

普通高等教育“十一五”国家级规划教材

情报学基础教程

Foundations of Information Science

叶 鹰 潘有能 潘 卫 编著

科 学 出 版 社

北 京

内 容 简 介

本书是普通高等教育“十一五”国家级规划教材。作为学习情报学的大学生、研究生以及科技情报、社科情报、图书情报工作人员编写的通识教材,本书的内容覆盖情报学理论与方法、情报管理与服务各相关领域,简明扼要地阐述了情报源、情报管理、情报检索、情报分析、情报技术等情报学核心分支的基本知识和新近发展。

本书配备多媒体教学课件等教学支持,适合作为大学生和研究生的教材及专业培训教材,也可供同等学力人员申请硕士学位学习参考。

图书在版编目(CIP)数据

情报学基础教程 / 叶鹰,潘有能,潘卫编著. —北京:科学出版社,2006
普通高等教育“十一五”国家级规划教材

ISBN 7-03-018043-7

I. 情… II. ①叶…②潘…③潘… III. 情报学-高等学校-教材
IV. G350

中国版本图书馆 CIP 数据核字(2006)第 108945 号

责任编辑:陈 亮 / 责任校对:邹慧卿

责任印制:张克忠 / 封面设计:陈 敬

科学出版社 出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

印刷

科学出版社发行 各地新华书店经销

*

2006 年 10 月第 一 版 开本:B5(720×1000)

2006 年 10 月第一次印刷 印张:19 1/2

印数:1—4 000 字数:377 000

定价:25.00 元

(如有印装质量问题,我社负责调换〈科印〉)



情报学作为一个独立的研究领域，是 20 世纪 50 年代开始确立的：50 年代初法语世界首先使用 L'informatique，1959 年英语世界正式提出 Information Science，从此，情报学逐步发展起来，至今仍处于动态发展之中。

如果在信息概念基础上认识情报，则情报是具有传递价值的信息；如果在知识概念基础上认识情报，则情报是具有传递价值的知识。因此，具有传递价值是情报的根本属性。也就是说，情报是针对特定对象的需要而提供的、在科学研究和各种活动中起继承、借鉴或参考作用的有价值的信息或知识。

情报学就是研究情报的产生、传递和利用规律的一门学科，现代情报学尤其关注利用现代化信息技术与手段，使情报交流过程、情报系统保持最佳效能状态，并帮助人们充分利用信息技术和手段提高情报产生、加工、贮存、检索、交流、利用的效率。

当然，国内外情报界对情报和情报学的认识多种多样，尚无统一标准，认为情报学“是研究情报传递工作的理论、方法和原理的学科”，“是研究如何开发、利用信息资源的规律及其技术、方法的学科”或“是研究人类如何创造、利用和交流各种形式的情报的学科”等观点均有代表性。但总的来看，情报学可以说是以情报工作的理论、原则和技术、方法为研究对象的学科。

作为一门近半个世纪才发展起来的新兴学科，情报学的主要研究任务包括：①揭示情报和情报工作的本质与特性；②探索情报工作的产生和发展规律；③探讨情报工作的内容环节及其在情报工作中的地位、作用，它们之间的内在联系以及各个环节的原则和方法；④研究各个时期情报工作的方针、政策、标准化的内容，以及现代化理论和技术；⑤阐明情报工作的科学管理理论与方法；等等。因此，情报学研究具有交叉学科或跨学科特点，是一门综合性的横断学科。中国情

报界也有人把“情报学”和“情报科学”看作两个不同的概念,认为情报学是研究情报工作的原理、工作方式与技术方法,情报的组织、管理、服务、评价及情报事业发展历史的学科。而情报科学则是综合研究由语言学、文字学、语文学、符号学、目录学、文献学、档案学、图书馆学、情报学等多学科知识的科学。

在中国,科技情报工作虽然早在 20 世纪 50 年代就与世界同时起步,但情报学作为一个独立的专业研究领域,却是 1978 年后随着科技情报专业教育的发展才得以确立的。因此,情报专业教育可以说是情报学的引擎。

为持续推动情报专业教育发展,浙江大学和上海交通大学的有关专业教师在吸收前辈老师和同辈学者的众多研究成果基础上,合作编写了这本《情报学基础教程》。

为避免繁杂的概念讨论,本书以三个简明扼要的科学化概念为基础:把信息作为物理实在并定义为信息密度的积分,将情报定义为有价值的信息(有效信息),再将知识定义为有效信息的积分。

在采用新的概念系统的同时,本书在体系组织方面的构造思路如下:

学习情报学,首先应当建立科学思维和量化观念,故将情报学经验规律作为第 1 章内容,这是文献计量学的一个子集,主要陈述布拉德福定律、洛特卡定律、齐夫定律、文献增长律、文献老化律的基本内容,并简要介绍其机理研究。

作为一门学科,应当有自己独立的基础理论体系,故第 2 章在扼要介绍国内外代表性情报学基础理论基础上,尝试用“一个轴心、三条原则、一组方程”构造情报学的基本理论轮廓和体系框架,使学生或读者能对情报学的总体结构有一个简明的整体把握。

情报学研究需要采用各种方法,除成熟的科学通用方法如逻辑分析法、数学分析法等外,也有本学科特有的引文分析法等独特方法,这构成本书第 3 章的核心内容。

情报源是情报工作和情报学研究的物质基础,故作为本书第 4 章。

围绕分类编目、主题标引、文摘索引等情报工作形成的情报管理不仅是情报工作的中心内容,而且是学习和研究情报学必不可少的基础,故在本书第 5 章进行系统陈述。

情报检索不仅是情报学的核心内容之一,而且是当代大学生、研究生、科研人员应当掌握的基本技能,故本书第 6 章对其理论基础和主要操作技能进行介绍。

情报分析也是情报学的核心内容之一,是最能体现情报学“研究”价值的分支,具有开拓潜力和发展前景,值得重视,这就形成本书第 7 章的实体内容。

伴随情报学进化,情报技术也日新月异地发展,尤其是在计算机技术、网络技术、通讯技术等组合推进下,情报技术正发生天翻地覆的变化,本书第 8 章

概要介绍现代主流情报技术。

此外，情报系统与情报服务、情报学与其他学科交叉形成的研究领域等内容也是情报学的重要组成部分，这就构成本书最后两章。

本书具有一定逻辑结构层次的编排希望能为初学情报学的学生或读者提供一个比较明晰的情报学概貌，适合用作研究生、大学生、专升本等类型学生的“情报学基础”或“情报学概论”类课程教材，特别适合具有其他学科背景的学生尤其是具有理、工、农、医专业背景的学生学习情报学使用，也可供图书情报工作人员参考。全书以业界公认的知识内容为主体，也适当增加了独立探索研究成果。本书特色是带有较强的科学实证思维方式，这也正是浙江大学和上海交通大学的学术传统，希望能有一定的启发价值。

为方便教师教学，本书配备了多媒体教学课件等立体化教学支持，凡选用本书作为教材的教师可与出版社取得联系，以获得相关支持。



前 言

第 1 章

情报学经验规律	1
1.1 布拉德福定律	1
1.2 洛特卡定律	7
1.3 齐夫定律	10
1.4 文献增长律	12
1.5 文献老化律	14
1.6 情报学经验规律的统一模型和机理研究	15
习题和思考题	17
主要参考文献	17

