

图书馆学译丛

tushuguanxueyicong

情报检索入门

辽宁省图书馆学会

455/05

情报检索入门

(日本)长泽雅男 著

抚顺市图书馆学会编译组 译

辽宁省图书馆学会

1983年8月

初等情報処理講座 3

情報検索入門

東京大学助教授

長沢雅男 著

森北出版株式会社 1976

东 京

情 报 检 索 入 门

(日本) 长泽雅男 著

抚顺市图书馆学会编译组译

辽宁省图书馆学会出版

辽宁省图书馆印刷厂印刷

*

$787 \times 1092 \frac{1}{32}$ 12万7千字

印数1—2500

主 编 者 序

在情报化的时代里，人们把计算机看成是主角。这可能是由于人们既认识到计算机所具有的某种特殊功能的巨大作用，又能亲眼看到计算机这个“实物”的缘故。这正如对一个驰名的研究所，由于门外汉不了解其详情，只能从表面看问题，因而认为这是一所进行着宏伟研究目标的极为壮观的建筑物而别无所知一样。从某一方面来说，计算机的功能确是非常出色的。如果我们对它的使用做到恰如其份，那么就是在现实的情况下它也是能够起到相当的作用的。我在这里使用了“相当的”这个词来形容其功能程度，而那些幻想着“计算机理想世界”的乐观论者会认为它具有无限能力而有可能冠上一个“极大的”形容词来夸耀其作用也未可知。这也许是个个人观感之不同吧。如果我们真正是站在利用它并以其不足为不足从而要求改进它的立场上，那我们就会说：

“计算机呀，莫骄傲吧！”无论如何，从现实来看，计算机若不能充分发挥作用，则确实是一大损失。计算机并非独立工作的东西，而是被人们赋予了某种功能的工具。我们应该把它作为我们头脑和手脚的补充，作为我们身体的延伸器官来大胆地广泛地运用它，如同眼睛成为我们人体的辅助部份那样。

计算机是供人们使用的工具。使用计算机的目的在于探讨情报处理的方法论，这是根本的问题。所以不掌握这方面的知识就不能有效地开展工作。这个方法论并非一目了然之“物”，所能目睹到的不过是那个受方法论支配而活动的

“机器人”即计算机之形态而已。由于它的活动引人注目，因而被大肆宣扬。就在众说纷纭有如梦呓般的奇谈怪论尚未统一的形势下，已将电子计算机付诸实际使用了。在这种局面下，人们开始认识到关于情报处理的方法论的重要性。在这里我想提醒大家，有人认为只要有了程序编制人员就能把计算机用好。这是一个大误解。为了解决问题，编制计算机程序时不仅要学会计算机语言，而且还必须熟悉分析一切事物现象用的情报处理的方法论。这与学会程序语言是两码事。

可是，当要钻研情报处理方法论时，几乎找不到适于阅读的书籍，即便能找到也异常费解。研究情报处理方法论，数学与统计学起着重要的作用，这个问题要求数据分析的意念必须与数据分析本身密切结合。数学与统计学的书籍很多，但从数据分析的立场来看还是非常不够的。这使人深切感到“死读书不会应用”是无裨于实际的。因此，从数据分析即利用计算机处理情报的观点看问题，应该筹划编写一套循序渐进、通俗易懂的基本知识讲座。这是一项极其艰巨的任务。首先，确定作者的人选就是个难题。要求作者不仅在理论上有较深的造诣，从事过许多实际工作，而且掌握数据解析的科学技术知识，熟悉计算机的功能并能自如进行操作，等等。这些都是必备的条件。另外还必须是能够掌握其奥妙并能够把必须知道的基本内容用通俗文体编写出来的人。

幸而我们找到了这样的作者，于是编辑成这一讲座。本讲座所阐述的基本观点和方法切合实际而又严谨，内容也通俗易懂。希望读者不要把它仅仅作为案头的知识，而要亲自

动手处理数据，借以理解情报处理的科学知识，希望应用这种知识能够恰如其份、得心应手地使用计算机，并把这种知识进一步演练提高而使自己成为“计算机表演能手”。只有这样，才能使计算机成为情报处理的“主角”而真正发挥出威力。

林 知己夫

1970年6月10日

前 言

当前，我们正处在一种被数量庞大的情报所包围却在急需之时又难以将所需的资料查找到的困境之中。情报的泛滥情况已经达到异常严重的程度，即使用“情报爆炸”或“情报泛滥”来形容，人们也不会认为是夸张，甚至世界上已流传着“情报公害”这种奇妙的说法。

