

论图书情报学中的相似原理

图书情报学中有一条维系理论与经验、规则与操作的纽带,这就是相似原理。本文希冀论述相似原理,并且简略地陈述该原理的具体展开与应用。这对学科的深化与现代化建设是有补益的。

一、相似与相似本质

自然界与人类社会存在并滋生着大量的相似现象。广袤无际的星系,无奇不有的人类社会,常规无法观察的原子结构,都充满着相似:表面与本质的相似,静态与动态的相似,行动与思想的相似。相似也是人类开展情报活动,建立情报作业手段,发展情报事业,开发图书情报学的思维基础、理论原则和操作依据。

(一) 相似现象。情报行为与活动中的相似现象比比皆是。例如,文献间的情报转换与传递,可以示意如下:

信息源 → 来源文献 → 转换规则 → 产品文献 → 接受者
(S) (M(S)) (M(R)) (U)

文献间的转换与传递过程:在众多信息源(S)中,仅把对口径或具有针对性的文献输出(或挑选)为来源文献〔M(S)〕。依据通行的标准转换规则——例如叙词标引规则,把〔M(S)〕先行主题分析,继之进行概念标引,再施符号标引——给出一个或几个标识(即叙词),以致成为情报产品〔M(R)〕。重组的情报产品——条目,已有别于原来的〔M(S)〕,它是〔M(S)〕

的替代品。在标引过程中也可能出于主客观原因的失误,使文献在变换与传递过程中失真欠准。最后均以传递至情报使用者而告终。

人们在人工处理文献情报的启迪下,根据相似原理设计并运行了机器信息变换与传递的过程。示意图如下:

信息源 → 编码器 → 信道 → 译码器 → 接收器
(S) (M(S)) (M(R)) (R)

这里,信息源(S)的信息被编码成为〔M(S)〕,然后作为物理信号通过某种载体(信道)加以传输。因译码而接收的信息〔M(R)〕可能不同于〔M(S)〕,所以信号接收可能受到某种干扰而畸变,最后也必定送达到接受者处。

显然,文献情报转换与传递系由手工操作的;机器间的信息变换与传递则是由机器执行的。两者有其相似之处,后者恰恰是模仿前者而出现的,相似原理启发人们实现了由手工操作到机器加工的跃进。

相似现象十分普遍。它们有时异常明显,有时却又特别隐蔽。为了研究的方便,人们把相似现象概分为纵向相似与横向相似两大类。客观事物在一定理论基础范围内,出自本质联系中的相似关系而演绎成系统这就是纵向相似。根据纵向相似就可以发展成一门学科的知识体系。横向相似是在那些跨越学科或领域,超时间、超观念相互作用中的相似现象。横向相似促进知识的相互渗透与转移,便出现了许多综合性(与原型的相

似集合)的新兴学科。图书情报学便是其一。

(二)相似的本质。既然相似现象十分普遍,那么“相似”应如何界定呢?由于相似无处不有,随处可见,以致“相似”的定义大有“随乡入俗”之势——在所在范围内就事论事,缺乏普遍的品格。

在自然科学中,“相似”被视为各种运动物理量相等或成比例的状态。这一概念从初等几何学中移植过来,如三角形相似。

在社会科学中,相似事件“物理量”——变量是不尽规范的。这是因为社会变量受人为性和不可重复性的制约。目前,整个社会科学的定量水平还未达到自然科学那么稳定,当然人类的社会运动较之自然运动要更为高级和复杂。社会的“相似”常常被解释为客观事物的时间、地点、人物、情节、前因、后果等一致性。自然,上述这6个要素要是完全一致那就成了一模一样的同一事物了。这里所讨论的相似是指这6个要素贡献于“相似”的份额不一,成分不等,比重不同。所以,社会及包括图书情报学在内的社会科学中的相似,乃是一个连续多值的判断对象,相似有大有小,有强有弱。相似的要素越多,相似程度越高,这种相似性就越强烈。当然,社会相似现象与相似程度的判断与个人的主观因素关系甚大,同样两个现象或事物的相似性常常会因判断者见仁见智。

