

信息学基础理论

符福垣等 编著

科学技术文献出版社



前 言

信息学是我国学术界正在热烈讨论的一门新兴学科。为了适应社会主义市场经济迅速发展的新形势，国家科委于1992年9月决定用“科技信息”取代“科技情报”，并将国家科委科技情报司更名为国家科委科技信息司。这对于科技信息界来说，是具有非常重大而深远意义的事件。当然，这不只是名称上的改变，更重要的是工作内涵与外延方面都将有重大的扩展，进一步推动我国科技信息服务业与产业的发展，标志着我国科技信息体制改革与信息事业的发展已进入一个崭新的历史时期。

“情报”或“信息”是人类社会和自然界中普遍存在的现象，是物质形态及其运动规律的体现，出现在自然、社会和人类思维活动之中。“情报”与“信息”国际上都通用“Information”这个词，多数国家译成“信息”，日本叫“情报”，其含义国际标准定为“被传递的知识或事实”，我国常称为“情况报导”，或译为“信息是指对消息接收者来说，预先不知道的报道”。其实，“情报”与“信息”是意思重合但叫法不同的两个词而已。我国之所以长期使用“情报”这一术语，也是顺应历史潮流之需要，在当时发挥了有效的作用。进入80年代以来，“信息”一词在我国已被广泛认同和使用，信息业迅猛发展，现又被列入第三产业中的一个先导行业而受到举国上下的重视。在这种情况下，如再用“情报”或叠用“情报信息”，已不适应迅速发展的新形势。因此，用“信息”一词取代“情报”是适时的，顺应了历史的潮流和国家产业发展政策，有利于改革开放，拓宽工作范围，打开新局面。无疑，这对科技信息事业的发展必将起到积极的促进作用。

既然国家已将科技情报更名为科技信息，作为学科建制“情报学”扩展为“信息学”也是很自然的，可以说是顺理成章。这种更名，不仅在其概念上与整个科技信息工作保持一致，与国际术语系统更为接近，以利于开展学术研究和学术交流活动，而且也有利于学科的自身建设，推动学科体系结构向着更加完整的方向发展。因为，学科的发展主要根据社会发展的需要而变化，社会经济与科技的发展必然要影响到学科的发展，学科发展的动力要靠行业的发展，事业不发展，学科建设也难以深化。而学科建设的深化与发展，又将影响和促进事业的发展，学科的建设与事业的发展不宜分开考虑，固守已有的情报学，对于进一步推动学科建设和信息事业的发展都颇为不利，所以，情报学扩展为信息学反映了历史发展的必然趋势。邓小平同志指出的“开发信息资源，服务四化建设”应作为信息工作的出发点和归宿。“情报学”更名为“信息学”后，既保持原有的研究内容、对象和范围，同时在其内涵与外延上都将随着信息工作的内容与服务领域的不断扩展而延伸。它将吸引更多的学者去研究信息学，探索信息学的基本理论、方法和技术问题，为信息学学科体系的完善与发展开拓更加广阔的前景。

信息学的产生、形成与发展，是现代科学技术，特别是高新技术迅速发展的产物，它是研究信息工作的理论、技术和方法的一门综合性横断学科。信息是现代社会三大资源之一，信息对于促进科学技术和国民经济的发展具有非常重大的作用。信息工作不但作为一种独立的行业而存在，而且是现代科学技术事业发展中不可缺少的重要组成部分，目前已进入由国家统一组织协调发展的崭新阶段。加强信息学基础理论研究，普及信息学基本知识，是开发、

利用信息资源，推动国家科学技术发展和经济社会进步的一项战略措施。

信息学研究的内容与范围极其广泛，既有理论方面，也有方法和技术方面。本书着重从基础理论方面探讨并揭示信息工作中广泛关注与研究的一些主要问题。如信息学的概念与学科性质；信息学的研究对象、范围与主要流派；信息学的产生、形成与发展趋势；信息学的哲学基础；信息学的数学基础；信息学的经济理论；信息学的学科体系结构；信息学系统论；信息学方法论，以及信息学应用技术等。

这里还必须对信息学和信息论的关系问题、两者在研究内容与范围方面有什么区别、其发展趋势如何，以及信息论对信息学的产生、形成与发展的作用 and 影响等做出明确而科学的回答。

信息论是研究信息现象及其运动规律的科学理论，分为狭义信息论和广义信息论。狭义信息论是关于通信技术的理论，也叫通信理论，或统计通信论。广义信息论是广泛研究信息的本质特性和信息的获取、计量、传输、存储、处理、控制与利用的一般规律问题，通常叫作信息科学。从狭义信息论到广义信息论，其研究范围就是由通信工程扩大到人类认识世界和改造世界的整个活动领域，并在实践中逐步形成一种独特的、具有普遍适用意义的新方法——信息方法。

显然，信息学作为一门独立的学科，在具体研究信息的搜集、加工、存储、检索、报道和有效地提供使用的理论、技术、方法及其交流全过程的规律性方面，与信息论有着天然的联系，既有共性，也有各自的特性。但从学科体系结构的总体上看，信息论是信息学形成与发展的基础。其不同点是，信息论以通信系统模型为研究对象，其重点是通信联系，而且这种联系广泛存在于自然界和人类社会之中。信息学研究的是—种特定的信息，即知识性信息。这种信息只存在于人类社会之中，其主体结构是知识。从知识的活化入手，研究知识信息的生产、构成、传递和交流结构与全过程的规律性，不断地为科技进步和经济社会的发展注入新的血液与动力。可以说，信息学与信息论尽管有交集之处，但是，两者在研究范围、研究目的、研究效果、研究的可靠性及其时间要求等方面都有很大的区别。

然而，随着现代高新技术的迅猛发展及其在信息工作中日益广泛的应用，使信息论与信息学的研究对象、内容与范围愈来愈密切。从学科结构的演化与发展趋势来看，信息论与信息学在向各自领域扩展与延伸的同时，有可能逐步发展成为—门完整的信息科学。这些问题在本书的第一章、第二章和第五章中已有详细的论述，在此不再作更多的说明。

