

庆 祝
中华人民共和国
成立四十周年



杨廷郊论文选

吉林省图书馆学会
四川省图书馆学会 联合编辑
成都东方图书馆学研究所

杨廷郊选集

吉林省图书馆学会
四川省图书馆学会 合编
成都东方图书馆学研究所
一九八八年

杨廷郊论文选

吉林省图书馆学会

四川省图书馆学会

主编

成都东方图书馆学研究所

成都东方图书馆学研究所出版

四川省图书馆学会发行组发行

(成都市总府街六号)

成都市自力印刷厂印刷 1988年

字数 100千字 印数1—1000册

四川省新闻出版局准印证88048

3.00元

简 历



杨延郊 男，1946年1月生，沈阳人。一九七〇年毕业于北京外国语学院英语系。同年末，分配到中国科学院沈阳自动化研究所，从事情报工作，开始广泛接触自动化科学技术的理论与实践，参与了诸多人工智能，机器人，高能加速器自动控制 and 粒子径迹识别，自动绘图机，DNC系统等十几项课题的情报调研或方案论证工作，先后写有《人工智能》，《遥感及遥感图像信息处理系统》，《国外机加工自动化发展概况》，《自动绘图机技术综述》等十几篇综述性文章，在内部或公开发表。在此期间，还编过内部刊物《自动化情报》和公开刊物《国外自动化》。一九七九年秋，考入武汉大学图书馆学系情报学专业，攻读研究生。一九八二年秋毕业，并获硕士学位。曾先后在《图书情报工作》，《科技情报工作》，《情报学刊》，《情报科学》，《图书馆学研究》等刊物上发表了十多篇文章和译文。一九八二年十月至今，在辽宁省科委科技政策经济处工作，开始研究世界新产业革命和技术进步经济学。发表有《技术进步与技术开发》，《信息化社会的社会结构》（译文）。编有《辽宁省科技法规选编》。

著者的话

承蒙《中国图书馆学论丛》编辑部的盛情相邀，我把近年来在图书馆学、情报学期刊上发表过的文章收集拢来，从中选择了九篇汇编成了这个小册子。作为在情报学这门知识的瀚海中涉足未深的年轻人，深感学识尚浅，水平有限。我之所以斗胆而这样做，完全是为了得到各位老师和同行的指导和帮助。

本选集包括了三个方面的九篇文章。它们是 1) 情报学基础；2) 索引法和索引的检索；3) 文献计量学方法及应用。都是些知识性、实用性较强的文章。近几年，图书馆学、情报学在理论和实践上都取得了长足的进步。其研究方法也日臻完善。人们可以预见，研究方法的成熟与发展对这门学科进一步发展将起到的作用。我把它当作了一个重要的工作领域。

借此机会，我想对《中国图书馆学论丛》编辑部对我的信任和鼓励，表示由衷的感谢！

杨延郊 1985.元月

目 录

简历

著者的话

情报科学的产生和发展.....	(1)
试论科学情报学及其若干基本理论问题	(18)
知识、情报和科学情报学	(40)
机编关键词文摘索引的形态及特征	(44)
研究微观科学结构的方法——引文分析	(60)
《科学引文索引》的检索策略.....	(73)
主题索引及其计算机编制	(91)
布拉德福定律的基本原理及应用.....	(109)
征询法在图书馆学研究中的应用.....	(120)