那么是否能给“相似”以一个更为抽象的普遍定义呢?我们必须试行上升到哲学高度来认识相似的本质。

在客观事物发展过程中,同时存在着“相同”与“异化”两种倾向。维护并坚持相同,才会有继续与继承,才能体现传统。反之,突出异化事物,才会有所发展,才有突破与创新。相似当然不是相同,况且自然界与人类社会绝然相同的事物是根本没有的。全世界几十亿人口中没有完全相同的指纹——所以指纹成了个人的绝对特征标志。自

然界成千上万的树叶也绝然没有两片是相同的。甚至就是在同一个氢原子内,核外电子运动的轨道也是不尽相同的。因此,可以这样理喻:100%的相似就是相同,这是根本不存在的理想境界;低于100%的相同便是程度不等的相似。何以相同一定要低于100%呢?原因就是其间不可避免地要渗入异化,异化是事物发展过程中背离相同的趋势。这样一来,相似正是客观事物存在的相同与差异的对立统一,是相同与异化达到某种平衡或妥协的状态。由于相同与异化在客观事物中成分不等,所以它们的相似程度也就会有所区别。

相同与异化在事物的发展中所显现出的作用也是不一样的。强调与突出了哪一种趋势,就会造成程度不等的相似出现,以致相似有强弱之分。这就是人为制造相似的基础,也是自然科学,工程技术,思维活动,社会实践以及情报行为,都可以依需要来建立相似观点的出发点。

二、相似规律

这里讨论的相似规律是从定性角度来展开的,意在说明相似的产生、发展与功效。

(一)相似产生的基本依据。这个基本依据是说,只有相似单元、相似层次的构造才能产生相似。单元是指组成事物内部结构中最基本的单位,科技文献的单位就是自然语言的词汇,词汇是形成文献的最基本的语言符号。由单元进而组成层次。层次是指事物内部的相互作用,相互联系与相互制约的最为紧密的相对独立部分。它在微观上是结构、运动和时空的统一表现形式。一般来说,事物越复杂,它的层次也就越多。

情报操作中这一规律的体现是不胜枚举的。我们常说的检索工具主要有两种功能,一是存储情报,二是检索情报。粗略地说,情报检索主要是文献检索。检索工具主要包

括揭示文献信息和标识两部分。标识应按检索目的有规律地排列起来。目前,检索工具可以分为卡片式、书本式和磁介质式。这3种不同的情报检索工具之所以具有按用户的情报提问输出文献的相似检索功能,是因为三者的文献信息揭示部分和标识都是对应相似的,即由相似单元和相似层次构成的,以致一个文献部门的普通卡片盒与电子计算机一样,均能从大量文献集合中,查找出所需要的文献。

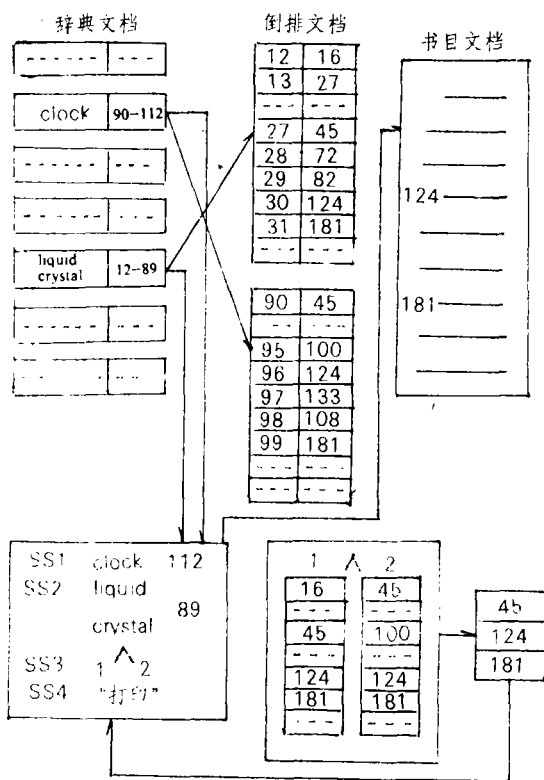
(二) 相似产生的主要条件。上述“基本依据”只回答了相似产生的可能,但不等于这样的相似必然会出现。只有具备相似过程和相似环境,相似才会出现。两个客观事物有了相似的“基因”后,还需要有相似过程和相似环境来配合。否则,纵有相似“基因”,它们的过程与环境也会大相径庭,相似也不会出现。在图书情报学中,虽有相似的“基本依据”,但因无产生相似的“主要条件”而导致相似程度淡泊,是不乏其例的。同是用主题法标引与检索文献,但是手工与机检过程却相差甚远,虽然它们有其相似“基因”,但过程与环境不尽相似,以致会令人有面目全非之感。