第 2 章

情报学基础理论	18
2.1 情报学概念系统	19
2.2 国外代表性情报学基础理论	29
2.3 国内代表性情报学基础理论	34
2.4 情报学基础理论新探	36

2.5 情报学理论原则	39
2.6 情报学相关学科	43
习题和思考题	44
主要参考文献	45

第3章

情报学方法论	46
3.1 情报学方法论概述	46
3.2 通用科学方法	48
3.3 情报学专门方法	70
习题和思考题	84
主要参考文献	84

第4章

情报源	86
4.1 情报源导引	86
4.2 文献情报源及其数字化资源	87
4.3 非文献情报源	103
4.4 情报源评价	109
习题和思考题	111
主要参考文献	112

第5章

情报管理	113
5.1 分类编目	113
5.2 主题标引	123
5.3 文摘索引	131
5.4 元数据	135
5.5 知识组织	143
习题和思考题	147

主要参考文献	147
--------------	-----

第 6 章

情报检索	148
6.1 情报检索理论基础和检索评价参数	148
6.2 全文数据库的发展及其检索利用	151
6.3 重要手工检索工具及其对应数据库	154
6.4 重要联机检索系统 DIALOG 及其操作	161
6.5 网络信息检索技术和主要搜索引擎	173
习题和思考题	183
主要参考文献	184

第 7 章

情报分析	185
7.1 情报分析导引	185
7.2 典型社科情报分析——竞争情报分析	198
7.3 典型科技情报分析——专利情报分析	213
习题和思考题	221
主要参考文献	221

第 8 章

情报技术	223
8.1 计算机技术基础	223
8.2 网络技术基础	228
8.3 数据库技术	235
8.4 数据仓库技术	242
8.5 数据挖掘技术	247
8.6 其他情报技术	252
习题和思考题	257
主要参考文献	257

第 9 章

情报系统与情报服务	258
9.1 从 MIS 到 ERP	258
9.2 从 DSS 到 CIS	266
9.3 从 OAS 到电子政务系统	271
习题和思考题	279
主要参考文献	279

第 10 章

情报学扩展	280
10.1 比较情报学	280
10.2 信息管理学	283
10.3 信息经济学	285
10.4 信息法学	289
10.5 信息伦理学	291
10.6 情报学发展方向	296
习题和思考题	298
主要参考文献	299

后记	300
----------	-----



第1章

情报学经验规律

经验规律是学科赖以建立的基石，也是定量化研究的起点。情报学虽然是一门“年轻”的学科，但在经验规律方面已有一定积累，其基础主要由下列文献计量规律（Egghe and Rousseau，1990）奠定。

1.1 布拉德福定律

布拉德福定律是描述专业论文在期刊中分布情况的经验规律，由英国著名文献学家布拉德福（S. C. Bradford，1878—1948）于1934年提出。

1.1.1 基础数据和规律表述

布拉德福毕业于英国伦敦大学，所学专业是化学，1922年获科学博士学位，但他热爱图书馆工作和图书馆事业，1925～1937年曾长期担任英国南肯辛顿科学图书馆馆长。在工作中，他发现一个学科的专业论文常分散发表在各种期刊杂志上，于是他以应用地球物理学和润滑专业为实例，研究了期刊论文分布的情况，提出了一个描述文献分散的经验定律。布拉德福采集的1929～1932年应用地球物理学原始数据见表1-1。

表 1-1 布拉德福原始数据

期刊数	相关论文数	累积期刊数	累积论文数	期刊数	相关论文数	累积期刊数	累积论文数
1	93	1	93	1	56	3	235
1	86	2	176	1	48	4	283

续表

期刊数	相关论文数	累积期刊数	累积论文数	期刊数	相关论文数	累积期刊数	累积论文数
1	46	5	329	5	10	27	662
1	35	6	354	3	9	30	689
1	28	7	392	8	8	38	753
1	20	8	412	7	7	45	802
1	17	9	429	11	6	56	868
4	16	13	493	12	5	68	928
1	15	14	508	17	4	85	996
5	14	19	578	23	3	108	1065
1	12	20	590	49	2	157	1163
2	11	22	612	169	1	326	1332

数据来源: Bradford S C. Documentation. Washington, D. C.: Public Affairs Press, 1950; 转引自 Garfield E. Bradford's law and related statistical patterns. Current Contents, 1980 (19): 5-12

如果将以上数据分成如下 3 个区域:

分区	期刊数	相关论文数	每刊年载相关论文数
C	9 (n_c)	429	>4
1	59 (n_1)	499	1~4
2	258 (n_2)	404	1

则近似有:

$$n_c : n_1 : n_2 = 1 : a : a^2, \quad a \approx 5$$

显然这只是一条近似规律。

用文字表述就是: 如果将期刊按其刊载某专业论文数量的多寡以递减顺序排列, 则可分出一个核心区和相继的几个领域, 当每区刊载的论文量相等时, 核心期刊数 n_c 和外围一区期刊数 n_1 、外围二区期刊数 n_2 成 $n_c : n_1 : n_2 = 1 : a : a^2$ 关系。其中, a 称为布拉德福常数。

若以累积期刊数 n 的对数为横坐标、累积相关论文数 $R(n)$ 为纵坐标, 可作出如图 1-1 所示的曲线。

因此, 后来布鲁克斯 (B. C. Brookes) 用公式将布拉德福定律表述为:

$$R(n) = \begin{cases} \alpha n^\beta, & (1 \leq n \leq n_c) \\ k \lg(n/s), & (n_c \leq n \leq N) \end{cases} \quad (1-1)$$

其中, $R(n)$ 为相关论文累积数; n 是期刊累积数; α 是 $n=1$ 时的 $R(n)$; β, k, s

为参数； N 为期刊总数。曲线右上端的弯折称为格鲁斯（Q. V. Groos）下降。

1.1.2 改进和扩展

布拉德福 1934 年提出上述经验规律后，并未引起人们注意。1948 年，其专著《文献学》（*Documentation*）出版，在第 9 章中全文收录了 1934 年的发现，才引起另一位英国文献学家维克

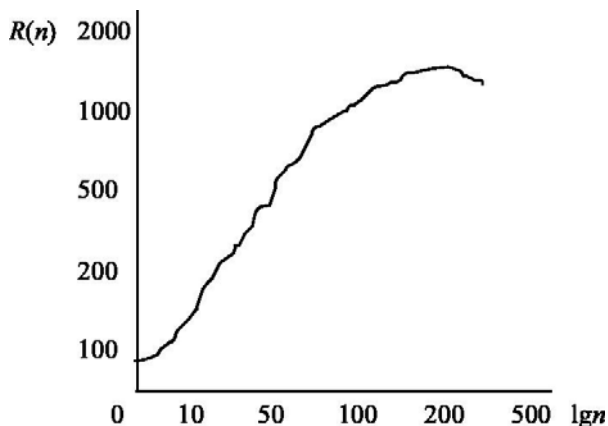


图 1-1 布拉德福分布曲线示意

利（B. C. Vickery）的重视。维克利对之进行研究后，将外围区扩展到多个区域，并首次用“布拉德福分散定律”命名这一经验规律（Vickery, 1987）：