《信息学基础理论》的出版，在我国还是第一次。本书内容涉及诸多专业，但主要侧重于基础理论，力求从理论与实践的结合上使人们对信息学有—个比较系统、全面的认识，其目的是为正在迅速兴起和蓬勃发展的信息学研究提供—些基本的理论和方法。作为—本专业性著作，它适用于广大的信息工作者、图书文献、档案资料人员及高等院校师生阅读。

本书在撰写过程中，得到不少专家同行积极、热情的支持与帮助，在此表示诚挚的谢意。书中不妥之处望读者批评指正。

符福顺

一九九三年初春

目 录

前 言

第一章 信息学导论 (1)

- 第一节 信息学发展史 (1)
- 第二节 信息学研究领域 (3)
- 第三节 国外信息学发展和主要流派 (17)
- 第四节 信息学在我国的发展 (21)
- 第五节 信息学发展趋势和展望 (23)

第二章 信息学哲学基础 (26)

- 第一节 引 论 (26)
- 第二节 信息学与哲学的关系 (26)
- 第三节 信息学哲学范畴的属性 (30)
- 第四节 信息的运动、运动形式及其在空间和时间上的存在形式 (35)
- 第五节 唯物论的信息学观与波普尔的“三个世界”理论 (40)
- 第六节 信息学的普遍联系和发展 (45)
- 第七节 信息学的矛盾运动规律 (50)
- 第八节 信息学的质量互变规律 (56)
- 第九节 信息学的遗传与变异 (62)
- 第十节 信息学与唯物辩证法的基本范畴 (69)
- 第十一节 信息学的认识和实践 (76)
- 第十二节 信息学的产业观 (84)

第三章 信息学数学基础 (94)

- 第一节 引 言 (94)
- 第二节 线性代数在信息学中的应用 (95)
- 第三节 概率分布与随机过程在信息学中的应用 (131)
- 第四节 数理统计与多元分析在信息学中的应用 (159)
- 第五节 集合论与图论在信息学中的应用 (197)
- 第六节 模糊数学在信息学中的应用 (224)

第四章 信息学经济理论 (259)

- 第一节 导 言 (259)
- 第二节 信息商品 (260)
- 第三节 信息商品的交换和流通 (268)

第四节	信息生产	(275)
第五节	信息产业和信息市场	(290)
第六节	信息商品的剩余价值、利润和资本的积累	(307)
第七节	信息工作经济管理	(318)
第五章	信息学学科体系结构	(324)
第一节	确立信息学分支学科体系的原则	(324)
第二节	信息学分支学科体系结构模式	(329)
第三节	理论信息学和应用信息学	(339)
第四节	信息学分支学科体系	(346)
第五节	信息学学科体系的发展趋势	(364)
第六章	信息学系统论	(367)
第一节	系统科学在信息学系统论中的应用	(367)
第二节	信息学系统论概说	(374)
第三节	信息学系统论的学科结构与功能	(378)
第四节	信息学系统的设计	(386)
第五节	信息学系统的决策	(390)
第六节	信息学系统的管理	(393)
第七章	信息学方法论	(401)
第一节	总论	(401)
第二节	信息学方法论体系	(407)
第三节	哲学方法	(411)
第四节	横断科学方法	(417)
第五节	经验科学方法	(429)
第六节	理性思维方法	(435)
第七节	信息学的专有方法	(448)
第八节	信息学研究中的移植方法	(470)
第八章	信息学应用技术	(476)
第一节	信息学、信息技术与信息工作	(476)
第二节	信息存储技术	(481)
第三节	计算机技术	(489)
第四节	数据库技术	(497)
第五节	通信技术	(511)
第六节	声像技术	(519)
第七节	现代复印技术	(523)
第八节	信息检索技术	(527)
第九节	其他信息技术	(563)

第一章 信息学导论

第一节 信息学发展史

一门学科的发展大多经历实验性描述阶段、理论综合阶段、数理论证阶段。

信息学的产生,是人们通过长期实践,不断探索、总结、升华的结果。自从有了大自然,就有了自然、生物信息,而人类社会的信息活动也经历了漫长的岁月。

一、人类信息活动的实践

人类的信息活动可以追溯到遥远的古代。早在6000年前,人类就发明了文字,用以记录、保存和传递信息,开始了有意识的信息活动。公元前25世纪,古埃及人把纸草(papyrus,一种植物,英文paper来源此拉丁文)的茎剖成薄片,压平后用作书写的材料;若干片粘成长幅,卷在木杆上形成卷轴(这种最早的“书籍”在古埃及已发现很多)。后来古希腊人在每个卷轴上,系上标题(titulus,即英文的title),并按字顺排列。公元前1700年,在巴比伦国王哈姆拉比时期,人们把楔形文字写在粘土板上。在我国,大约公元前1400年的殷代甲骨文,就记载了农业和医学的知识。文字的产生标志着人类思想交流进入第二阶段,已能够把知识和信息记载在一定的载体上,不过早期多记录在甲骨、金石、简策、木头、布帛等载体上。从古埃及人用纸草书写,到我国春秋、东汉(公元前8世纪到公元2世纪)多用竹简、布帛书写,直到公元105年蔡伦发明纸张为止,人类只是用手抄的方式记录信息。在世界的文字历史中,汉语3000年前就有甲骨文的文献,拉丁语在公元前6世纪起有文献,希腊语在公元前8世纪起有文献,阿拉伯语则有悠久的文献记载历史,而英语在公元7世纪起才有文献记载。

从文献工作角度出发,公元前669~630年的亚叙班尼佩尔(Assurbanipal)时期就有刻在粘土版上的目录。我国春秋末期的孔子(公元前479~551年)在整编“六经”时,就创立了揭示文献提要的方法。希腊的伟大哲学家亚里士多德(公元前384~322年)是目录学理论的创始人。他编写的第一部百科全书,是一本综合性分类学的著作。我国西汉(公元前206~公元25年)刘向、刘歆父子经过20多年的工作,完成了《别录》和《七略》。这是我国最早的书目编撰,也是世界最古老的大型综合目录。以后,班固(公元32~92年)编撰的《汉书·艺文志》在分类、目录、提要等文献工作中,做了有意义的工作。