情报科学的产生和发展

一、引言

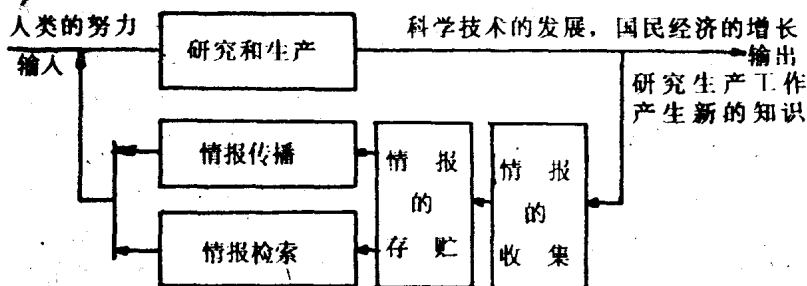
情报科学或者情报学，是四十年代中期适应现代科学技术的发展，为解决科学技术的有效通讯，在数学、逻辑学、语言学、信息论、控制论、心理学、计算机技术、运筹学、图书馆学和管理学等学科的交叉下产生和发展起来的一门边缘学科。一般地说，它是研究人类通讯的现象和通讯系统特性的一门科学；特殊地说，它是关于科学的有效通讯及其一般规律，即研究通过一切通讯方式有效地收集、加工、存贮、传播和利用科学情报的过程的一门科学。对于情报的定义，多少年来，众说纷纭，在有些语言例如英语中，其同义词竟达一百多个。但通过研究，人们逐渐明确了情报一词的基本意义是“获知”或“告知”，即情报的意思，由于不断地使用而将其名词化了。现在，人们将情报划分为生物情报、机器情报和社会情报三种。广义的情报科学是以这三种情报的共同理论问题为研究对象的。我们所说的情报科学，仅以社会情报特别是科学情报作为对象，这是目前整个学科的基本面貌。在人类社会，知识的积累、传播和借鉴，是人类文明的重要表现。美国学者 Brookes 提出了反映人的知识的这一过程的基本理论：

$$\Delta I + (s) = (s + \Delta s)$$

式中， Δ 表示增量， s 为已有的知识结构， ΔI 为情报输入， $s + \Delta s$ 为新的知识结构。这就是说，知识是通过情报的获得而增加的。现在如果我们假设，情报是传送的消息中的部分量，

则可推知，所谓情报，乃是消息中使知识发生变化的那些东西。它一般表现为传送中所需的那部分知识。社会情报，一般都是人们行为的一种结果。人的思想行为的结果或存在于人的意识中（表现为知识），或存在某种物质媒介或载体，例如：文献上作为意识数据的一种记录（知识被记录下来就变成了公开知识）。它们通过人类的通讯过程而被收集、存贮、加工、利用和创造，从而表现为情报过程。我们可用图一所示的闭环系统来说明这一过程及其与人们的社会实践的关系。

人们获取和利用大量的情报（输入），通过科研活动产生科研成果，从而促进了科学技术的发展和国民经济的增长，同时也产生了新的知识。这是一个周而复始，螺旋上升的过程。而整个情报过程恰是一个反馈环节。这个环节反馈得愈快，整个过程的效率就愈高。情报科学的任务，也正在于反映上述过程的一般规律并提高情报的社会利用率，即提高情报的利用效率。



图一 情报过程及其在人类生产活动中的作用

三十多年来，情报科学作为一门科学，伴随着现代科学技术的高速发展，经历了产生、技术发展和上升到理论的发展阶段。四十年代中期，是它的产生期。在这一时期，信息论和控制论的出现，为其开辟了美好的前景。它在整个科学技术发展中占据了

一定的位置；五十年代到六十年代，是它的发展期，为解决“情报爆炸”而发展起来的情报检索和情报系统，成为其发展中心。使情报工作开始成为一门服务性较强的行业；到七十年代，人们开始更多地关注理论研究工作，企图把它从“没有科学的科学”中摆脱出来，奠定牢固的理论基础，使其自立于科学之林。

从这一领域的科学论文发表的数量，也可看出目前发展概况。美国《情报科学和技术年译》创刊于1966年，到1975年的十年间的十卷中，就发表了约六千名学者的14,988篇论文。其中大约20%的作者，30%的论文是关于理论研究方面的。美国和苏联都编辑出版了情报学文摘，其中美国的《情报科学文摘》选用了255种杂志。苏联《情报科学文摘》选用了78种杂志。此外，还有《图书馆与情报科学文摘》等多种涉及情报科学的刊物。据统计全世界约有30个国家的60种杂志是情报科学的专业杂志，其中有十种是关于情报科学的理论研究的。