把期刊按照刊载某学科相关论文的数量排列，可得到专门登载该学科论文的核心区期刊和若干外围区期刊，当各区所含该学科相关论文相等时，各区期刊数成

$$n_c : n_1 : n_2 : \dots = 1 : a : a^2 : \dots \quad (a > 1) \quad (1-2)$$

其中， a 是布拉德福常数。

还有一个改进称为维克利推论：

$$n_c : (n_c + n_1) : (n_c + n_1 + n_2) : \dots = 1 : b : b^2 : \dots \quad (b > 1) \quad (1-3)$$

其中， b 叫做维克利系数。

1977 年，苏联情报学家斯莫里科夫（И. А. СМОЛЬКОВ）对布鲁克斯的公式描述进行研究后，提出用一个统一的方程来表述布拉德福定律：

$$R(n) = K \lg(q_1 n + q_2 e^{-\beta n}) \quad (1-4)$$

其中， $R(n)$ 为相关论文累积数； n 是期刊累积数； K ， β ， q_1 ， q_2 为参数。

1.1.3 应用提示

布拉德福定律的主要用途是用于确定核心期刊，以指导期刊订购和期刊利用，并由此扩展到核心馆藏维护、核心检索工具选择等。有人认为布拉德福定律可以看作是社会科学中普遍存在的“二八律”的一种表现：20% 的核心期刊上刊载了 80% 的重要论文。此外，布拉德福定律还可应用于核心出版社的研究等。

上海交通大学王国龙等通过对 SCI 和 SSCI 引用期刊的研究，发现影响因子（IF, impact factor）的分布也符合布拉德福定律。他们的数据如表 1.2～表 1.5 所示^①。

^① 王国龙，李佩．SCI 源期刊与 JCR．上海交通大学学报，2003（增刊）：287-290；王国龙．SSCI 源期刊概述．中国索引，2004，2（2）：20-25

SCI 源期刊数据见表 1-2。

表 1-2 SCI 源期刊数据 (2001 年)

序号	国家/地区	期刊种数	影响因子 (IF)	期刊所占 比例 (%)	序号	国 家 名	期刊种数	影响因子 (IF)	期刊所占 比例 (%)
1	美国	2220	1.83	38.60	30	比利时	13	0.43	0.23
2	英格兰	1139	1.54	19.80	31	斯洛伐克	12	0.38	0.21
3	荷兰	565	0.94	9.82	32	芬兰	11	0.75	0.19
4	德国	432	0.90	7.51	33	以色列	10	0.63	0.17
5	瑞士	154	1.03	2.68	34	克罗地亚	10	0.18	0.17
6	法国	148	0.59	2.57	35	墨西哥	7	0.37	0.12
7	日本	147	0.57	2.56	36	乌克兰	7	0.22	0.12
8	俄罗斯	99	0.31	1.72	37	智利	6	0.47	0.11
9	加拿大	76	1.32	1.32	38	罗马尼亚	6	0.23	0.11
10	意大利	69	0.76	1.20	39	阿根廷	6	0.04	0.11
11	丹麦	60	1.24	1.04	40	威尔士	5	0.42	0.09
12	澳大利亚	59	0.65	1.03	41	希腊	4	1.48	0.07
13	中国	57	0.37	0.99	42	委内瑞拉	4	0.15	0.07
14	印度	45	0.22	0.78	43	南斯拉夫	3	0.27	0.05
15	波兰	37	0.52	0.64	44	伊朗	3	0.18	0.05
16	挪威	34	0.97	0.59	45	土耳其	3	0.12	0.05
17	新加坡	29	0.67	0.50	46	沙特	3	0.11	0.05
18	苏格兰	26	0.87	0.45	47	孟加拉国	3	0.10	0.05
19	西班牙	26	0.54	0.45	48	斯洛文尼亚	3	0.04	0.05
20	奥地利	25	0.94	0.44	49	埃塞俄比亚	2	0.14	0.03
21	捷克	22	0.41	0.38	50	肯尼亚	2	0.13	0.03
22	新西兰	21	1.31	0.37	51	保加利亚	2	0.12	0.03
23	南非	19	0.39	0.33	52	巴基斯坦	2	0.08	0.03
24	瑞典	18	1.04	0.31	53	泰国	1	0.38	0.02
25	韩国	18	0.59	0.31	54	拉脱维亚	1	0.37	0.02
26	匈牙利	18	0.29	0.31	55	乌兹别 克斯坦	1	0.32	0.02
27	爱尔兰	16	1.03	0.28	56	爱沙尼亚	1	0.26	0.02
28	中国台湾	16	0.54	0.28	57	牙买加	1	0.26	0.02
29	巴西	15	0.35	0.26					

续表

序号	国家/地区	期刊种数	影响因子 (IF)	期刊所占 比例 (%)	序号	国家名	期刊种数	影响因子 (IF)	期刊所占 比例 (%)
58	乌克兰	1	0.20	0.02	62	特立尼达 和多巴哥	1	0.08	0.02
59	哥斯达黎加	1	0.10	0.02					
60	北爱尔兰	1	0.09	0.02	63	厄瓜多尔	1	0.04	0.02
61	科威特	1	0.08	0.02	64	古巴	1	0.02	0.02
	其他	3		0.05	总计		5752	1.44	100

根据表 1-2 数据归纳整理出的结论见表 1-3。

表 1-3 SCI 源期刊 IF 的布拉德福分布

布拉德福分区	核心区	准核心区	外围区
期刊 (种)	328	1173	4251
IF 的范围	4.14~46.23	1.56~4.13	<1.56
总影响因子 (TIF)	2760.3	2758	2759.7

SSCI 源期刊数据见表 1-4。

表 1-4 SSCI 源期刊数据 (2002 年)

序 号	国家/地区	期刊种数	影响因子 (IF)		期刊所占比例 (%)
			总 数	平 均	
1	美国	1003	968	0.97	58.69
2	英格兰	386	291.6	0.76	22.59
3	荷兰	99	75.4	0.76	5.79
4	德国	54	28.5	0.53	3.16
5	加拿大	29	13.98	0.48	1.70
6	法国	18	5.6	0.31	1.05
7	瑞士	17	13.8	0.81	0.99
7	澳大利亚	17	9.55	0.56	0.99
9	日本	9	2.30	0.26	0.53
10	俄罗斯	7	1.62	0.23	0.41
11	挪威	6	2.97	0.50	0.35
12	苏格兰	6	2.61	0.44	0.35
13	捷克	4	0.74	0.18	0.23

续表

序 号	国家/地区	期 刊 种 数	影响因子 (IF)		期刊所占比例 (%)
			总 数	平 均	
13	新西兰	4	0.93	0.23	0.23
13	爱尔兰	4	2.40	0.60	0.23
13	墨西哥	4	1.18	0.30	0.23
17	瑞典	3	0.82	0.27	0.18
17	中国	3	0.05	0.02	0.18
17	南非	3	1.08	0.36	0.18
17	奥地利	3	1.23	0.41	0.18
17	印度	3	1.40	0.37	0.18
22	克罗地亚	2	0.40	0.20	0.12
22	比利时	2	1.11	0.56	0.12
22	丹麦	2	2.32	1.16	0.12
22	韩国	2	0.27	0.14	0.12
22	巴西	2	0.50	0.25	0.12
22	西班牙	2	1.25	0.63	0.12
22	斯洛伐克	2	0.24	0.12	0.12
29	中国台湾	1	0.29	0.29	0.06
29	斯洛文尼亚	1	0.30	0.30	0.06
29	新加坡	1	0.04	0.04	0.06
29	芬兰	1	0.08	0.08	0.06
29	以色列	1	0.30	0.30	0.06
29	意大利	1	0.78	0.78	0.06
29	土耳其	1	0.28	0.28	0.06
29	哥伦比亚	1	1.10	1.10	0.06
29	智利	1	0.47	0.47	0.06
29	葡萄牙	1	0.29	0.29	0.06
29	阿根廷	1	0.10	0.10	0.06
29	南斯拉夫	1	0.14	0.14	0.06
	其他	1			0.06
总计		1709	1436.1	0.84	100