以查找液晶钟表的专利文献为例。若用单元词或叙词来组配检索,虽然检索方式的原理未变,但因过程与环境大变,表现出的相似性几乎消失殆尽。下表是模拟美国《化学专利单元词索引》的样例。

单元词	1	2	3	4	5	6	7	8	9	0
Liquid	181	72		24	45	16	27		39	
Crystal	201	80				176			109	
Clock	181		133	124	45			78		100

手工检索只能在该《索引》不同的页码(单元词按字顺排序)上查找到“Liquid Crystal”和“Clock”两个款目词,按个位数

核对文献号,结果检索出 181、124 和 45,共 3 篇文献。查找方式完全是眼看手翻,再复杂一些的情报提问就更难进行了。可见检索词组配检索在手工作业下是很难迅速与准确进行的。可是运用电子计算机检索就能发挥优势了。下面我们按兰开斯特的典型范例来模拟上述的文献计算机检索(见下图)。



电子计算机检索示意图

在机检的情况下,索引档是由辞典档和倒排档构成的。辞典档是标识按字顺排列而成的表,该档收录两种情报:一是标识所代表的方向编号——在文献里的序位(在图中,表明“Clock”存贮在倒排档 90 号以后);二是情报的数量,有“Clock”属性的文献约 112 篇。同理,“Liquid Crystal”存贮在倒排档 12 号之后,共有文献 89 份。倒排档以标识为基础,收录以该标识所标引的文献号。例如,“Clock”的文献编号为 16、27、……45、72、82、124、181、……。检

索过程是在标识之间进行逻辑运算时参照倒排档。针对 Clock、Liquid Crystal 在倒排档中互相吻合一致, 因而检出文献编号为 45、124、181, 共 3 篇。单元书目文档是按文献编号顺序排列的文献档。该档文献是根据“打印”指令而输出检索结果的。

手工印刷型检索工具的检索, 机检的文献检索, 有其相似的“基因”, 可是检索过程与检索环境都已改变, 因而检索的方式、策略、效率和输出形式都大不相同。

(三) 相似属性、特征与功能。一般而言, 如果客观事物中相似属性和相似特征越多、越强烈, 那么这种相似的功能也就越多、越大。这是因为相似特征和属性的富集实际上是一种选择性的综合, 它保留并相互辉映了相似原型的特长或优势。由于相似原型得到多向反映, 所以相似功能就越丰富、越发达。

这种例子在生活中是相当多的, 在情报产品中大量存在。例如, 各种手册就是典型的相似属性、特性繁多的情报产品, 它属于工具参考书之列。手册一般都是汇编、聚集并提供大量的从不同情报源搜集并以可用数量表示的原始数据, 这其实就是一种广泛的相似。手册对于回答那些要求特定情报或数据的“即时咨询”的提问来说, 也是非常有用的。有人统计过, 在科技图书馆中提出的 50~80% 的咨询, 均可以在手册中找出答案。由于手册中包括了大量的不同数据, 人们经常将这些数据浓缩成单行本, 所以手册常被喻为“单行本、图书馆”。许多手册都精心编制索引, 以进一步保证迅速查出所需的数据单元。其实, 手册是相似聚集有价值数据的结果, 它与其来源甚为相似。一个手册相似属性多, 相似特征明显, 相似面广泛, 那么这个手册就几乎无所不包, 什么情报都几乎可一查即得, 这乃是广泛相似的结果。