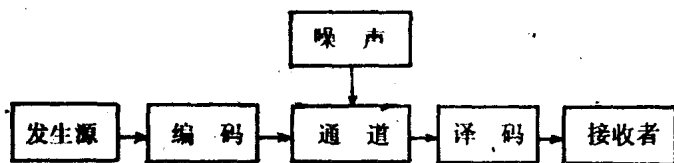
二、情报科学是关于科学技术 有效通讯的科学

1. 情报科学作为研究通讯问题的一个领域

通讯是一个极为复杂的过程，它不仅是一种涉及到人类相互交往所发生的社会现象，而且还被认为发生于人或动物的体内。历史上，从有学问开始，就有许多领域研究通讯的问题。到四十年代，由于信息论和控制论的出现，使得当时成为通讯发展的全盛时期而被誉为“通讯的时代”。这时出现了一类新的通讯领域，如信息论、控制论、计算机科学、博弈论和决策论、新语言学理论、广义系统理论等学科。而情报科学也正是在这一类通讯领域的影响下，在四十年代，特别是二次世界大战之后，受科学技术的发展、推动而产生和发展起来的一个通讯领域。它是关于知识

通讯的科学。

按照广义的通讯理论，通讯系统由下面四个部份组成：(1) 发生源；(2) 接收者；(3) 传送对象：消息或情报；(4) 传送媒介或载体。其模型如图二的框图所示。



图二 通讯的模型

消息和情报从发生源发出，通过收集、加工、存贮、传送通道，再经译码而被接收者所接收。

通讯，按照发生源和接收者之间的关系，分为直接通讯和间接通讯两种。直接通讯可认为是一次通讯，它产生于发生源与接收者直接接触时，通过传送媒介——语言来实现；间接通讯是二次通讯过程，在发生源与接收者之间插入了中间源或者特殊形式的情报载体——文献、照片、磁带等这些可载消息或情报的中间物。情报科学所研究的通讯问题包括上述二者或它们的组合的全部形式，但多指的是二次通讯系统。在 17 世纪，科学技术杂志一出现，便作为一种特殊的情报传送媒介或载体，成为了科学技术通讯的一种手段和通讯系统中的中间源。它有选择地并有系统地刊登科学技术论文，不仅对于科学技术成果的归纳积累，而且对于科学技术的传播借鉴，都起到了举足轻重的作用，解决了人类通讯的一个至为重要的方面——科学技术的通讯问题。科学技术知识或情报被某种语言表示，并记录在文献内，然后从通讯系统中的发生源送到接收者，便发生了通讯，而触发了接收者的行

为。通讯的效率取决于被传送的情报与接收者的需求之间的相关程度及消除非需要情报——即噪声的程度。科学工作者在研究工作结束后把结果写成论文，通过杂志等文献形式，向外传送，比传统的传播方式效率高得多了。随着科学的发展，文献的种类、数量繁增，质量益佳，文献变成了科学通讯的主要形式。四十年代特别是第二次世界大战后，由于生产发展和战争的需要，科学技术进入了一个空前发展的时期。科学技术的高速发展，科技文献每年的出版量成指数增长，产生了情报危机：

1) 文献出版量成指数增长。据统计，世界文献量每年约增长 7~8%，平均每 15 年翻一翻，某些尖端领域如原子能、空间技术，2~3 年就翻一翻。这样从大量的出版物中选择相关情报有如瀚海寻宝，海洋越大，越是困难。

2) 学科之间的界限不断突破，学科的专门化程度越来越高。学科的细分，学科内部的语言隔阂越大。同属于一个学科的不同专业，这个专业的人对另一专业竟成了外行。同时，学科之间又不断交叉而产生边缘学科。科学工作者的密切交往和了解已是加速科学发展进程至为重要的条件。科学通讯面临着不断改革的要求。