根据表 1-4 数据归纳整理出的结论见表 1-5。

表 1-5 SSCI 源期刊 IF 的布拉德福分布

布拉德福分区	核心区	准核心区	外围区
期刊(种)	197	436	1076
IF 的范围	1.676~11.622	0.789~1.670	<0.789
总影响因子(TIF)	478.8	478.6	478.7

这一研究将布拉德福定律的应用范围扩展到影响因子,具有参考意义。

1.2 洛特卡定律

洛特卡定律是描述作者与论文数量之间关系的经验规律,由美国情报学家洛特卡(A. J. Lotka, 1880—1949)于1926年提出。

1.2.1 基础数据和规律表述

洛特卡的原始研究是对物理学和化学两大学科中科学家们发表论文的情况进行统计分析。对于物理学,他使用德国奥尔巴赫(Aürbach)《物理学史一览表》(Geschichtstafelen der Physik, 1910)的人名索引中,对其全部1325位作者进行统计分析,依次列出发表1篇、2篇和多篇论文的作者数;对于化学,他采用《化学文摘》(Chemical Abstracts, 1907~1917)十年累积索引中的姓氏以A、B开头的6891位作者进行了统计,分别列出发表1篇、2篇、3篇直至346篇论文的作者数。结果发现物理学1325位作者中,发表1篇论文的为784人,占59.2%;化学A字母姓氏作者1543人,发表1篇论文的为890人,占57.7%;B字母姓氏作者5348人,发表1篇论文的为3101人,占57.98%。于是他提出作者的百分比分布应符合如下公式:

$$f(x) = \frac{c}{x^a} \quad (1-5)$$

其中, $f(x)$ 是发表 x 篇论文的作者占作者总数的百分比(作者频率),常数 $a > 1$ 。上式取对数,有

$$\log f(x) = -a \log x + \log c \quad (1-6)$$

故以 $\log f(x)$ 为纵坐标、 $\log x$ 为横坐标作图应是一斜率为 $-a$ 、截距为 $\log c$ 的直线。

洛特卡的原始数据中,Aürbach数据部分见表1-6(计算百分比是根据洛特卡平方反比律计算所得)。

表 1-6 洛特卡原始数据片段

论文数	统计数据			作图数据	
	撰写人数	实际百分比	计算百分比	数据点个数	斜率
总计	1325				
1	784	59.17	60.79	13	-2.0411
2	204	15.40	15.20	14	-2.0533
3	127	9.58	6.75	15	-2.0182
4	50	3.77	3.80	16	-2.0255
5	33	2.49	2.43	17	-2.0210
6	28	2.11	1.69	18	-2.0953
7	19	1.43	1.24	19	-2.1385
8	19	1.43	0.95	20	-2.0786
9	6	0.45	0.75	21	-2.0726
10	7	0.53	0.61	22	-1.9887
11	6	0.45	0.50	23	-1.9989
12	7	0.53	0.42	24	-1.9946
13	4	0.30	0.36	25	-1.9754
14	4	0.30	0.31	26	-1.9490
15	5	0.38	0.27	27	-1.8485
⋮	⋮	⋮	⋮	⋮	⋮

数据来源：邱均平．文献计量学．北京：科学技术文献出版社，1988，175-177，181

基于这些数据绘制的 $\log f(x) \sim \log x$ 直线如图 1-2 所示：

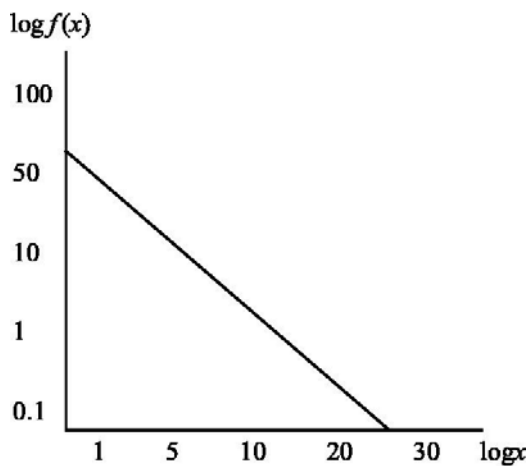


图 1-2 洛特卡分布曲线示意

既然大部分数据点（尤其是前 13~25 个点）获得的斜率都接近 -2，故洛特卡定律近似平方反比律：

$$f(x) = \frac{c}{x^2} \quad (1-7)$$

其中， c 是常数。

由于 $1 = \sum_{n=1}^{\infty} f(n) = \sum_{n=1}^{\infty} \frac{c}{n^2} = c \frac{\pi^2}{6}$ ，故 $c = \frac{6}{\pi^2} \approx 60.79\%$ ， $f(1) = \frac{c}{1^2} = c$ ，进而有：

$$f(2) = \frac{c}{2^2} = \frac{f(1)}{2^2} \quad (1-8)$$

.....

$$f(n) = \frac{c}{n^2} = \frac{f(1)}{n^2} \quad (1-9)$$

也就是说：发表 1 篇论文的作者约占作者总数的 60.79%，发表 2 篇论文的作者占发表 1 篇论文作者数量的 1/4，……，发表 n 篇论文的作者占发表 1 篇论文作者数量的 $1/n^2$ 。

洛特卡定律描述了作者人数与其发表论文数量之间的关系，首次揭示了两者之间存在的规律。后经研究，发现物理学等学科领域的作者与论文之间的关系基本符合平方反比律，而生物、工程、计算机等领域则不符合平方反比关系。一般来说，人文学科、社会科学中， a 值将变大；规模较大、科研合作程度较高的学科中， a 值会变小。

1.2.2 改进与扩展

虽然洛特卡定律早在 1926 年就提出，但直到 1949 年才被称为“洛特卡定律”。后来，不少学者深入探讨了洛特卡定律的形成机理和改善表达，著名科学家普赖斯（D. Price）就在洛特卡定律基础上进一步研究了科学家人数和科学论文数量之间的关联，在《小科学，大科学》一书中提出了“普赖斯定律”：科学家总数是杰出科学家人数的平方（或者说杰出科学家人数是科学家总数的平方根）。普赖斯定律的另一定量表达是：杰出科学家发表了全部科学论文的 50%。

设 $n_{\max}$ 为杰出科学家中最高产作者发表论文数， m 为杰出科学家中最低产作者发表论文数， R 为杰出科学家人数与全体科学家总数之比，则普赖斯推导出了以下两式：

$$m = 0.749 \sqrt{n_{\max}} \quad (1-10)$$

$$R = 0.812 / \sqrt{n_{\max}} \quad (1-11)$$

也就是说,杰出科学家中最低产作者发表论文数是最高产作者发表论文数平方根的 0.749 倍;杰出科学家人数与全体科学家总数之比是杰出科学家中最高产作者发表论文数平方根倒数的 0.812 倍。