此外, 自然科学中尚有相似三定理。相似第 1 定理称: 对于相似现象, 其相似指标

等于 1。相似第 2 定理称: 设一物理系统有 n 个物理量, 若其中有 K 个物理量的量纲是相互独立的, 那么这 n 个物理量可表示成相似准则 $\pi_1, \pi_2, \dots, \pi_{n-k}$ 之间的函数关系。相似第 3 定理称: 对于同一类物理现象, 如果单值量相似, 而且由单值量所组成的相似准则在数值上相等, 则这类现象相似。上述相似三定理在自然科学中运用得炉火纯青, 但包括图书情报学在内的社会科学对它还比较陌生。

三、相似原理的展开与应用

人类情报行为与过程贯穿着相似规律与定理的体现与应用, 相当多的情报单元操作或综合作业就是依据相似原则来展开的, 而且运用得很成功。

(一) 文献的耦合聚类。1961 年, 美国情报科学家 M·M·开斯勒在麻省理工学院做了一场报告, 首次提出“文献耦合”(Bibliographic Coupling) 的概念。它可使人们从表面相似深入到内在的主题相似, 从而形成文献聚类。引文耦合的相关强度是可以计算的, 文献〔1〕提供了较为先进的计算方法。若设 d_i 与 d_j 为著文, 则:

$R^i = \{r_1^i, r_2^i, \dots, r_{L_i}^i\}$ 为著文 d_i 的引文集合;

$R^j = \{r_1^j, r_2^j, \dots, r_{L_j}^j\}$ 为著文 d_j 的引文集合。

其中, L_i 为 d_i 的引文数量; L_j 为 d_j 的引文的数量。

如果 $r^i \cap R^j = \{r_1^j, r_2^j, \dots, r_{L_y}^j\}$ 为 d_i 与 d_j 中共同拥有引文的集合, L_y 为引文偶的数量。又令 $CS_y^K = L_y = |R^i \cap R^j|$; r_s 为所收集的数据中受引的任一引文 j_{rs}^D 为 r_s 在所收集的数据中被引用的绝对频次。

则 CS_v^{lf} 为既考虑了受引文献的数量、又考虑了受引文献频次的耦合强度:

$$CS_v^{lf} = \frac{\sum_{n=1}^{L_i} 1 / f_{rky}^D}{L_i + L_j} \cdots \cdots (3-1)$$

其中, L_i 与 L_j 为受引论文的数量。

这些都是基于相似原理而演绎出来的。

(二) 术语相关分析。这种相关分析是基于文献中重要术语共同出现频次多寡来确定两篇文献的相关程度的。术语相关一般通过文献的内容分析来获得。首先, 把研究文献中频繁出现的名词术语作为重要的名词术语; 然后, 测定这些重要术语在各篇文献中出现的频次; 接着计算相似度, 以达到相关分析的目的。相似度是这样定义的:

设 $W = \{W_1, W_2, \cdots, W_n\}$ 为重要术语的集合, $T = \{t_1, t_2, \cdots, t_m\}$ 为被研究的文献集合。术语 W_k 出现在 t_i 中的频次为 X_{ki} , 那么文献 t_i 与 t_j 的相似度为:

$$S(t_i, t_j) = \frac{\sum_k X_{ki} \cdot X_{kj}}{\sqrt{\left(\sum_k X_{ki}^2\right) \left(\sum_k X_{kj}^2\right)}} \cdots \cdots (3-2)$$

用相似度 S 可以表征两篇文献的相似程度。规定了 S 值后, 就可以进一步实行相关分析了。

我国道教的经典——《黄帝阴符经》的注释多达 18 种, 如何确定注释之间的相关程度呢? 日本松本浩一等就是利用相似度来进行相关分析而加以判断的^[2]。首先统计 18 种注释中的 91 个 (基本上是一个术语一个汉字) 重要术语, 接着测量 91 个术语在 18 个注释中的出现频次, 然后, 两两计算相似度, 便可得知注释之间的相关性, 亦即相似程度。