现代图书目录被认为始于1564年,威勒(G. Willer)出版了书本式目录。在这以前,1041~1049年间,宋朝的毕升发明了胶泥活字版,直到1450年德国人古登堡(J. Gutenberg)综合了中国油墨技术和朝鲜的铜活字技术,发明了活字印刷术。这种技术使文献成为科技交流的主体,并有可能出现市售的书本式目录。

1654年,盖斯勒(K. Gessner)编辑了《世界目录大全》(Bibliotheca Universalis)。最早的科学杂志是1665年在伦敦出版的《哲学汇刊》(Philosophical Transactions)和法国出版《学者杂志》(Journal des Savants)。后者是周刊,是向读者介绍全欧出版的最重

要书刊。1682年德国出版了《Acta Eruditorum》书刊介绍,作成文摘。17世纪,在欧洲期刊已有了重要的地位,但这些期刊不是专业性的,而是综合性的、信息简介性刊物。这在某种意义上讲,是文献信息工作的雏形。

18世纪,文摘期刊在德国取得了进一步发展。1751年出版的《医药图书目录》,是最早的专业性文摘期刊。1765年又出版了《德国图书目录》。因此,我国科技信息界通常认为德国的《医药文摘》是世界上第一种文摘期刊,这是没有依据的。该刊1830年才问世。

目录学是在人们大量从事书目实践的基础上逐渐形成的。直到18世纪,大量文摘期刊的出现,才使目录学有了进一步发展。

虽然我国在西汉年代就有大型综合目录《别录》和《七略》,但在相当长的历史时期内,没有进一步发展,直到清朝(1773~1783年)才完成了《四库全书总目提要》,这是一本规模巨大的评论性文摘著作。近代的科学文摘在我国,直到1934年《化学》杂志上才开辟有“中国化学提要”专栏,比欧洲晚150年,比日本晚50年。

二、相关学科的发展为信息学奠定了基础

如前所述,目录学(Bibliography)是一门古老的学科,直到17世纪,人们对书目这一术语解释为“关于书籍的描述”。目录学是研究对文献进行描述性加工、处理的知识集合。在18世纪以前,人类的大量编制书目的实践,逐步促进了目录学的形成。

图书馆学(Library Science)直到19世纪初才逐步形成,其标志是1887年,当年德国的哥廷根(Göttingen)大学开办了图书馆课程,同年美国哥伦比亚大学杜威教授也在该校创办了图书馆学校。在这以前,1808年德国学者斯莱廷格(M. W. Schrettinger)首先提出图书馆学这一术语。1836年帕尼兹(A. Panizzi)为大英博物馆制定了目录规则。1867年在美国学者亨利(J. Henry)的倡议下,出版了《科学文献目录》,收入了1800年以后出版的期刊。1876年美国杜威(M. Dewey)首先采用十进分类法。在此基础上,比利时两位律师保尔·奥特莱(P. Otlet)和亨利·拉芳丹(H. Lafontaine)进一步发展了十进分类法,建立了国际十进分类法(UDC)。从某种意义上讲,文献工作的发展史实际上是国际十进分类法的发展史。我国运用这一术语始于20世纪初。

Documentation可以译成文献、文献工作,也可指文献学。这一术语是奥特莱1905年提出的。1908年在第四届国际书目协会(IIB)上首次接受这一术语。1930年IIB创办了《国际文献工作》会刊。1931年德国文献学家普赖斯(J. A. Price)参加领导工作,改名为“国际文献协会”(IID)。1937年又改名为“国际文献联合会”(FID)。1960年FID将文献学定义为“指对一切类型信息的收集和存贮、分类和选择、传播和利用。”30年代IIB、IID的主要工作是推广UDC,1934年布拉德福(S. C. Bradford)提出文献分散原理,认为应采用UDC来标引文献,以提高文献的标引质量。

目录学是一门古老的学科。本世纪初,目录学、文献学、图书馆学同时并存,直至40年代,才出现科学信息,信息检索,50年代才出现信息科学等术语。

三、信息学的萌芽和形成

“科学信息”这一术语是在1945年6月公开发表的布什(V. Bush)给美国当局的信件中提出的。他总结了战时科学信息的形成和发展,并对信息管理与应用提出了新的技术设计。他在应用缩微技术、存贮技术、代码技术的基础上,设计了“Memex”装置,提出了采

用机械化缩微检索方法, 查找文献。科学信息的概念后来被许多学者所接受, 并与科学交流相联系, 发展成信息学的交流学派。但处于信息时代的今天, 科学信息的概念, 逐渐被科学技术信息、科学技术社会信息所取代。

如果说文献学的历史是UDC发展的历史, 那么在一段时期, 信息检索已成为信息学的别名。1950年在一次国际数学会上, 莫尔斯(C. N. Mooers)提出了“信息检索”这一术语, 给信息查询活动以科学的概念。当时他提出这一概念是针对穿孔卡片检索而言的。

“信息检索”、“科学信息”这两个学术概念的出现推动和发展了信息学学科的建立, 被认为是信息学诞生的先声。1948年申农(C. Shannon)和维纳(N. Wiener)先后发表的《通信的数学原理》和《控制论: 或动物和机器中控制和通信的科学》两篇著作, 为信息学的建立奠定了理论基础。同时值得一提的是同年在伦敦召开了科学信息会议, 这是各国学者参加的第一次国际科学会议, 会议的议题涉及信息工作的方方面面。1958年由美国文献学会(ADI)、美国科学基金会、美国科学院等联合召开华盛顿国际科学信息会议。这两次重要会议为信息学在世界范围内的建立打下了牢固的基础。

1946年世界第一台数字计算机问世。1959年, Information Science一词在美国正式提出。1962年法国创造了Informatique一词, 1966年被法兰西科学院承认。1962年前苏联学者提出информатика英文为Informatics的概念和术语。在美国提出Information Science的前两年, 我国就确立了情报科学的概念。1968年美国文献学会改名为美国情报科学学会(ASIS), 1970年其会刊也作了相应的变更。1977年在美国奥本尼大学讨论会上, 提出了图书馆信息学的概念, 随后在美国又出现了Information and Computer Science和Library and Information Science两门学科, 显然前者侧重于计算机技术本身, 但它对信息学的建立与发展也起了很大的促进作用。