3) 由于科学通讯的不力，科技工作和情报文献重复量也增加了。日本物理学研究课题，据说在大学中重复率达 44%，民间 48%，国家研究机关 35%。浪费了资金，造成人力、物力、时间的不可估量的损失。情报文献的重复也是严重的。如美国的 AD 报告 60%，科学基金会 95% 的报告，又转换成论文在期刊上发表，这就更加大了文献的数量。

因此，就要求人们改善科学通讯的方式，提高通讯效率，加快情报过程。

科学技术的有效通讯，伴随着一类新的通讯领域的发展，引起了社会和各国政府的注意和重视。问题是显而易见的：要加速

发展经济和强大的国防，就必须以强大的科学技术作为后盾，科学技术的强有力发展，又在于科学技术情报的有效通讯。它是改善科学效率的中心。在这种背景下，作为研究科学有效通讯的一门学科——情报科学应运而生并发展起来了。四十年代和五十年代，主要是想通过情报检索解决“情报爆炸”的问题；其后，情报检索与网络技术相结合，形成大系统化；七十年代又开始构成所谓的“情报化工业”模式，情报科学的定量研究和实验，也发展了起来，并开始指导情报科学实践。

2. 情报科学研究的三大基本问题

综上所述，作为情报科学，它主要研究知识特别是公开知识或文献的通讯问题。那么，首先要研究的是一种共同的通讯过程，其次是文献载体。通讯过程又都是通过系统来实现的（一般叫通讯系统）。不同的通讯系统，有着利用不同情报特点的不同情报系统。即使对于相同的通讯系统，其结构和形式也不尽相同。所以，情报科学还研究情报系统的结构、目的、功能、特点等。因此，情报科学的研究对象主要有下述三个方面。

- 1) 通讯：通讯理论和实验的基础研究；
- 2) 文献：产生了文献计量学和科学计量学；
- 3) 情报系统（包括图书馆系统）：产生了情报检索原理和技术，以及情报网络。

1. 通讯问题

情报科学的通讯研究是指人类知识通讯的研究，而人类知识通讯是最重要的通讯过程。

本世纪五十年代，电子学与统计数学的结合产生了Shannon的信息论。按着信息论的观点，通讯分为三级：

第一级，技术问题：通讯信号从发生源传送到接收端的精度；

第二级，语义问题：所传送信号表达需求的意义的准确性问

题，它指对情报意义的共同理解程度。

第三级，系统的行为：即接收的意义按要求方式对接收者行为的影响程度，亦即效率。

信息论研究通讯系统中信息的传输过程，主要领域是通讯工程、计算机和电报等，它研究第一级的问题。在信息论中，信息（或情报）只限于信号（电信号）的特性上。消息为编成代码的信号的集合。一个消息中所含信息，可减少接收部分对于所传送消息的不肯定性。其信息量的大小，用在接到消息前对问题答案的不肯定程度（概率）来度量。这就产生了一个描写不肯定程度的判据“熵”。熵愈大，不肯定性愈大。信息熵（H）与各种可能出现的答案的概率 P_i 的关系是：

$$H = -P_1 \log P_1 - P_2 \log P_2 - P_3 \log P_3 \dots \\ P_N \log P_N \text{ 即 } H = -\sum P_i \log P_i$$

然而，情报（信息）并不只限于这种信号流，它还有另一种截然相反的属性，正如 Fairthorne 所指出的，情报类似于一种从海编中挤水似的那样的现象。这就是说，情报与接收者（部分）的预先状态不是没有关系，而是受到后者的深刻影响。例如，一本有关情报科学的书籍，情报专业的学生与其它专业的学生同样看过，取得的知识就不同，即是说，情报量取决于接收者已有的知识。语言学家们通过对情报意义的研究提出：“预先的知识可增加一个消息中的情报含量”。以这个理论为指导，当我们谈及情报价值时，光就情报本身来定量，就不那么完善了，因为情报用户的状态也是衡量情报价值的一个不可缺少的条件。

情报科学还主要研究情报通讯的效率。情报科学所阐述的基本问题是公开知识的通讯的效率。我们都知道，科学研究的一个重要特点，就是有选择地吸收所需要的那部份知识。这种选择甚至发生于科学研究和科学通讯的每一阶段。由于这种选择性，便