(三) 文献经验规律的相似性。文献计量学中的经验规律可以分为两大类, 其中一部分是与时间无关的“静态”规律, 它们是布

拉德福定律、洛特卡定律和齐普夫定律^[3]。这三个定律的共性是异常明显而又强烈的: 其分均呈反“J”型, 高度倾斜, 可一般地表示为通式 $f(x) = c(\alpha + x)^k$, c 、 α 、 k 为常数^[4]。简言之, 它们都强烈地体现着不均匀的“成功产生成功”。之所以出现如此之多的共性, 均因它们内在的相似性强烈, 也可说是相似原则使然。

1、布拉德福定律。设有 M 个来源 A , 产生项目 Y 为 N 个。依据递降的原则将项目 Y 排序为: $x_1, x_2, \cdots, x_m$, 而项目则呈 $y_1 > y_2 > \cdots y_M$, 来源为 $A_1, A_2, \cdots, A_N$, 则:

$$\sum_{i=1}^M y_i = N$$

依布拉德福定律 (区域法), 可以得出以下结果:

$$\sum_{i=1}^{M_1} y_i = \sum_{i=M_1+1}^{M_2} y_i = \sum_{i=M_2+1}^{M_3} y_i = \frac{N}{3}$$

当然 $M_1 + M_2 + M_3 = M$, 若把来源 M 视为期刊, 项目可视为这些期刊相应的载文量, 可有:

$$M_1 : M_2 : M_3 = 1 : \alpha : \alpha^2 \cdots \cdots (3-3)$$

$\alpha > 1$, 俗称为布拉德福系数, M_1, M_2 和 M_3 分别为核心区、相关区和边沿区的期刊数。

2、洛特卡定律。若把每个来源看成是科学论文的作者, 相应的项目为期刊论文。如果撰写 n 篇论文作者的份额为 $f(n)$, 则:

$$f(n) = \frac{C}{n^2} \quad (C = \frac{b}{\pi^2} = 0.6079)$$

那么写 n 篇论文的人数:

$$b(n) = Mf(n) = \frac{MC}{n^2}$$

同理, 写 n 篇论文的作者共撰写的论文数量为:

$$a(n) = nb(n) = \frac{MC}{n}$$

我们假设 $S(t, t')$ 为发表 t 篇到 t' 篇论文的作者所撰写的论文数, 那么便有:

$$S(t, t') = \sum_{n=t}^{t'} \frac{MC}{n} = \int_t^{t'} \frac{dn}{n} MC$$

$$= MC[\ln t' - \ln t] \cdots \cdots (3-4)$$

如果 T 为这个作者群中发文最多作者的发文数, 那么式 (3-4) 便有:

$$S(1, T) = MCLnT = N$$

$$T = e^{\frac{N}{MC}} \cdots \cdots (3-5)$$

如果我们对洛特卡定律也仿布拉德福定律分为三区, 并规定 a_1 为第一区发表论文的最高数, a_2 为第二区发表论文的最高数, a_3 为第三区发表论文的最高数。同时, 令 $T = a_1$; $a_1 > a_2 > a_3$, 则满足 $S(1, a_3) = S(a_3, a_2) = S(a_2, a_1) = \frac{N}{3}$; $MCLna_3 = MC(Lna_2 - Lna_3) = MC(Lna_1 - Lna_2)$ 。

$$\text{因 } a_3 = \frac{a_2}{a_3} = \frac{a_1}{a_2}, \text{ 所以 } a_2 = a_1^{\frac{2}{3}}$$

$$= T^{\frac{2}{3}}, a_3 = a_1^{\frac{1}{3}} = T^{\frac{1}{3}}。 \text{ 同时, 便}$$

$$\text{有: } a_1 : a_2 : a_3 = T : T^{\frac{2}{3}} : T^{\frac{1}{3}}。$$

假设 $R(t, t')$ 为发表 t 篇到 t' 篇论文的作者数, 则:

$$R(t, t') = \sum_{n=t}^{t'} b(n) = \sum_{n=t}^{t'} \frac{MC}{n^2} = \int_t^{t'} \frac{MC}{n^2} dn$$

$$= MC \left(\frac{1}{t} - \frac{1}{t'} \right)$$

这样便有:

$$R\left(T^{\frac{2}{3}}, T\right) = MC\left(T^{-\frac{2}{3}} - T^{-1}\right)$$

$$R\left(T^{\frac{1}{3}}, T^{\frac{2}{3}}\right) = MC\left(T^{-\frac{1}{3}} - T^{-\frac{2}{3}}\right)$$

$$R\left(1, T^{\frac{1}{3}}\right) = MC\left(1 - T^{-\frac{1}{3}}\right)$$

$$R\left(T^{\frac{2}{3}}, T\right) : R\left(T^{\frac{1}{3}}, T^{\frac{2}{3}}\right) : R\left(1, T^{\frac{1}{3}}\right)$$

$$= 1 : T^{\frac{1}{3}} : T^{\frac{2}{3}} \cdots \cdots (3-6)$$

如果令 $T^{\frac{1}{3}} = \alpha$, 则 $T^{\frac{2}{3}} = \alpha^2$ 。那么, 洛特卡定律的布拉德福式处理也有类似的 $1 : \alpha : \alpha^2$ 。

3、齐普夫定律。假如把每个来源看成一个词汇, 相应的项目看成该词汇在某文集中的出现次数。根据齐普夫定律可知: 该词汇出现的频次为 $f(r)$, r 为依频次排列的顺序, 那么便有:

$$f(r) = \frac{C}{r} \cdots \cdots (3-7)$$

如果假定一个词只对应一个顺序, 也就是不能有并存相同频次的词, 那么最高频次, 亦即最低序号的词为 M 。若令 $Q(t, t')$ 为顺序 t 到 t' 的词汇出现的总频次:

$$Q(t, t') = \sum_{r=t}^{t'} \frac{C}{r} = C(\ln t' - \ln t) \cdots \cdots (3-8)$$

同时, 将所有频次也等分为三个区, 亦即按布拉德福式处理, 则第 1 区为高频—低序区, 其序号为 $1 - b_1$ 。同理, 第 2 区为 $b_1 - b_2$, 第 3 区为 $b_2 - b_3$ 。令 $b_3 = H$, 则:

$$Q(1, b_1) : Q(b_1, b_2) : Q(b_2, b_3)$$

$$= C \ln b_1 : C(\ln b_2 - \ln b_1) : C(\ln b_3 - \ln b_2)$$

$$= b_1 : \frac{b_2}{b_1} : \frac{H}{b_2} = 1 : 1 : 1$$

$$\text{解: } b_1 = H^{\frac{1}{3}}, b_2 = H^{\frac{2}{3}}, b_3 = H。 \text{ 亦}$$

$$\text{即: } b_1 : b_2 : b_3 = 1 : H^{\frac{1}{3}} : H^{\frac{2}{3}}。$$

设 $H^{\frac{1}{3}}$ 为 α , $H^{\frac{2}{3}}$ 为 α^2 , 那么齐普夫定律也与布拉德福定律类似。可见这三个定律若将其项目三等分的话, 其来源均呈等比级

数。这便是相似原理在它们中所起的作用——真可谓“原理是一个原理”。

四、研究与利用相似的意义

既然相似原理支配与控制着众多的情报活动,而且许多图书馆与情报中心的操作又都据此而建立起来,那么充分研究相似现象,挖掘相似的深层涵义,合理地利用相似,对开发图书情报资源是大有好处的。它能启发思维;活化智力。具体来说,有如下的积极意义⁽⁵⁾:

(一) 有助于强化认识,深化认识。面对不胜其繁的客观事物,人们要想一蹴而就地完成认识,是根本不可能的。认识是一个渐进而又永无止境的过程。简化认识、催化认识、优化认识是人类永恒的任务。研究相似现象,利用相似规律,有助于我们在认识客观事物的征途上,加快节奏,加大步伐。我们业已知道,相似是“同”与“异”的并存与共现,研究事物的异中之同,能使貌似无缘的事物依照它们的相似特征加以重组与归类,使无序的研究对象有序化。反之,研究事物之间的同中之异,又能根据它们的不相似属性,查明研究对象的多样性,灵活性,使思维更为活跃,并有助于我们认识现代科学的整体化趋势。重视客观事物的相互联系和兼容渗透,对图书情报学学科理论建设也是非常必要的。

例如,体系分类法是一种直接体现知识等级概念的标识系统,它是情报交流载体——文献有序化的重要依据和工具。这种有序化不仅适用于文献所含的信息方面,而且也适用于文献的物理载体。体系分类法的知识基础是概念逻辑和知识分类。人类知识分类的历史是不断地用相似原理去进行科学概念划分与概括的发展史。19世纪,自然科学呈现出一种大跃进的态势。迈尔-赫尔姆霍茨-焦耳的能量守恒与转换定律,施旺与

施莱登的细胞理论,达尔文的生物进化论,门捷列夫的元素周期律等都是这个时期重大发现。此外,数学、物理、化学、天文学、地质学、医学也都取得了重大成果。如何克服科学在蓬勃发展过程中出现的无序与紊乱,对科学本身进行分类一直被认为是有效的措施,也是异常引人瞩目的重要课题。费·培根、圣西门、康德、黑格尔、孔德、安培等人都在进行过科学分类的研究,把五花八门的科学现象按照一定原则,运用相似的思维方式,把科学现象进行区分与归类。

早在1858年科学的分类就引起了恩格斯的注意,1873年他拟定了科学分类法,把科学分类的水平推向一个崭新的阶段。他把研究事物性质,揭示其相互连结又相互排斥的本质矛盾作为分类的基础,并以这种科学的唯物主义原则代替了只按事物外部的、附加的,甚至常常是凭空臆造的特征去划分科学的、形而上学的理论。恩格斯拟定的科学分类基础是:物质运动不同形式的发展与转化;自然界历史发展的整体性;人类对自然界认识的顺序发展。为此,他列出了以下的知识分类表:

天文学	力学	数学	物理学
化学	地质学	古生物学	矿物学
植物生理学	动物生理学		
解剖学	治疗学	诊断学	

这种排列基本上再现了物质运动不同形式连续出现的顺序,因此也就再现了自然发展的不同阶段。恩格斯根据众多科学现象、过程、原理的相似性,进行相似概括,区分不同的学科。无疑这对科学的认识是有利的,它使庞杂而又纷乱的科学变成简明清晰的体系。在这里,恩格斯突出了科学的相同,回避了相异,对整个科学进行了归类,并且演绎出体系。

(二) 有助于优化思维,推动创新。研究相似有利于开展积极的思维,可以举一反三,触类旁通地有所发现,有所创新。善于

运用比较、分类、类比等抽象思维形式,也易于产生形象思维的联想。在情报科学范围内,许多观点的建立,新技术的开发,都可以说受益于相似的研究与运用。例如,1926年美国统计学家 A·J·洛特卡发表了被后人誉为“洛特卡定律”的论文——《科学生产率的频次分布》。据洛特卡自称,他本人深受同代生物学家 C·J·维利斯的影响,发现科学论文作者分布与植物的“属种”关系极为相似。以菊科植物为例:

- 有 446 个属含有 1 个种;
- 有 140 个属含有 2 个种;
- 有 97 个属含有 3 个种;
- 有 43 个属含有 4 个种;
- 有 55 个属含有 5 个种。

可见含有 1 个“种”的“属”,一般数量都很大,大体上占据整个群系 1/3;含有 2 个“种”的“属”的频次仅居其后;含有“种”数较多的“属”依次变少……。这种相似启发了洛特卡,使他在科学论著作者分布中发现了同样的情形:

- 有 3991 个人写 1 篇论文;
- 有 1059 个人写 2 篇论文;
- 有 493 个人写 3 篇论文;
- 有 287 个人写 4 篇论文;
- 有 184 个人写 5 篇论文。

洛特卡正是接受了维利斯的学术观点,从研究相似出发,发现和创立了至今仍在文献计量学中深有影响的一条定律。

(三) 有助于横向借鉴,多向移植。图书情报学是发展中的新兴学科,人类情报活动虽然有悠久的传统,但其历史却很短暂。目前这个学科的学术地位倒是确定了,但理论体系却没有完整建立起来。除了加强其内部科学逻辑结构外,尚应从相关学科去寻找借鉴,从中移植先进的理论与成果,以滋补这一学科的发展。

图书情报学的相关学科很多,它们之间具有极大的相似性。现代汉语的词频统计是一件非常重要而又有意义的工作。我们可以利用它来编纂和管理汉语叙词表,进行文献自动标引。这就是基于相似的相关学科的借鉴,这种借鉴为图书情报学理论体系的建立与完善提供了捷径。

反之,多向移植也是一个学科发展的重要措施。许多学科看上去毫不相关,但仍然可以移植,移植的基础仍然是相似原则。这种移植的相似性有大有小,哪怕一点点相似也可能奏效。比如文献增长定律就是移植金融学的利息增长规律而建立的。

当然,借鉴也好,移植也好,都是相互而双向的。图书情报学的成就被其他兄弟学科看重并应用也为数不少。例如,引文分析就被社会学家青睐而变成一种有效的社会评价工具。

参 考 文 献

- (1) Vladutz, G. and Cook, J. Bibliographic Coupling and Subject Relatedness. Proceedings of the 47th ASIS Annual Meeting, 1984.(21)
- (2) 松本浩一ほか,汉文资料の计量的分析:《黄帝阴符经》を注对象として, Library and Information Science, 1984, (22):1-29
- (3) 王崇德编著,文献计量学教程,天津:南开大学出版社,1990
- (4) (印)拉维昌德拉·劳著,王崇德、黄月译,图书情报学定量方法,北京:兵器工业出版社,1989
- (5) 王崇德,社会科学研究方法要论,上海:学林出版社,1990

(作者单位:南开大学分校图书馆学情报学系。来稿时间:1991.1。编发者:丘峰。)

ABSTRACTS IN PUBLICATION(159~173)

A Quantitative Analysis of the Trends of Library and Information Research Subjects in China / Qin Jun-ping // Bulletin of the Library Science in China / China Society of the Library Science.—1991,17(3).3~11

Taking 4,791 pieces of articles from 9 journal titles published in the past 10 years as objects of investigation, the article figures out 4,791 for the first subject and 1,360 for the second one. The total is 6,151. It has quantitatively profiled the general picture and the course of development of the library and information sciences in China during the past 10 years; revealed the centralization and dispersedness in the quantitative distribution of subjects; fixed the popular subjects for library and information research; analysed the 5 types of trends of subjects; reflected the interests, the focal points, the ups and downs and trends of changes of every subject. It has also played a positive role in developing and perfecting the document metrical analysis. 1 illus. 8 tables. 4 references.

Library science — Trends of research

Information science — China

Journals articles — Quantitative analysis

G250-1

An Approach to the Similarity Principle of Library and Information Science / Wang Chongde // Bulletin of the Library Science in China / China Society of the Library Science.—1991,17(3).—12~19

The similarity principle is a tie which links theories to experiences and rules and regulations to operations. The articles approaches some fundamentals of the similarity principle and explains by way of some typical examples to apply the principle to library and information science. indicating that to make use of the principle is an important orientation in developing library and information resources and going deep into library and information science. 5 references.

Library and information science — Theories

Similarity principle — Applications

Document metrology — Laws

G250

Bulletin of the Library Science in China: the Past, the Present and the Future / Qiu Feng // Bulletin of the Library Science in China / China Society of the Library Science.—1991,17(3).—20~28, 70

Up to April 1991, the Bulletin of the Library Science in China, being sponsored by the China Society of