第二节 信息学研究领域

纵观信息学发展的历史, 分析信息学研究的现状, 可以看出, 信息学是一门新兴的学科, 正处在形成、发展与完善之中。它是以普遍存在的社会现象——信息与信息过程作为研究对象的。

一、信息的基本概念

目前人们对信息的定义认识还不一致, 对Information这一词的解释有几十种, 直接与知识信息有关的定义有37种。早在1959~1971年期间, 联合国教科文组织(UNESCO)负责科技信息的机构PCI就对37种定义进行过研究。这些定义(即概念种差和邻近属概念)的释文是由词和词组组成。这37种定义共计使用自然语言词汇81个, 如知识、应用、技术、描述等。词频统计的结果, 使用频率最高的词依次是: 传递(Transmission)、语言、流(Flow)、处理、交流(Communication)。在37种定义中, 兼融这5个高频词汇的定义者只有3人。因此, 调查统计的结果表明, 有关知识信息的术语尚未一致公认和定义。

在我国, 80年代初对信息学赖以存在的最基本概念——“信息”, 进行了热烈的争论, 结果仍然没有一个令人共同接受的定义。但这次争论也不是无意义的, 它为信息学研究开拓了崭新的领域。

事实上, 自50年代以来, Information一词使用十分广泛, 几乎包括了所有学科、专业,

乃至日常生活的各个方面。究其原因,是因为它涉及到科学思想的发展、技术发明与应用以及科技进步造成的社会变革等因素。因而相关的术语、学术观点、科学理论也在迅速发展。虽然目前关于“知识信息”的概念尚未取得一致的认识,但其中3点是被共同强调的:

1. 为了一定目的而系统收集起来的正确的信息和事实所构成的知识,也就是说,经过传递而又为人们所吸收的知识构成信息。

2. 通过传递而进入人类社会交流体系的运动着的知识构成的信息,即一方面信息可以通过一定的物质载体或其它交流形式进行传递,另一方面获取信息必须经过传递。

3. 信息不仅是传递中的知识,而且必须是有效用的知识。凡是人类社会的信息均有一定的效用,或与接收者原来知识状态吻合而被认可;或有相异,从而发生知识的重构,生成新的知识结构。

钱学森称信息“是激活了、活化了的”知识,信息“就是为了解决一个特定的问题所需要的知识。”^{〔1〕}他对信息概念的解释对科技信息界关于信息定义的争鸣是极为重要的、有益的。

因此,可以说,在社会交流过程中,知识变为信息,信息转化为知识,它们总是处在相互作用之中。知识是静态的,以实践为基础,通过抽象思维,对客观事物规律性给予概括。知识信息是人类社会中客观存在着的部分社会信息,是以知识形态为主,还包括数据、新闻、消息等非知识形态的社会信息。所以说,知识信息是知识的激活,知识是知识信息的条理化。

Information一词来源于拉丁语Informatio,表示传达的过程和内容,最初传达仅仅意味着告知,而被告知的只能是以语言为外壳的人的思想内容。也就是说,Information是人类社会特有的现象。随着科学技术的发展,Information的语用范围也随之扩大了。1948年,N.维纳以《控制论或关于动物和机器中的控制与通信》为标题,阐述控制论的基本原理。尔后,维纳关于Infiromation的论断被广泛引证,于是人们开始把Infomation与诸如空间、时间、运动和能量之类的普遍概念联系在一起。这样,Information(信息)不是一个概念,而是一组概念群(见图1-2-1)。N.维纳断言:“信息就是信息,既非物质,也非能量。”^{〔2〕}美国学者萨拉塞维克说:“在研究信息时,我们面对着一个展示出其物理的、生物的和行为的多种特性的复杂现象。”^{〔3〕}他把问题引向了物理学、生物学和行为科学的多种不同的层次上。英国学者B.C.布鲁克斯将信息看成是知识结构的一个小小的组成部分,它可以改变知识的结构^{〔4〕}。

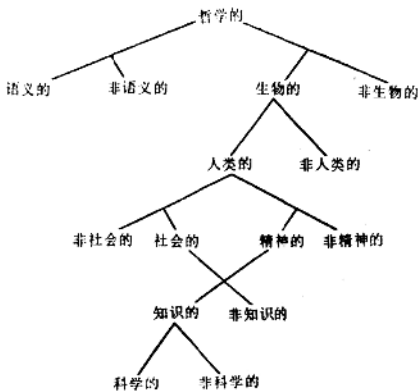


图1-2-1 信息定义域

因此,信息处于一个多维与多层次的定义域里。

Information作为比较抽象、比较正式的学术名称,主要有两种译名——信息、情报,台湾学者译为“资料”、“资讯”。一般说来,信息是Information比较恰当的对应概念。目前,形成以下几种关于信息的语用现象:

(1)信息可以表示一切交流、传递现象,诸如语义的、非语义的、人类的、生物界的等等。它可以在抽象的意义上使用。其指称范围超越了人类社会、生物界,甚至成为哲学研究的课题。

(2)知识信息只能表示人类社会的具有语义的信息交流和传递,尤其用来表示人际间的知识信息交流。它仍在谍报意义上被使用,如军事情报。

(3)在人类社会的许多领域里,信息与情报的概念变得模糊,有时在我国大陆信息和情报作为同义词使用。在海外,英、法、德语中均称为Information,俄语为 информация,日语也通称“情报”,均为一词,但都有特定的含义。

(4)仅在特定的领域里,信息被做为正式术语,在我国大陆已被广泛使用。如科技信息、社科信息、图书信息、文献信息以及这些领域中的信息科学。

二、信息学有关概念体系

上一节,探讨了信息的概念。信息学是以自然和社会中普遍存在的信息现象和信息过程为研究对象。我们在这里将信息学作为术语来探讨其概念体系。

术语学(Terminology)是在一个专门主题领域里,或所有主题领域的集合里,涉及概念的构成与命名的科学。术语(Term)则是表示专门知识领域的概念,反映概念间相互关系的词或词组。因此,从术语学的观点出发,术语是从概念出发,研究表达概念所需的名称(即称谓),而非术语的词或词组是从名称出发,研究名称所表示的意义。虽然术语和词都是语言现象,但出发点,研究对象却不一样。因此,研究术语要划分和确定概念,给予明确的定义,分清概念的范畴、关系,然后将概念按学科进行系统的分类。本节从术语学的观点出发,对信息学的概念体系作必要的阐述。