处于这样的时代，我们若忽视情报就不能生存下去。不仅如此，如果不积极收集和选择情报或者掌握不住利用情报的技术，那么在形势急剧变化和日趋庞大与复杂的情报化的社会里就不能强有力地克服各种不利环境而生存下去。这就正如掌握敌情确定战略一样，如果不了解情报的实际情况，适当地发挥出对情报的处理能力，则势必被大量的情报所压倒。如果能很好地支配大量的情报，则情报不仅不是累赘，反而会在激烈的生存竞争中成为克敌制胜的得力伙伴。

情报本身也许是一种难以驾驭的东西，通过资料化的方法，就能使之成为具体的处理对象。图书馆历来就是进行搜集、整理和收藏那些具有代表性的资料情报的图书以供人们利用的。然而目前除图书外还有各种各样的文献在大量出版，这样，图书的情报价值就相对地有所降低。因此，图书馆若仍然继续坚持传统的图书工作方法，从情报利用角度来看，则势必会降低其应有的作用。

为此，不应该满足于这种传统作法，而应该从面向情报的角度出发来存贮情报资料，并创造出多种能在需要时即把适合情报迅速找到手的新的情报检索方法。近年来这种方法

骤然引人注目，电子计算机的问世则加剧了这种倾向。可以说，今天的情报检索技术是同电子计算机的发展相结合而精炼出来的。

但是，高度发展的各种技术并非在所有场合都能发挥其有效作用。按拘泥于情报检索系统的主要因素而言，外人乍看起来很土里土气的图书馆式的做法有时还最适用。同时，新的机械化方法在很大程度上还要依赖于传统方法，它们的基本原理上也有许多共同之处。

因此，本书从消除图书馆经过漫长岁月考验的传统方法与新方法之间的鸿沟，从促进二者之间交流的立场出发，力求把情报检索的基本问题阐述得通俗易懂，阐明在进行机械处理之前所必须解决的问题。

在编辑本书过程中，承蒙安西郁夫先生、藤川正信先生惠予了宝贵的启示，特别是佐藤和贵先生在收集参考资料和校对等方面给予了大力协助，更为主要的是得到了林知己夫先生的亲切教导以及勉励编撰本书的森北出版社的柳泽茂入先生和编辑部诸位的多方协助，在此一并表示谢意。

长泽雅男

1977年3月3日

目 次

主编者序..... (1)

前 言..... (4)

第一章 情报检索

1.1 概述 (10)

1.2 情报 (10)

1.2.1 事实 (11)

1.2.2 数据 (11)

1.2.3 情报 (12)

1.3 情报处理 (13)

1.3.1 情报时代 (13)

1.3.2 情报处理 (15)

1.4 情报检索 (18)

1.4.1 什么是情报检索 (18)

1.4.2 情报检索系统 (19)

第二章 情报的存贮

2.1 概述 (25)

2.2 构成情报源的资料 (26)

2.2.1 情报源 (26)

2.2.2 资料 (27)

2.3 资料的存贮 (31)

2.3.1 著录单元 (31)

2.3.2 资料的存贮 (34)

2.4 排档 (35)

2.4.1	什么是排档	(35)
2.4.2	字、词的顺序	(37)
2.4.3	排列规则	(40)

第三章 情报的压缩与变换

3.1	概述	(45)
3.2	资料的缩微化	(47)
3.2.1	缩微资料的种类	(47)
3.2.2	缩微资料的特征	(50)
3.3	文摘的编制	(52)
3.3.1	什么是文摘	(53)
3.3.2	文摘的种类	(53)
3.3.3	文摘的编辑	(55)
3.4	文献的分析	(56)
3.4.1	内容分析	(57)
3.4.2	主题分析	(58)

第四章 分类法

4.1	概述	(60)
4.2	等级分类	(62)
4.2.1	什么是等级分类	(62)
4.2.2	杜威十进分类法	(63)
4.2.3	日本十进分类法	(66)
4.2.4	国际十进分类法	(68)
4.3	组配分类	(75)
4.3.1	冒号分类法	(75)
4.3.2	组配分类法的构成	(80)
4.4	对分类法的评价	(85)

4.4.1	分类表的条件	(85)
4.4.2	标记法	(86)
4.4.3	分类法的评价	(92)