1.2.3 应用提示

洛特卡定律主要用于研究“科学生产率”,可用于预测发表不同篇数论文的作者数量和特定学科的论文总量,或根据作者数量估计科学论文数量等。

为合理评价洛特卡定律的适用性,美国情报学家科尔(R. C. Coile)于 1977 年提出用 K-S (Kolmogorov-Smirnov) 检验法对其进行鉴定,其步骤为:

(1) 设 A 为统计的作者总数,计算 K-S 值: $KS=1.63/\sqrt{A}$;

(2) 计算最大偏差值 D : $D=\max |F_0(x)-S_n(x)|$; 其中, $F_0(x)$ 为累积作者频率理论值, $S_n(x)$ 为累积作者频率观察值;

(3) 比较 D 与 KS : 若 $D < KS$, 则抽样分布符合洛特卡定律; 若 $D > KS$, 则抽样分布不符合洛特卡定律。

科尔用该方法检测了洛特卡《物理学史一览表》和《化学文摘》的原始数据,发现《物理学史一览表》的数据符合洛特卡定律,而《化学文摘》的数据则不完全符合洛特卡定律。科尔还用该方法对美国国会图书馆的 MARC 作者统计数据 and 伊利诺大学图书馆学院卡片目录作者统计数据进行了检验,结果发现美国国会图书馆的 MARC 作者统计数据不符合洛特卡定律,而伊利诺大学图书馆学院卡片目录作者统计数据则完全符合洛特卡定律。K-S 检验法从此成为检验洛特卡定律适用性的有效方法。

国内也有不少洛特卡定律研究成果,如福建农林大学张贤澳围绕洛特卡定律发表过多篇论文^①,并出版有专著《广义洛特卡定律——估计、推论及其在管理中的应用》,可供参考。

1.3 齐夫定律

齐夫定律是描述文献中的词与其出现频次之间关系的经验规律,由齐夫(G. K. Zipf)于 1935 年提出,是一条与语言学密切相关的文献学规律。

1.3.1 基础数据和规律表述

1935 年,哈佛大学语言学教授齐夫统计了乔伊士(J. Joice)的中篇小说《尤

^① 参见刘婧. 文献作者分布规律研究——对近十五年来国内洛特卡定律、普赖斯定律研究成果综述. 情报科学. 2004, 22 (1): 123-128

利西斯》(Ulysses) 中使用的 29 899 个词汇, 按使用频率(词频) 排列成序, 并对应编定词级——使用频率最高的词汇的词级为 1, 依次类推, 则词级和词频的乘积接近于常数。他的统计数据见表 1-7。

表 1-7 齐夫定律原始数据

词级 r	词频 f	词级 $r \times$ 词频 f	词级 r	词频 f	词级 $r \times$ 词频 f
10	2653	26 530	500	50	25 000
20	1311	26 220	1000	26	26 000
30	926	27 780	2000	12	24 000
40	717	28 680	3000	8	24 000
50	556	27 800	4000	6	24 000
100	256	25 600	5000	5	25 000
200	133	26 600	10 000	2	20 000
300	84	25 200	20 000	1	20 000
400	62	24 800	29 899	1	29 899

数据来源: Zipf G K. Human Behavior and the Principle of Least Effort. New York: Hafner, 1972; 转引自 Garfield E. Bradford's Law and Related Statistical Patterns. Current Contents, 1980 (19): 5-12

如果以 $\lg f \sim \lg r$ 作图, 则为一直线, 如图 1-3 所示。

齐夫在 1949 年出版的《人类行为与最省力法则》(*Human Behavior and the Principle of Least Effort*) 专著中, 更系统地将他的发现整理为一条规律:

设有一篇含有 N 个词的文献 ($N \geq 5000$), 用自然数 1, 2, 3, ... 给文献中的词编级, 出现频次最高的为 1, 其次为 2, 直到 r ($r < N$), 则高频词的频次 f 与词级 r 的乘积是一个常数:

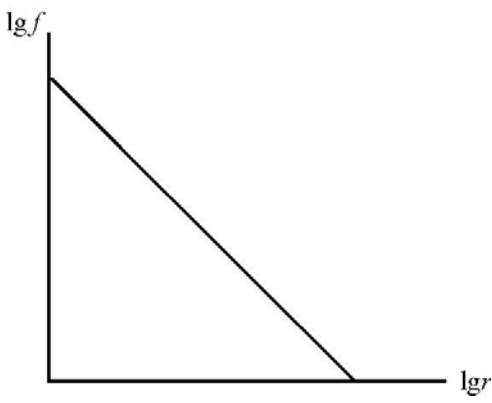


图 1-3 齐夫分布示意图

$$fr = c \tag{1-12}$$

其中, $0 < c < 0.1$ 。这就是齐夫定律。齐夫本人用最省力法则作为该定律的机理。

1.3.2 改进和扩展

1936 年, 美国语言学家朱斯 (M. Joos) 提出含两个参数的齐夫公式:

$$fr^b = c \tag{1-13}$$

这是双参数词频分布律, 其中, $b > 0$, $c > 0$ 。当 $b = 1$ 时, 等价于齐夫单参数

公式。

1952年,美籍法国数学家曼德布罗(B. Mandelbrot)运用信息论原理和概率论方法研究了词频分布规律,重新解释了齐夫定律,提出了三参数词频分布律:

$$f(r+a)^b = c \quad (1-14)$$

其中, $0 \leq a < 1$, $b > 0$, $c > 0$ 。该公式当 $a=0$ 时,就是朱斯双参数公式;当 $a=0$, $b=1$ 时,就是齐夫单参数公式,因而三参数公式具有更广的适用性。

但以上改善均针对高频词,低频词的分布规律是由布斯(B. Booth)首先导出的:

设出现1次的词数量为 I_1 , 出现 n 次的词数量为 I_n , 则低频词满足公式:

$$\frac{I_n}{I_1} = \frac{2}{n(n+1)} \quad (1-15)$$

式(1-15)也称为齐夫第二定律,式(1-12)相应地称为齐夫第一定律。

1.3.3 应用提示

齐夫定律主要可用于指导文献标引、词表编制和情报检索等,尤其是在自动标引中具有应用价值,因为使用频率太高的词和使用频率太低的词都没有多少检索价值,故应当使用词频适中、词义强健的词作为标引词。

1.4 文献增长律

文献增长律是描述文献信息量随时间而不断增加的规律,主要有以下模型。

1.4.1 指数曲线

早在1949年,普赖斯(D. Price)就发现“十年一叠的《哲学汇刊》靠墙竟堆成了一条完美的指数曲线”。1950年,他发表了第一篇关于“指数增长”的论文。随后,他对1665年《哲学汇刊》创刊以来世界科学杂志的增长数据、《化学文摘》、《生物学文摘》、《科学文摘》等摘录的期刊论文数量进行了统计,在1961年正式出版的《巴比伦以来的科学》中,系统地提出了科技文献按指数规律增长的经验规律,内容如下:

设 t 时间的文献信息数量为 $Q = F(t)$, 则有:

$$Q = F(t) = ae^{bt} = a\exp(bt) \quad (1-16)$$

其中, $a > 0$ 为条件常数,代表 $t=0$ 时刻的 Q ; $b > 0$ 为时间常数,代表持续增长率; $e=2.718$ 。 t 通常以年计算。

支持文献指数增长规律的部分数据参见表1-8。

用表 1-8 数据, 以公元年代作为横坐标时间 t , 以期刊数作为纵坐标文献量 $F(t)$, 作成图 1-4 曲线, 即普赖斯曲线:

表 1-8 世界期刊增长数据

年代	期刊种数
1750	10
1800	100
1850	1000
1900	10 000
1950	50 000
2000	100 000

数据来源: 根据多种信息渠道汇总估计

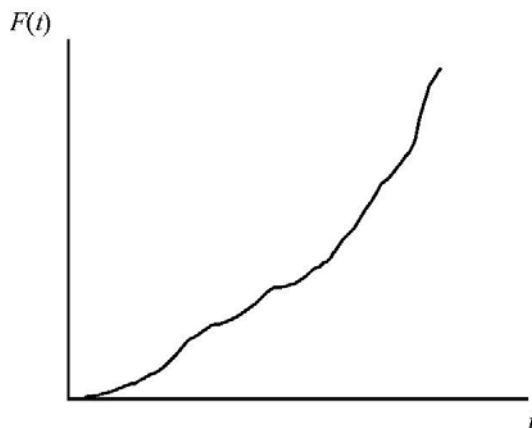


图 1-4 普赖斯曲线示意图

按照指数增长律, 如果初始文献量 $a=1000$ (件), $b=10\%$, 则:

10 年后文献量 $F(10) = 1000\exp(0.1 \times 10) = 2718$ (件)

100 年后文献量 $F(100) = 1000\exp(0.1 \times 100) = 22\,026\,466$ (件)

因此, 有“信息爆炸”之说。然而, 实际上文献信息的增长是有限的, 这表明指数增长律只是一条近似规律, 故产生以下改进模型。

1.4.2 逻辑曲线

苏联科学学家纳里莫夫 (В. Налимов) 等在仔细研究指数增长律后, 发现文献增长是分阶段进行的, 在经过一段急剧增长过程后, 会有一段缓慢增长, 也就是指数曲线会变为逻辑曲线 (一种生长曲线), 其数学模型为:

$$Q = F(t) = \frac{k}{1 + ae^{-bt}} = \frac{k}{1 + a\exp(-bt)} \quad (1-17)$$

其中, k 为最大上界即 $t \rightarrow \infty$ 时的文献累积量。其形状如图 1-5 所示。

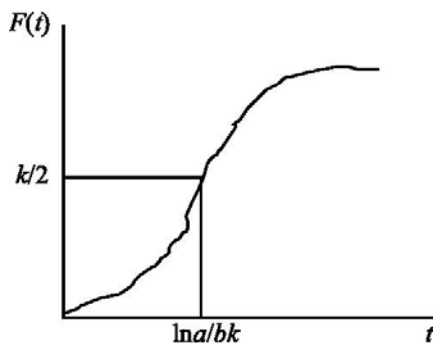


图 1-5 逻辑曲线示意图

逻辑曲线模型表明文献增长存在一个最大上界, 这也不符合发展实际, 因而也只是一条近似规律。

1.4.3 勒歇尔模型

美国科学史家和情报学家勒歇尔 (N. Rescher) 1978 年在《科学的进步》一书中引进文献质量等级指标 $\lambda \in [0, 1]$, 将普赖斯指数增长定律改进为:

$$F(t) = (ae^{bt})^\lambda \quad (1-18)$$

并约定： $\lambda=1 \rightarrow$ 常规文献

$\lambda=3/4 \rightarrow$ 有意义文献

$\lambda=1/2 \rightarrow$ 重要文献（“重要文献是文献总数的平方根”）

$\lambda=1/4 \rightarrow$ 非常重要文献

$\lambda=0 \rightarrow$ 此时定义 $F(t) = \ln a + bt$ ，是最重要文献

勒歇尔模型延缓了重要文献增长的速率，并让最重要文献呈线性增长，改善了文献增长律，深化了对文献指数增长的认识。

1.5 文献老化律

文献老化律是描述文献信息价值随时间推移而不断下降的规律，主要有以下模型。

1.5.1 负指数老化律

与指数增加律对应，负指数老化律的数学模型如下

$$C(t) = ke^{-dt} \quad (1-19)$$

其中， $C(t)$ 代表发表 t 年后的文献被引次数， k 是因学科而异的学科常数， d 为老化率。其示意图见图 1-6。

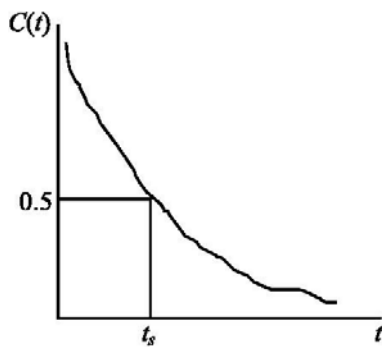


图 1-6 负指数老化曲线示意

图 1-6 中，与 $C(t) = 0.5$ 相对应的时间 t_s 称为相应学科文献的半衰期或半生期，其含义是经过 t_s 年后已有 50% 的文献不再被引用、或者说只有 50% 的文献仍在被引用。

1.5.2 巴尔顿—凯卜勒老化方程

美国图书馆员巴尔顿 (R. E. Burton) 和凯卜勒 (R. W. Keblér) 合作对科技文献的老化进行研究后，于 1960 年提出如下老化方程：

$$y = 1 - \left[\frac{a}{e^x} + \frac{b}{e^{2x}} \right] \quad (1-20)$$

其中， y 是经过一定时间尚在利用的文献的相对数量（ $y=0.5$ 对应半生期）， x 是以 10 年为单位计量的时间， $a+b=1$ 为常数。

巴尔顿-凯卜勒具体测定了一些学科的半生期，参见表 1-9。

表 1-9 一些学科文献的半生期

学 科	半生期	学 科	半生期
生物医学	3.0	生理学	7.2
冶金工程	3.9	化 学	8.1
物 理 学	4.6	植物学	10.0
化学工程	4.8	数 学	10.5
社 会 学	5.0	地质学	11.8
机 械 工 程	5.2	地理学	16.0

数据来源：邱均平．文献计量学．北京：科学技术文献出版社，1988，94

苏联学者莫捷列夫 (B. M. Мотылев) 仔细研究巴尔顿-凯卜勒老化方程后，发现实际统计数据与理论计算数据在 χ^2 检验时差异显著，于是提出了如下修正公式：

$$y = 1 - \left[\frac{a}{e^{x-0.1}} + \frac{b}{e^{2x-0.2}} \right] \quad (1-21)$$

并用下式计算半生期：

$$x_s = 10 \times [\ln(a + \sqrt{a^2 + 2b}) + 0.1] \quad (\text{年}) \quad (1-22)$$

这些成果对深化文献老化律的研究具有重要意义。

1.6 情报学经验规律的统一模型和机理研究

由于布拉德福定律 $R(n) = k \log(n/s)$ 对 n 微分可得： $\frac{dR(n)}{dn} = \frac{ks}{n} = \frac{c}{n}$ ，因此，布拉德福定律与洛特卡定律 $f(x) = \frac{c}{x^\alpha}$ 和双参数齐夫定律 $f = \frac{c}{r^b}$ 可合并成统一的形式：

$$f(n) = \frac{C}{n^\alpha} \quad (1-23)$$

其中， c, α 均为常数， $1 \leq \alpha \leq 2$ 。 $f(n)$ 为分布密度。这就是所谓布-洛-齐分布的统一机理。

如果要用一个数学模型来统一布-洛-齐分布和文献增长律、文献老化律，则可用如下具有普遍意义的信息分析方程组来描述（叶鹰，1999）：

$$\left. \begin{aligned} \frac{\partial^2 I}{\partial t^2} - c^2 \Delta^2 I &= \frac{c}{p} I \\ \frac{\partial I}{\partial t} - k \Delta^2 I &= \frac{k}{pb} I \\ \frac{\partial I}{\partial t} - \Delta I &= a \end{aligned} \right\} \quad (1-24)$$