1. 外来各种信息学名词的辨析

随着工业社会向信息社会的转化,尤其在西方发达国家,信息技术已成为整个技术领域中的先导技术,对信息、知识信息的研究占突出的位置,涉及并形成了各个新的科学门类 and 结构。因此,对国外迄今为止的涉及知识信息、信息的学科名称,进行学术鉴别是无意义的。

(1) Information Science (信息学)

此词早在1959年在美国便已出现。1961年在美国乔治亚工程大学召开的信息学会议上,信息学有好几个定义被提出来。泰劳尔(Talor)指出:“信息学包括了在交流领域中,从信息产生到最终使用的各个环节的所有职能。”而学者海斯(Hayes)则是从信息产生系统的观点来认识,他说:“信息学是一门研究特定信息系统中信息传递、选择、处理和分析过程的学科。”1967年当时任美国文献学会主席的L.杰鲁尔(Jerool)在给会员一封信中写道:“信息学是研究信息的特性和活动、管理传递过程的手段以及保证信息的最有效利用所必需的信息处理技术。信息学的研究范围包括:利用自然的和人工的系统提供信息的方法,利用号码有效进行信息的传递、存贮和检索,研究信息处理设备和方法,如计算机及其程序设计系统。”他指出的定义,经会员略作修改后,即被美国文献学会(ADI)采纳。随后,1968年

该会改名为美国情报科学学会 (ASIS)。至今已有21年历史的《美国文献工作》(American Documentation) 自1970年被改名为《美国情报科学学会会刊》(Journal of ASIS)。从此, 信息学在美国扎下根, 并在不断完善和发展。

在前苏联与Information Science对应的术语是информатика, 按米哈依诺夫的解释, 信息学是“研究科学信息的结构和基本性质, 并研究科学交流所有过程的一般规律性的一门科学。”^[5]在此以前, 前苏联也用过информационная Наука表示信息学。

在德国, 人们采用Informations-und Dokumentationswissenschaft(信息与文献科学)一词表示信息学, 在日本被称为情报科学, 在我国大陆译为情报学、信息学, 在台湾译为资料科学资讯科学。

(2) Informatics (信息学或信息自动化处理技术)

此词最早出现在法国, 法文为Informatique, 是1962年由Philippe Dreyfus 工程师创造的新词, 它由INFORMATION加autoMATIQUE构成, 表示信息处理自动化的含义。尔后, 在报刊杂志上使用频率较高。同年苏联科学院院士A. A. 纳尔凯维奇译成俄语, 称为Информатика, 组词法相同, 后来, 米哈依洛夫赋予该词新的内容(见上述)。1966年, 法兰西科学院(语言学方面的权威机构)予以承认, 定义为, “一门在技术、经济和社会领域里, 用计算机处理被视为人类知识载体的信息的一门学科和技术的总体。”因而, Informatique在汉语中译为信息处理技术更确切。在日本称为“情报工学”。而美国词典则注释此词Informatics译自俄文информатика, 而不是法文Informatique。英、美国有的学者则认为Informatics是由法语派生的词, 意为信息自动化处理研究,^[13]等于automatic data processing。因此, 在美国信息界大多数人仍坚持用Information Science这一术语。但俄文的 Информатика与Information Science和Informatics完全同义。在我国大陆将前者译为信息科学, 将后者译为信息学, 试问这两者在字面含义上有什么区别。

这门新学科(Informatique)的基础理论不是直接跟计算机的使用相联系的, 而是建立在逻辑学、数学、语言学、形式语法、编译以及计算机结构的基础上, 其学科体系见图1-2-2。

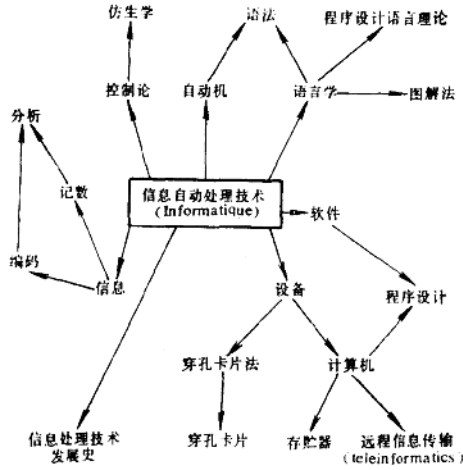


图1-2-2 信息自动处理技术应用学科的体系结构

此外,在这一领域里,还有 Teleinformatics (远程信息传输技术) 和 Telematics (远程信息处理技术) 是在这一门应用科学技术领域内与远程通信技术相关的技术学科。

(3) Information Sciences (信息科学)

这一术语是指和信息处理与系统的研究有关的所有科学,也就是指所有信息(包括人类信息和非人类信息,生物信息和非生物信息,语义和非语义信息)研究的整合。如同使用信息技术 (Information Technology) 一词一样,使用信息科学一词,是试图打破传统的技术和科学基本部类的界限。

信息科学不是交叉科学,而象自然科学和社会科学一样作为一个科学基本部类的名称。美国著名学者F·马克卢普所编写的《信息研究:学科之间的通讯》^[7]一书认为信息科学应包括:哲学中有关信息范畴的研究,控制论和系统论中有关信息一般功能的研究,认识论,认知科学,大脑理论中有关人类精神信息的研究,符号学、语义学、象征学、图书馆学、文献学、档案学、博物馆学、传播学、新闻学、教育学等,信息科学的边缘学科,如信息社会学、信息经济学、信息管理学、信息心理学等,还有人工智能以及信息技术所包含的各种技术领域有关的研究。因而,信息科学是一门包罗所有信息研究的综合学科。

(4) Library and Information Science (图书馆与信息学)

