程中所经历的不同具体阶段、具体形态和过程,将两者结合起来反映对象本质和规律。因此,研究情报学方法论,寻求情报学特征方法的社会意义就是为了从追求理论的普遍适用性方面将学科体系纳入到整个科学知识背景中,为解决具体研究课题或研究过程中的科学方法问题起到指导作用,即如何根据情报研究对象的性质和特点,选择、引鉴和创造相应的科学方法。按照性质研究本质特征及其矛盾的概念范围,我们从情报的概念分析可以表明对情报学的认识:情报学是研究社会情报现象,即根据用户需要选择、传递信息的科学,研究选择传递信息的一般原理、规律和方式方法的科学。这样限定的意义和范围体现了本学科的特殊性,可能更符合社会实际情况。

第三,通过对情报学方法论的研究,可以从整体上认识情报学体系及其辩证关系,提供正确的思维方式和思维形式,并且利用科学的思维方法、严密的逻辑方法对情报学理论加以论证。利用方法论观点还能正确引导人们自觉培养辩证的思维能力,以唯物辩证法的规律去正确地进行情报预测。

二、情报学研究方法的内容和特点

一般来说,任何一门学科的研究方法按照其抽象化程度和适用程度,可分为三个层次:哲学方法、一般方法和特殊方法。

1. 哲学方法

哲学方法是研究一切科学最普遍的方法。对于情报学来说,当然也应当应用哲学方法来进行研究。马克思主义哲学所创立的从生动的直观过渡到抽象的思维和从抽象的思维过渡到实践的辩证