第五章 标引

5.1	概述	(96)
5.2	索引	(98)
5.2.1	索引的种类	(98)
5.2.2	标引的条件	(100)
5.3	先组索引	(101)
5.3.1	字顺主题索引	(101)
5.3.2	等级分类索引	(107)
5.4	后组索引	(109)
5.4.1	什么是后组索引	(110)
5.4.2	文献方式与研究方向方式	(113)
5.4.3	边缘穿孔卡片	(115)
5.4.4	单元词卡片	(123)
5.4.5	重叠比孔卡片	(127)

第六章 标引语言

6.1	概述	(130)
6.2	标题表	(132)
6.2.1	标题	(132)
6.2.2	标题表	(134)
6.3	叙词表	(137)
6.3.1	什么是叙词表	(137)
6.3.2	叙词表的结构	(139)
6.4	补正手段	(145)

6.4.1 连号	(145)
6.4.2 职能号	(147)
6.4.3 加权	(148)

第七章 情报查寻

7.1 概述	(150)
7.2 情报需要与情报服务	(151)
7.2.1 情报需要	(151)
7.2.2 情报服务	(155)
7.3 查寻	(157)
7.3.1 查寻方法	(157)
7.3.2 查寻程序	(160)
7.4 检索系统的评价	(162)
7.4.1 检索效率	(162)
7.4.2 经济性与迅速性	(166)

第八章 机械的利用

8.1 概述	(167)
8.2 关键词索引	(170)
8.2.1 关键词索引的种类	(170)
8.2.2 关键词索引的用途	(177)
8.3 二次文献的编制	(178)
8.3.1 机器标引	(179)
8.3.2 自动标引	(180)
8.4 定题情报提供	(183)
8.4.1 定题情报提供系统	(183)
8.4.2 定题情报提供的作用	(185)

译后记

第一章 情报检索

1.1 概述

什么是情报？所谓情报一词，虽然人们日常自认为已经理解而且常用，可是一旦要给它下个定义却很难。

实际上，即使不给情报一词下个定义，人们不论在谈话中还是在写文章时，从文理上也能比较正确地理解其意义。然而对于本书所讨论的情报检索来说，情报这个词是基本概念，所以为了巩固共同理解的基础，则有必要首先阐明什么是情报这一问题。

对情报不易下定义的原因之一，是这个词就现代意义来说得到广泛普及还是最近之事。很久之前，人们就有从物质或能量角度去认识事物和现象的习惯，然而从情报角度去认识事物和现象，却是从二十世纪中才开始的。

因为从情报方面来观察各种事物和现象往往是很方便的，所以人们对情报重视和关心的程度也就是日益增强起来。于一鸣惊人的情报，人们便以各自的观点下了广义和狭义的对各种定义。可以说，在情报的理解上呈现出来的混乱状况正是这种情况的反映。

为此，下面首先介绍几个定义，然后从情报检索的立场来说明应该如何理解所要讨论的情报。

1.2 情报

与情报一词有密切关联的有“事实”、“资料”和“知识”这样一些词。在日常使用时，这些词往往被混淆。正是由

于这种情况才给情报一词的意义造成混乱而难以确定。因此，只有弄清“事实”、“资料”和“知识”这些词的含义，进而明确其与情报的关系，才有助于为情报做出定义。

1.2.1 事实

首先让我们分析一下“事实” (fact) 这个词。人们常说“这是无可争辩的事实”的话，在这种场合，“事实”这个词往往具有真象这一含义。但由于对各种各样事物和现象看到或听到的人之不同，因而在接受上就有可能产生种种不同情况，也就是对事物和现象如何去领会，这将因人而异。事实是不是真象，这要取决于人们各自不同的认识与体验，所以在这里有必要把“事实”和“真象”两个词加以区别。

韦氏英语大词典对“事实”的解释是：“或多或少可以认为客观存在的事物，从而成为人们所主张的言论、陈述以至情报”。可以说，这是用人们的理智明确下来的事物和现象。

若按上述那样来解释“事实”一词，则其见解不过是一己之见，终究是从主观见解来解释而已。可是，如果经由多数人从相同的事物和现象中发现并确认出相同的事实，那么，这个事实甚至带有征求别人从理性上承认它是正确的这样一种有说服性的力量。对人们来说，只有这样的事实才能成为具有说服力的客观事实。