其中 c 为波速, b 为媒体常数, a 为接触系数, k 为媒体的信息扩散率, p 为信息压缩率, t 为时间。信息空间除物理空间外, 也包括由学科、语种、地域等“空间”型变量构成的信息空间。式 (1-24) 中第一个方程描述情报的载波传播, 第二个方程描述情报的扩散传播, 第三个方程描述情报的接触传播。

在信息科技和现代社会中, 起主导作用的情报传播模式是载波传播和扩散传播, 故基本信息分析方程组由式 (1-24) 中的前二式构成:

$$\frac{\partial^2 I}{\partial t^2} - c^2 \Delta^2 I = \frac{c^2}{p} I \quad (1-25)$$

$$\frac{\partial I}{\partial t} - k \Delta^2 I = \frac{k}{pb} I \quad (1-26)$$

当只考虑信息随时间变化规律时, 用 $\frac{1}{c} \times$ 式 (1-25) $-\frac{1}{k} \times$ 式 (1-26) 消去空间变量并写成常微分方程形式为:

$$\frac{d^2 I}{dt^2} - \frac{c^2}{k} \frac{dI}{dt} = \frac{c^2}{p} \left[1 - \frac{1}{b} \right] I \quad (1-27)$$

式 (1-27) 的特征方程为 $r^2 - \frac{c^2}{k} r - \frac{c^2}{p} \left[1 - \frac{1}{b} \right] = 0$, 其解为二实根 $r_{1,2} = \frac{\frac{c^2}{k} \pm \sqrt{\frac{c^4}{k^2} + 4 \frac{c^2}{p} \left[1 - \frac{1}{b} \right]}}{2}$, 于是式 (1-27) 的通解为:

$$I = C_1 \exp \left[\left[\frac{\frac{c^2}{k} + \sqrt{\frac{c^4}{k^2} + 4 \frac{c^2}{p} \left[1 - \frac{1}{b} \right]}}{2} \right] t \right] + C_2 \exp \left[\left[\frac{\frac{c^2}{k} - \sqrt{\frac{c^4}{k^2} + 4 \frac{c^2}{p} \left[1 - \frac{1}{b} \right]}}{2} \right] t \right] \quad (1-28)$$

其中, C_1 和 C_2 为两积分常数。

式 (1-28) 表明: 当前一指数项占主导地位时, I 随时间呈指数上升, 此即信息增长的普赖斯曲线的理论机制; 而当后一指数项占主导地位且 $\sqrt{\frac{c^4}{k^2} + 4 \frac{c^2}{p} \left[1 - \frac{1}{b} \right]} > \frac{c^2}{k}$ 时, I 随时间呈指数下降, 此即信息老化的负指数曲线的理论机制 (当然, 信息老化主要是质量问题而非数量问题, 这里体现的只是动力机制), 一般情况下, I 随时间变化的规律由式 (1-27) 中两指数项共同描述, 这也就是为什么普赖斯曲线和负指数曲线都不能全面表述信息随时间变化规律的根本原因。

当只考虑信息的空间分布而设 $\frac{\partial I}{\partial t} = 0$, $\frac{\partial^2 I}{\partial t^2} = 0$ 时, 用式 (1-25) $+$ 式

(1-26)得:

$$\Delta^2 I = -\frac{c^2 b + k}{pb(c^2 + k)} I \quad (1-29)$$

这是类似定态 Schrodinger 方程的二阶偏微分方程, 可以根据实际情况和约束条件用分离变量法或积分变换法等方法求解, 其解一般具有较复杂的形式, 布拉德福定律、洛特卡定律、齐夫定律等文献“空间”分布经验规律的共同内在机理就包括在其中, 这也就是 Sichel 提出的 GIGP 模型 (Sichel, 1985) 等复杂形式的文献计量分布可能具有的解析表达。

对情报学经验规律的统一模型研究和机理研究目前仍在动态进行中, 这是一个值得继续深入探索的理论课题。

习题和思考题

1. 按照布拉德福定律, 设布拉德福常数为 5, 当核心期刊数量为 20 种时, 外围一区 and 外围二区期刊数量各是多少?
2. 按照洛特卡定律, 当某领域中写 6 篇文献的作者数量为 1000 人时, 该领域中写 1 篇文献的作者数是多少? 估计该领域中作者总量约是多少?
3. 如果普赖斯定律 $Q(t) = a \exp(bt)$ 成立, 一个初始文献量 $Q(0) = 100$ 的学术领域, 当常数 $b=1$ 时, 估计 10 年后文献量 $Q(10)$ 将达到多少?

主要参考文献

- 邱均平. 1988. 文献计量学. 北京: 科学技术文献出版社
- 王崇德. 1994. 情报学引论. 天津: 天津大学出版社
- 严怡民. 1994. 情报学概论. 修订版. 武汉: 武汉大学出版社
- 叶鹰. 1985. 文献增长律与老化律研究. 情报科学, 6 (4): 30-34
- 叶鹰. 1987. 论文献规律. 情报业务研究, 4 (4): 289-295
- 叶鹰. 1999. 信息科技基础理论的分析建构. 情报学报, 18 (2): 160-166
- Egghe L, Rousseau R. 1990. Introduction to Informetrics. Elsevier Science Publisher
- Garfield E. 1980. Bradford's Law and Related Statistical Patterns. Current Contents, (19): 5-12
- Sichel H S. 1985. A Bibliometric System Which Really Works. Journal of the American Society for Information Science, 36 (5): 314-321
- Vickery B C, Vickery A. 1987. Information Science in Theory and Practice. Butterworth & Co. Ltd.



第2章

情报学基础理论

情报学基础理论要解决的主要问题是情报学的概念系统、理论原则、体系结构等学科架构问题。对于基本概念，情报学基础理论首先需要澄清情报概念和信息概念及其关系。

物质、能量和信息并称为当今社会的三大支柱，因而信息概念应该作为一个基本概念。可是，信息究竟是什么？却迄今为止还没有一个公认的定义。虽然中国古诗中即有“梦断美人沉信息，目穿长路倚楼台”（南唐李中《暮春怀故人》）、“不乞随珠与和璧，只乞乡关新信息”（南宋李清照《上枢密韩肖胄诗》）等佳句，但这只是文学性描述而非科学性论断。按照申农（C. E. Shannon）在其经典论著《通讯的数学理论》中的原始定义：“信息是用来消除不确定性的东西”；而控制论创始人维纳（N. Wiener）在其名著《人有人的用处》中也给出一个经典定义^①：“信息是人们适应外部世界并且使这种适应反作用于外部世界的过程中，同外部世界进行互相交换的内容和名称”。现在，不仅信息科技各分支可以对信息取不同定义，从科学技术、社会科学等角度也可以各有其说，在《中国大百科全书》中就可见一斑：《自动控制与系统工程》卷中定义信息为“符号、信号或消息所包含的内容，用来消除对客观事物认识的不确定性”，而《新闻出版》卷中则定义信息为“事物运动状态的陈述，物与物、物与人、人与人之间的传输特征”，最后还是《哲学》卷归纳得较为完善：“关于信息概念，在科学界尚无统一