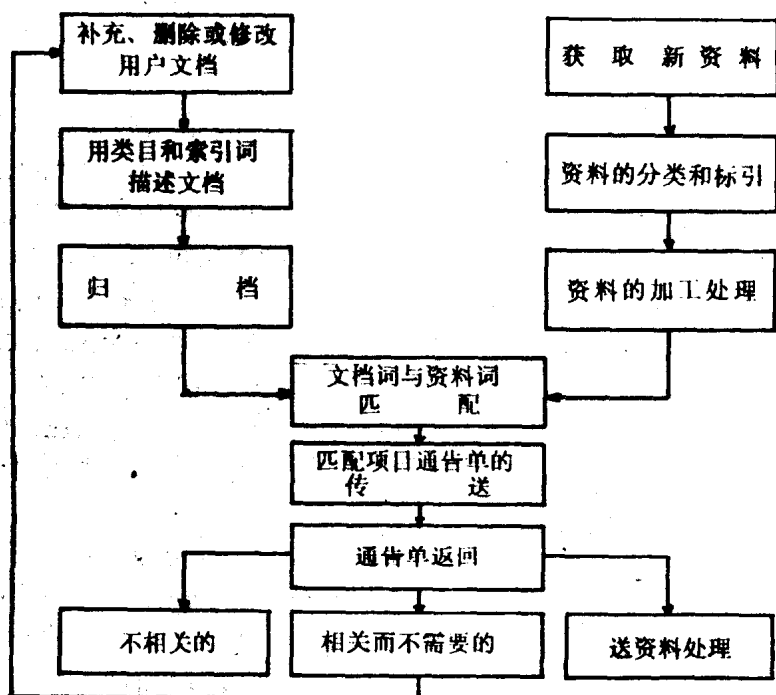
导致仅仅关注相关的情报。为了提高效率，科学通讯最好是仅仅处理相关的情报。随着情报量成指数增长和科研人员数量的增加，确定相关情报愈来愈困难。因而引出通讯系统中情报源与接收者接触效率的判据——相关率的概念，它可在通讯系统的许多点上测出，就是看接收者行动的改变情况。已出现了定量研究情报效率的重要进展。Mackay 提出，任何数据项之值都是接到数据后和接到数据前系统性能之比的以 10 底的对数。Charry 提出情报的判据是选择的预先概率与得到情报后概率之比的对数。Goffman 对系统非冗余情报进行了分析，提出了另一种情报的判据，它恰好满足 Shannon 信息熵定律。我国的情报工作在 1958 年提出了广、快、精、准的方针，正确地反映了情报效率的要求。但由于尚未提升到定量的水平，所以未能更好地发挥其积极的指导作用。当然，上述研究，并不十分成熟，尚有待于情报科学工作者的进一步突破。情报学的理论正在向定量方向发展，情报计量学 (informetrics) 吸引了越来越多的研究者。

在情报科学中，与相关率有关的知识通讯过程的研究，近二十多年来出现了两种截然不同的方向。一种是使用统计和数学的方法研究情报用户及其情报的需求。

文献与科学技术的发展有着极为密切的关系。它的普遍性规律，如文献增长、老化、分散和聚类等规律，正确地反映了科学发展的规律。了解和掌握这些规律性，对于改善科学通讯和提高情报系统的效率，是十分必要的。目前，对文献从定性到定量研究的飞跃，已形成了一门新的情报学的分支——文献计量学。它是研究记录知识特点和行为的一门科学。

情报的使用者对于情报的需求，是情报通讯的目标，它直接影响到情报源的选择，情报的贮藏，馆藏的完整性和排架方法，参考和咨询服务的方式，情报加工整理等整个情报通讯过程。情报用户及需求的研究主要通过社会调查的方法进行。例如有些研