美国学者雷沃德 (W.B. Rayward) 在《图书馆与信息学》一文中,对有关术语进行了考证,并列出了 Librarianship→Library Science→Information Science→Informatic 的学科易名顺序。有关学者也曾有“信息学产生于图书馆学”或“信息学等于图书馆学加信息技术”的说法,还有的观点认为“信息学是现代图书馆学”。但也有的学者不同意上述论述,认为:“信息学是社会信息化的产物”,“不是图书馆学自然的延伸”。^[8]加拿大一学者卡洛林在《信息时代的信息科学》^[9]一文中认为:“信息学不是图书馆学和文献学的发展,而是一门与新通信领域相关的学科。”T. 萨拉塞维克划出了10种信息学专业期刊,统计结果表明:论述图书馆学的论文百分比逐渐减少,论述计算机应用所占的百分比日益增加。因此,信息学 (Information Science) 与信息处理技术 (Informatics) 有深刻的联系。

日本图书馆学家齐藤作了引文分析研究,证明图书馆学是比信息学范围更大的学科。在信息学的期刊论文引文中,图书馆学和信息学文献占48%,其次是信息技术方面的文献占18%,再其次是其它学科。而在图书馆学期刊论文引文中,图书馆学和信息学的文献被引率高达78%,而其它学科的引文率均未超过3%。

由此可见,Library and Information Science这一术语表示图书馆学与信息学是两个平行的学科,彼此相互联系,但又是各自独立的学科。目前美国大学的学院、科系的名称多取此意义,但与计算机科学是有差异的。我国也大体相同,国家学位评审委员会也将图书馆与情报学作为理科A级学科。但在德国和西欧有些国家则采用Information and Documentation Science (信息和文献学)的称谓。这说明图书、文献、信息部门都以知识信息作为研究对象,在信息技术应用方面也几乎相互重叠,向信息概念靠拢。

以上对外来术语的讨论以及其译名的探讨是非常重要的,交待了这些典型术语出现的背景、研究的范围和限度以及翻译上的准确性问题,与信息学直接相关的学术名称的分歧和不统一性,一方面说明这一学科还正在形成与发展之中,另一方面也表示我们讨论对象的模糊性。但随着信息学理论体系的逐步建立,研究成果的不断涌现,那么信息学术语的科学概括就会趋于统一化和规范化。

2. 信息学的定义

如信息的概念没有明确的定论一样,信息学的定义,在国内外学者中,也没有一致的意见;甚至出现两个以上的称谓,这说明信息学的概念是含糊的。

前苏联学者米哈依洛夫指出,信息学是“研究科学信息的结构和基本性质,并研究科学交流所有过程一般规律的一门科学。”^[1]简而言之,他认为信息学是关于科学交流和科学信息的科学。信息学的产生是由于科学信息活动,是从科学研究活动和科学组织活动中分离出来,因而形成相对的领域。因此,信息学既非图书馆学,也不是来自图书馆学母体,同样也非计算机科学。很明显,米哈依洛夫在定义信息学概念时,将其研究对象仅限于科学信息和科学交流的过程,因而存在着明显的局限性。应该说,从当今社会的“大信息观”来衡量,其严格限定研究对象为科学信息未免太窄,但他最大的特点在于重视信息的社会性,强调提高社会中科学交流效率的重要性,以及计算机不可能成为信息发展的关键因素,这些观点是有积极作用的。他作为知识交流学派的重要人物也是有代表意义的。

B.C.布鲁克斯主张“信息的任务是探索和组织客观知识”。^[4]他提出从哲学的高度探索信息学基础理论,提出信息作用于知识的方程式,以及主张活化静态的知识结构,提出把文献孕含着的知识绘成“知识地图”。其学术观点鲜明并具有新意。但他采用了波普世界3的多元论哲学观点,提出的知识结构也有不成熟之处。他研究课题是关于知识在交流中的特性、功能、生成、转换和发展等。而美国教授F.W.兰卡斯特(Lancaster)从事信息检索、词表、文献计量学、信息系统和信息服务的研究。他认为信息学研究信息技术应用于知识交流的基本原理和规律,以及服务于知识交流的信息系统的设计原理。他们被称之为客观知识学派的代表人物。

这是比较有代表性的前苏联和英美信息学者对信息学概念作的解释。此外,国际标准组织和我国的国家标准也都有明确的定义。

1979年ISO提出,“信息学是对信息的功能、结构、传递的研究和信息系统管理的研究。”我国标准(GB4894—85)关于《情报与文献工作词汇基本术语》把信息学定义为“研究情报获取,传递与使用的理论、规律与方法以及情报系统管理的学科。”

我们认为应该从术语学的观点去研究信息学的定义。术语是由名称和概念组成的语言符号。名称是术语外在的语言形式,是表达概念的;概念(notion)和哲学的概念(concept)相近,其主要特征有外形、功能和性质三个方面。

因此,概念是从一组实体(事物或观念)的主要特征的总和中,减去它们的个别特征而得到的抽象概括。

从术语学的观点出发,信息学研究的是人类社会中部分社会信息——以知识形态为主的,也包括新闻,消息等非知识形态的那部分信息。而信息科学或广义信息学则研究一切信息——人类的、非人类的,生物的,非生物的,包括哲学、控制论和系统论中有关信息的研究。而信息学阐明社会信息现象、信息过程和信息行为的范畴。虽然图书馆学和信息学在研究内容上有相同之处,采用的原理、方法、技术手段也比较相近;但由于图书馆的社会功能以保存文化遗产,提高社会科学文化水平为主,而信息机构则侧重于开发和利用信息资源,因而决定了信息学和图书馆学研究内容的明显差别。信息学在信息源(侧重于科技报告、专利、会议录,而图书馆学侧重图书),在信息检索(图书馆学侧重研究目录学),在信息用户和需求(图书馆学仅研究读者、阅览和参考服务),在信息系统——含文献支持系统(图书馆学研究藏书结构,以藏为主)等研究范畴里都有根本性的差别。

由此可见,信息学是社会信息化的产物,其研究对象是社会信息现象、信息过程和信息行为,探索社会信息化的规律性。这就是信息学特定的研究领域。也就是说,信息学是研究信息获取、整序、传递和利用的规律性的一门科学。

3. 信息学研究的具体问题

信息学研究包括以下几个方面的内容:

(1) 基本概念问题:信息的概念、属性、特性、与科技、经济、社会的关系;信息学的研究对象、学科性质与相关学科的关系。

(2) 基本原理和理论体系:联系以上诸单元基本概念的判断、思维的展开或组合形式。例如文献计量学6个定律:作者分布定律、文献分散定律、词频分布定律、文献老化定律、文献增长定律、文献聚类规律,以及信息计量学的研究——其基础是科学计量学。知识组织结构的研究会引起信息学理论体系的重大突破。

(3) 信息产生源的研究,涉及客观知识、知识性信息;信息序化的研究,涉及知识、信息从混沌到序化的过程和规律;信息传递的研究,涉及传播、交流理论,信息技术的应用以及信息价值转化的方式和过程的研究。

(4) 信息增值的研究,涉及信息用户、信息需求的类型、特点与规律的研究,信息转化为生产力的理论与实践的研究。

(5) 信息系统的设计、运行和评价。

(6) 信息的管理、组织问题:信息政策、信息体系、运行机制、信息立法以及信息教育等。

(7) 信息技术应用问题:计算机、通信、存储等技术的应用等。

三、信息学的基本原理

建立一门科学理论要有3个主要要素有机构成。本章在第二节中论述了信息学的研究对象和基本概念。基本概念是科学理论的初始思维单元,基本概念群或概念体系构成单元的集合,为科学理论的建立,提供原始的前提。本节在基本概念的基础上演绎、衍生、推论出信息学的基本原理。

纵观信息学研究范畴,概括信息学的客观现象和基本规律,确认其基本原理如下:

1. 信息产生原理

信息产生原理涉及到信息基本概念、信息与其它各种社会现象的相互联系和作用,信息源等基本概念群。

这一原理有比较成熟的理论基础,即文献计量学的6个定律:

(1) 作者分布定律:文献统计中的洛特卡定律(Lotka's law),是一个平方反比律。它的作用在于指示作者与文献数量的关系,描述科学生产率问题,即科技工作者论著的数量。

$$\text{洛特卡定律可表达为 } f(x) = \frac{C}{x^2}$$

C为常数,论文数X与作者频率 $f(x)$ 之间存在着一定的关系。这就是科学生产的平方反比率。

(2) 文献分散定律:英国人布拉德福(S.C. Bradford) 1933年进行了大量的统计调查,1934年发表了《特定主题的信息源》首次用数学模式描述了文献分散的客观规律。

其公式为: $R(n) = K/n(an+1)$

布氏定律及其区域分析法、图象观察法、以及假设推理 8 种不同的方法,对信息检索、信息搜集和选择核心期刊及核心文献都有实际的应用价值。此后,布鲁克斯又从数学上对布氏定律的实际应用技术进行探讨和完善。

(3) 词频分布定律:美国语言学家齐普夫(G.K.Zipf)收集了大量统计材料,试图证明自然语言词汇的分布服从一个简单的定律,他称为“省力原则”。

$$r^m \cdot f = C$$

齐氏定律在标引和词表编制理论以及计算机信息检索中应用比较广泛。

(4) 文献增长定律:美国科学史家 D. 普赖斯在他的名著《巴比伦以来的科学》一书中,以科学杂志和学术论文为知识发展的两个重要标志,进而对知识量的增长率进行了推算。通过曲线分析,普氏注意到科技文献增长与时间指数函数的关系。如果用 $F(t)$ 表示时刻 t 的文献量,则指数定律为: $F(t) = ae^{bt}$

a 是统计初始的文献量, e 是自然对数底, b 是时间常数(持续增长率)。

考虑到某些因素,如物质的、经济的、智力的因素对文献增长率的影响,有的学者提出文献逻辑曲线增长理论,其方程为:

$$F(t) = \frac{K}{1 + ae^{bt}}$$

(5) 文献老化定律:目前描述文献老化规律的数学公式一般都采用 $C(t) = Ke^{-at}$ 式中 $C(t)$ 表示发表了 t 年的文献引用频率, K 是常数, e 是自然对数的底, a 是老化率。

R·巴尔顿和 R·凯普勒提出用下列解析式来描述老化: $Y = 1 - \left(\frac{a}{e^x} + \frac{b}{e^{2x}} \right)$

式中: $a + b = 1$, Y 经过一定时间被利用的某一学科或专业全部文献的相对部分; X 时间,以 10 年为单位

(6) 文献聚类定律:涉及文献耦合(document coupling),同被引(co-citation),自引、引用文献(citing document),被引文献(cited document)等概念。

引证分析法用来确定核心文献(含期刊)一般采用引用次数法、被引次数法、引用被引用法、引用被引比较法,同时还要考虑影响因数法,文献耦合原理以及综合法(被引、被摘及流通)等方法。

文献计量学的理论方法,自 1923 年英国 Hulme 提出书目统计学(Statistical Bibliography)以来,已建立了扎实的理论基础。文献计量学在科学计量学(Scientometrics)和信息计量学(Informetrics)的影响下,将会进一步发展,改变信息学现状,使信息学理论体系更加完善和健全。如建立联合概率模型,即把传统的独立模型组配起来。1983 年南非的 H. S. 希切尔(Sichel)提出广义负高斯泊松分布(GIGPD-Generalized Inverse Gaussian Poisson Distribution),将包含多种概率分布统一解释文献计量几个基本定律。

由于当今信息源已不再是单一的纸介质,而是多种媒体。数据库的出现使以信息要素为单元的信息计量学得以充分的发展。目前开发的智能软件能检索、统计知识单元、信息单元,使信息分析定量化成为现实。此外,文献计量学还涉及知识结构、思想基因等原理。

2. 信息序化原理

知识序化理论和方法是信息学研究的核心。在信息产生(信息源)、序化、传递、增值(吸收、利用)这 4 个主要现象中,序化是起主导作用的,也是信息这一社会现象中最本质的现象。序化就是将杂乱无章、随机的知识,加以整序、分析综合成人们解决问题的形态。

从哲学上讲人们认识和改造客观事物是一种从混沌到有序,从初级有序到高级有序的进化过程。

信息序化原理是依据 I. 普利高津 (Prigogine) 的耗散结构理论来支配的。耗散结构理论同样是人类信息现象和行为的基本原理。普利高津假设,一个系统形成耗散结构至少要有 4 个条件:①系统必须是一个开放系统,②系统必须处于远离平衡的状态,③系统内部各要素之间存在着非线性的相互作用,④涨落导致有序。