^① 维纳．人有人的用处．北京：商务印书馆，1978；转引自傅祖芸．信息论——基础理论与应用．北京：电子工业出版社，2001，3

的、精确的定义。在哲学界……有4种观点：（1）信息是精神实体的特性；（2）信息是物质的普遍属性；（3）信息是与物质和精神并列的第三种存在；（4）信息是物质载体与意识成分的特殊结合。”因此，情报学基础理论研究自然应从信息概念开始。

2.1 情报学概念系统

既然物质、能量、信息是当今世界的三个基本物理量，物理理论中就应该有关于信息的理论，可是信息论只是一种通讯的数学理论，因而建立信息的物理理论非常必要和重要。

由于物质质量 M 可以表示成

$$M = \int \rho(s) ds \quad (2-1)$$

其中， $\rho(s)$ 是物质密度函数（国际单位制单位： kg/m^3 ）， s 是空间坐标。

能量 E 可以表示成

$$E = \int \epsilon(t) dt \quad (2-2)$$

其中， $\epsilon(t)$ 是能流密度函数（国际单位制单位： J/s ）， t 是时间坐标。

信息 I 类比应该可以表示成

$$I = \int \eta(r) dr \quad (2-3)$$

其中， $\eta(r)$ 是信息密度函数（建议单位：进制位/标度位，例如： bit/tet ）， r 就是新“发现”的物理学基础坐标，可称为序间坐标或量间坐标（简称序间或量标），它与时（ t ）、空（ s ）一道构成物理学的基本架标系统。

这样，在物理层次上，信息与物质和能量密不可分，其物理基础构成一个序间-空间-时间架构并蕴含对时空观的拓展。只有具备序间或量标，时间和空间才能序化或量化，时间-空间-序间不可分离，而数则成为它们之间的纽带。

信息在物理本质上是物质和能量系统有序化程度的量值，这就是在基础科学层次上的信息定义，该定义与申农信息（Shannon, 1949）^①、费歇尔信息（Frieden, 1990）^② 等不同。在一个系统内，信息与物质和能量紧密相关，物质和能量的变化将导致信息变化，信息变化也将导致物质和能量变化。

相对于“物质是标志客观实在的哲学范畴”、“能量是标志客观运动的哲学范畴”，也可以用“信息是标志客观知识的哲学范畴”作为信息的哲学定义（叶鹰，

① 申农信息熵 $H(P) = -\sum P \log P$ ， P 为信源概率

② 费歇尔信息 $I(p) = \int [p'(x)]^2 / p(x) dx$ ， $p(x)$ 为概率密度函数

1999)。

引入序间后的物理基础关系如下：

坐标	对称性	守恒量	特征常数	坐标积分
序间 (r)	对称性破缺	无守恒量	k	$I = \int \eta dr$
空间 (s)	空间平移和旋转	动量守恒	c	$m = \int \rho ds$
时间 (t)	时间平移和旋转	能量守恒	h	$E = \int \epsilon dt$

其中, c 为光速, h 为 Planck 常数, k 为 Boltzmann 常数。从信息等于负熵的角度考虑, 根据热力学第二定律:

$$dS \geq \frac{dQ}{T} \quad (2-4)$$

其中, S 为熵, Q 为热量, T 为温度。因 Q 、 S 都是外观物理量, 从系统内部看, 由于能量 $E = -Q$, 而 $I = -S$, 故有

$$dI \leq \frac{dE}{T} \quad \text{或} \quad E \geq \int T dI \quad (2-5)$$

类比

$$E = mc^2 \quad (2-6)$$

应有

$$E = ITh^2 \quad (2-7)$$

设 v 为波频, T 为温度, 从能量公式 $E = mc^2 = hv = kT$ 考虑, 符合序间—空间—时间物理逻辑框架的基本物理常数只能选择 c , h , k , 其含义是:

坐标	物理常数	坐标微分	坐标积分	坐标作用
序间 (r)	$u_{\min} = k$	$u = \frac{dr}{ds}$	$I = \int \eta(r) dr$	$D = \frac{d}{dt} (Iu)$
空间 (s)	$v_{\max} = c$	$v = \frac{ds}{dt}$	$m = \int \rho(s) ds$	$F = \frac{d}{ds} (mv)$
时间 (t)	$w_{\min} = h$	$w = \frac{dt}{dr}$	$E = \int \epsilon(t) dt$	$G = \frac{d}{dt} (Ew)$

取 $c = 2.9979 \times 10^8 \text{ m} \cdot \text{s}^{-1}$, $h = 6.62608 \times 10^{-34} \text{ J} \cdot \text{s}$, $k = 1.37951 \times 10^{-23} \text{ J} \cdot \text{K}^{-1}$, 有

$$e = \frac{h}{kc} = 1.60218 \times 10^{-19} \text{ C} \quad (2-8)$$

这就自然导出了基本电荷。

这样，信息—能量—质量的物理基础构成一个序间—空间—时间结构并蕴含对时空观的拓展：物质的坐标基础是空间、能量的坐标基础是时间，信息的坐标基础是序间（或量标），这不仅建立了信息的物理定义，而且明确了统一宏观与微观、统一相对论和量子论、统一引力论和热力学的基础。

2.1.1 概念逻辑

情报学的概念系统可以从信息这一基本概念出发，逻辑地扩展成包括信息、情报（资讯）、知识等在内的概念体系。其核心构造逻辑如下：

由式（2-3）定义的信息是物理信息或客观信息。真正有用或有效的信息 J 是信息 I 的价值修正，即有价值信息：

$$J = I^V \quad (2-9)$$

其中，信息的价值系数 $V \in [0, 1]$ ，当 $V=1$ 时， $J=I$ ，所有信息都是有价值的；而当 $V=0$ 时， $J=1$ ，只是表示有信息存有这一消息。 V 越大，表明有效信息越多。存在着的信息是 I ，而真正有价值或有用的只是 J ， J 是有效信息、社会信息或主观信息。

英文 information 的主流中文翻译是“信息”，而日文译成“情报”，中国港台地区又译为“资讯”，于是造成“信息”、“情报”、“资讯”混用的局面。如果不用“情报”或“资讯”概念，则“信息”概念可分解为客观信息和主观信息而自足；如果需要区别使用“信息”、“情报”和“资讯”，则“信息”是指 I ，而“情报=资讯”就是 J 。

知识 K 则可定义为有效信息 J 即情报（资讯）的积分：

$$K = \int \delta J dI = \int \delta I^V dI \quad (2-10)$$

其中， δ 是信息的知识转化系数。此时 K 的计量单位与 I 的计量单位一致。由式（2-10）决定的知识就是客观知识。

式（2-10）的微分形式是：

$$\frac{dK}{dI} = \delta J = \delta I^V \quad (2-11)$$

式（2-11）表明单位信息增量产生的单位知识增量与有价值信息量（有效信息量）成正比，这与式（2-10）所表明的知识是有效信息的积分兼容。式（2-10）和式（2-11）同时也表明：当对信息施行调控时必将影响相应的知识，所以可以通过控制信息来控制知识。

由于引入了价值系数，勒歇尔（Rescher）模型被自然统一进来：