究划分为四个方面:①读者的研究,包括图书资料目录的使用,参考咨询的分析;②各界(工程师,学生等)对于情报服务方式或设备的使用情况的研究;③用户的调查,哪些读者阅读或需要哪方面的情报;④不同专家情报需求的研究,根据不同的对象又可细分为机械工程师、化学家、城市规划专家、社会科学家等等。通过这种定量的研究,建立了不同模型,又反过来指导实践,大大地改善了情报传播的效率,改进了传统的服务方式。例如:近几年正在大力推行一种全新的情报传播方式——情报定题服务或者情报的选择传播(SDI),出现了SDI服务中心,其工作流程如图三所示。



图三 SDI 服务中心的工作流程

2. 文献问题

以前科学通讯主要靠图书，自十七世纪以后，科学期刊成为了科学技术传播的介质。由于图书期刊的时延性，而后又发展了索引、文摘、缩微胶卷、磁带等多种形式，发生了“不用纸的革命”。文献是一个综合的概念，文献计量学的任务就是专门定量地研究它的普遍规律。

在文献计量研究中最负盛名的要算是 Bradford 文献分散定律了。如果把一组期刊按照它们含有某一学科的论文数量的多少进行排列，再将这些期刊分为三个区，使每个区关于这一学科的论文篇数相等，就会看出核心区的期刊最少，而第二、第三区则以 $1:a:a^2$ (式中 $a \approx 5$) 的规律递增。从这一发现，可以找出文献密度最大的期刊来。如果多收藏这类期刊，情报效率显然是增加了。这个定律后来被许多人发展了，如 Vickery, Leimkuhler, Brookes 等。它的原理已被用于很多情报处理系统，大大改善了情报收集、贮藏及利用的效率。另一重要规律是半生期的发现，这就是著名的 Burton 和 Kepler 的文献老化规律。一些文献随着时间的发展，其情报内容被包含在别的文献中，或被后来的文献所超越，或者干脆失效，这就是所谓的老化。如果算一下科技文献中的一半是在多长时间内（半生期）发表的，则可得出一个规律来。半生期是情报老化速度的判据。表 1 列出了几个学科的文獻半生期，这个规律有着极大的应用价值，它指出了—个情报机构应尽量加工和传播在半生期内的文献，超过半生期的则可转到图书馆系统。

引文分析是研究文献规律性的另一条途径。引文一般出现在学术论文的后面，指出以前发表又被著者所借鉴的那些文章。引文与著文有着密切的联系。由此可见，引用同一篇文章的论文，也具有着某种关系。这种关系 Kessler 称为文献的耦合。如果有一组文章 G_A 是具有这种关系的，而与一篇给定的试验论文 P_o

又至少有一个耦合单位，则耦合关系的大小便可由每篇论文与Po之间耦合单位的多少测得。文献耦合的定量判据可用于指导哪些人读哪些文章。对于情报机构也可以知道把什么文章送给什么人。通过引文分析，还可确定某一学科中包含引文最多的期刊来，这是另一定律，称为期刊的聚类。Nairn用科学引文索引的期刊引文索引作工具，对每个期刊及其引用的期刊作了分析，找出了最常用期刊以及与它们有密切关系的期刊来，为人们有效和合理地利用期刊提供了依据。

表 1: 几个学科文献的半生期

学 科	半生期 (年)	学 科	半生期 (年)
冶金工程	3.9	物 理 学	4.6
化 工	4.8	机械工程	5.2
生 物 学	7.2	化 学	8.1
植 物 学	10.0	数 学	10.5
地 质 学	11.8		

3. 情报系统问题

在历史上，不同的时期有着不同的通讯和文献的问题，因而情报系统及其过程，也是不尽相同的。但从最早的图书馆到现代的数据库和情报检索网，以及它们的各种不同系统，都有着紧密的联系。当代，所有的情报系统，一个共同的要求，就是提高通讯和服务的效率。新的服务方式的出现，和情报系统中的新原理，如联机情报检索网，都可直接追溯到这个要求上。

一个情报系统，其作用就在于从现有的学科知识中选取所需要的情报，按某种方式进行存贮，最后有选择地将情报传送到接收者。其主要环节有：