系统的总熵增量 $ds = dis + des$

如果使 $des < 0$, 则有可能使系统的总熵值减少,以至达到低熵值。这时系统可能处于有序状态。

德国教授哈肯 (Haken) 提出协同论,认为:各种不同系统在一定的条件下,其内部各子系统之间通过非线性相互作用产生协同效应,从混沌状态走向有序,从低级有序走向高级有序,或者从有序转化为无序。

此外,在信息检索理论方面,也提出了一些基本原理和方法。

检索效率:1955年, J. W. 佩里 (Perry) 首先提出查全率 (R) 和查准率 (P)。1963年, C. W. Cleverdon 发现 R 和 P 之间的互逆相关关系。J. A. Swets 提出: R 和 P 并不是理想的检索效果评价指标。检索效率综合评价指标应为覆盖率、查全率、查准率、响应时间、用户费力程度、输出格式等。此外,不少学者试图用数学,如集合论、决策论、非线性优化方法、模糊数学等应用于检索理论。

信息检索理论还包括检索系统模式和检索语言的理论和方法。

序化原理是以相关特征为基础的。相关特征包括内容相关、形式相关两大类,细分还可包括功能相关、结构相关、隶属相关、并列相关等,形成系统的理论体系。

3. 信息传递原理

研究传递交流的行为和过程。研究传递、交流的模式、信息存取的理论和方法,研究正式交流、非正式交流的理论和方法;研究通信技术、联机检索系统、网络技术、接口技术、机器翻译、数字、图像、声音的传递技术和应用等。

信息传递必须处于激发状态,即 $I \geq I_0$ 。传递信息所需的时间 (T) 与其自身的价值 (I) 和信息用户对它的需求强度 (F) 成正比,与传递环境阻力 (i) 成反比。其数学描述式为:

$$T = K \cdot \frac{I}{F} \cdot f$$

在国外,米哈依洛夫对交流理论,申农对通信数学原理,维纳对控制论, T. 萨拉塞维克对人类知识通信等基本原理都有深入的研究。申农的早期合作者 W. 韦费尔 (Weaver) 把通信包括三个层次:①技术问题—传递通信符号,②语义问题—使符号精确表达语义,③有效性问题—接收者有效发挥作用。近10几年来,在通信的语义和有效性方面出现了一些新概念与理论。

4. 信息增殖原理

信息增值原理即信息的吸收利用原理。

布鲁克斯在信息作用于知识结构方面也提出了基本方程式,对探讨信息增殖原理具有重大意义。

如果我们进一步探讨知识的组织结构,这种知识的集合体最初的形式是图书馆的话,那

么在当今便是数据库,未来则是智能化的思想库。布氏曾提出分类法、主题法组织的不是知识,而是文献。他提出按“认识地图”来组织。认识地图系指分析文献中的逻辑内容,找到人们创造与思考的相互影响及联结点,象地图一样,直观地标示出来,展示知识的有机结构。随着自动标引和全文数据库的发展,人工智能软件的开发,一种认识地图的雏型软件 Leximape 已出现。它可将文献中出现的知识及其相关学科以树形结构表示,并标出有关学科的研究单位等。布氏还提出将“静态知识结构活化,构成人的体外大脑”建立思想库。智能信息检索系统不仅能帮助人们检索信息,而且还应该使人们分辨与吸收这些信息。

钱学森的学术观点认为信息是激活的知识,信息学是思维科学。M. C. 约维兹研究信息在决策过程中的影响和作用。他们研究的基点也在信息吸收和利用上。

应该说,信息学不单是一门单纯研究“物”(文献)的学问,也不单纯研究信息技术应用的学科,而是一门以人和社会为主体研究对象的学科。因此,要研究人的信息吸收、人的信息心理、人的信息行为以及人吸收知识的过程等一系列问题。

信息学的研究应该重视认知科学和知识工程这两门学科,并把它们作为重点基础科学来完善和发展自己的理论体系。目前信息学研究中涉及的认知科学内容主要限于用户方面,没有认知的要求就不存在对信息的需求。

四、信息学理论体系

信息学理论体系是指信息学研究对象,即信息这一社会现象和活动的内在联系和规律的几个具体理论的综合,并在一定的框架内将各种信息学概念、原理、分支学科形成一个整体。

在信息学理论建设上,应按研究对象→理论基础→学科体系→相关学科→方法论体系这样的模式推进理论研究。要防止两种干扰:唯理论研究的干扰、庸俗实用主义的干扰,避免基础理论研究、工程技术研究、应用研究的脱离、分割的状况,使信息学得以协调发展。同时,体系研究要使信息学各分支学科之间建立起逻辑联系,注意整体层次的研究,避免分化的局限性,保持学科体系发展的系统性和科学性。

1. 信息学理论体系研究的现状

一门学科理论体系的建立和发展一般要经历以下几个阶段:

①经验阶段或理论体系孕育形成阶段。处于技术、方法的经验描述和初步的理论总结,体系的孕育形成。这一阶段主要是描述性的理论和方法。

②理论体系的建立和稳定阶段。研究对象有关概念体系已明确,创建解释许多事实的原则和理论,并得到实践的证实,有成熟的理论和方法。

③理论体系成熟阶段。基本理论、分支理论、应用理论和技术方法,已相当成熟。

④理论体系预测阶段。在原有结构的基础上,派生出预测未来,代表未来发展前景的新层次——科学前沿。

由于信息学研究对象定义不统一,边界模糊,没有自己独特的方法,理论落后于实践等方面的原因,当前信息学正处在从阶段(1)向阶段(2)的过渡之中。

(1) 国外信息学理论体系研究

由于各个国家信息实践方式不同,信息工作的模式多样化,以及时代和历史、文化背景的差异,产生了不同的理论体系。如60~70年代以前苏联米哈依洛夫为代表的科学交流型理论体系,80年代以英国布鲁克斯为代表的知识利用型理论体系,70~80年代兴起的以美国萨尔顿和日本北川敏南为代表的信息技术应用型体系以及以兰卡斯特为代表的联机信息检索型