1.2.2 数据

什么是数据 (data)？查阅一下手头儿辞典，关于数据的定义是：“可以看作是直接观察到的事物的论据的事实。”从这个定义中可以看出数据与事实有着密切的关系。

上面已经讲过，事实属于个人所认识的问题，这些问题

以语言或文章的形式变成数据传给未见过或未弄清的问题人们。也就是说，数据将事实变换成某种符号系列而应用于交流过程中。

如果这样来理解，则事实与数据之间有着不可分割的密切关系。我们之所以利用数据，就是为了获得事实，并通过这一事实来了解该事实中所包含的事物和现象。因此，要想准确掌握住事物和现象，就必须尽量找到来自客观事实的数据。此间，人们对各种数据可靠性的辨别能力起着重要的作用。

1.2.3 情 报

我们往往把情报和数据两个词等同起来使用。实际上即使不加区别，因关系不大也无不可。但是严格地说，有时要求必须将数据和情报 (information) 严加区别。其原因是，尽管数据很多，但对特定用户来说，只有有用的数据才能算是情报。因此说，情报乃是“围绕着人们和社会团体的特定目的而评价出来的有用数据”，而数据本身则属于没有特定价值的符号系列。

如此说来，数据当被人们用来进行判断或做出决策时才被称为情报。例如，看到某个符号，如果它能用于做出决策，则这个符号就表现为情报。

以上所说的有关情报的观点，它与情报就是“对有的人在决定意图时给以一定预见性的影响”或“对某事项进行可能性选择方面起指导作用”等定义在理解上大体相同。总而言之，这些概念与“对自我保存系统（包括生物系、机械系、人类系和社会系）的特定目的有一定意义的符号及符号系列”就是情报这个广义的概念相反，可以称为狭义的概念。

下面谈谈情报与知识的异同问题。

这两个词一般是在接近相同的意义上来使用的。既然称呼不同，当然含义也就应该有所不同。从这个观点出发，列举二者的差异如下：

如上所述，如果把情报看成是按照特定目的所评价出来的数据，那么“知识”就是此情报具有更为广泛的概念。知识并非单纯的各自分类的情报，而是可以视为把某些情报以某种关联性为基础汇编起来的系统。在这个意义上来说，知识是稳定的情报。

为什么把某种知识不叫作知识而叫作情报呢？这是因为在情报交流的场所中获得了某种知识即所谓稳定情报的缘故。例如，正像将知识授予别人或从别人那里求得那样，以特定的目的来进行知识的传递或接受时，知识就被看成为情报了。

但正如人们所说，情报是“在人与人之间传递的所有符号的系列”，那么作为情报来传递的就不单纯只是知识了。

另外，情报的传递也在人与人的外界进行。但是若从情报检索角度来看，把情报传递看作是在人与人之间进行也未为不可。因此，本章姑且把情报作为一种人与人之间有意识进行交流、传递的信息的有关符号系列，表述如下。

1.3 情报处理

1.3.1 情报时代

由于以交通和通讯领域为先导，因而在科学技术领域出现了飞跃发展，使大量情报在很短时间内便能传播到世界各个角落。同时随着社会的扩大及其机构的日益高度化和复杂

化，社会对情报的需求也有急剧增加。

这种情况表现在：涉及所有领域的情报在世界各地不断产生，并通过各种渠道源源不断地涌现到人们的身边。因此许多人把这种情况看成是情报化时代的来临。

人们所接触的情报并不完全是为现在或将来所必需的。有的情报即使对某些人是有用的，但对其他人来说则可能是件麻烦事也未可知。因此，为了利用情报就必须从中选择出所需要者。根据从少量情报中选择出所需要者来进行分析、判断或创造的时代，正在成为过去。

当今正处于所谓情报时代，大量的情报正在毫无间断地产生出来。人们处于由情报本身又派生情报的社会现状下，哪里还谈得上去选择适用的情报，大有被来势迅猛、日益高涨的情报所淹没之虞。

情报洪水般地泛滥，想要得到需要的情报并加以利用，变得更加困难。当人们处于如此状态时，就有对情报产生厌烦、坠入自我疏远的险境。另外，在对情报的要求越来越迫切的社会里，假如情报不受任何约束而任其泛滥下去，则说不定情报会先于事实，反而限制人们思考，甚至出现少数控制情报的人来垄断、支配情报机构的危险。

目前之所以再三强调情报处理技术的重要性，就是为了取得恰当的情报并加以利用。这已经成为社会所关心的极其重大的问题。不仅如此，由于并非集中收集某种特定的情报，而是根据多数人的意见来收集其所需要的情报，这就造成了情报愈加难以有效利用。

情报的泛滥现象今后可能越来越严重，任其放任自流不仅对个人不利，而且对社会来说也是一个不容忽视的问题。