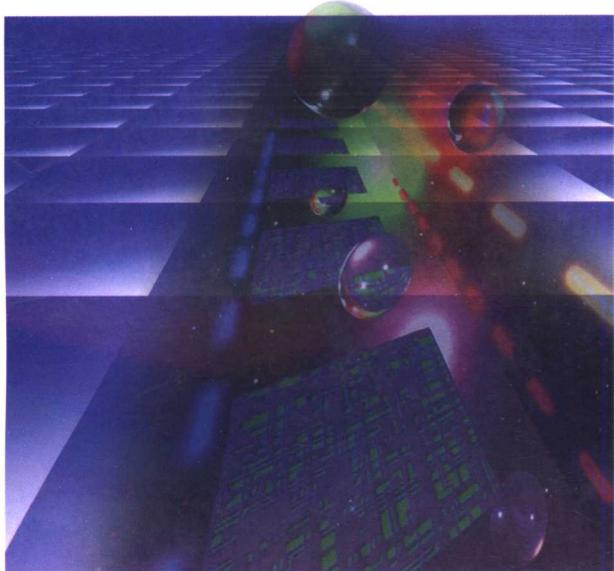


信息媒体及其采集

XinXiMeiTi JiQiCaiJi

刘兹恒 编著



北京大学出版社

信息媒体及其采集

刘兹恒 编著

北京 大学 出版 社
北 京

内 容 简 介

本书立足于信息媒体资源建设,从分析信息和信息媒体的概念、功能出发,探讨了纸质信息媒体、非纸质信息媒体、电子及网络信息媒体的类型、特点和作用;介绍了信息媒体的生产过程和管理、控制机制;并根据我国信息服务机构的实际,全面论述了信息媒体采集的原则、方法和程序;还对信息媒体自动化采访系统的建立及网络信息资源的采集进行了阐述。本书不仅可作为高等院校信息管理及相关专业的教材,也可供信息服务机构的信息采访人员参考。

图书在版编目(CIP)数据

信息媒体及其采集/刘兹恒编著. —北京:北京大学出版社,1998. 9

ISBN 7-301-03811-9

I. 信… II. 刘… III. ①信息-传播媒介-信息管理②信息-采集
IV.G202

中国版本图书馆 CIP 数据核字(98)第 17865 号

书 名: 信息媒体及其采集

著作责任者: 刘兹恒

责任编辑: 段晓青

标准书号: ISBN 7-301-03811-9/G · 0465

出版者: 北京大学出版社

地址: 北京市海淀区中关村北京大学校内 100871

电话: 出版部 62752015 发行部 62754140 编辑部 62752032

排 排 者: 兴盛达激光照排中心

印 刷 者: 北京飞达印刷厂

发 行 者: 北京大学出版社

经 销 者: 新华书店

850×1168 32 开本 9.25 印张 241 千字

1998 年 8 月第一版 1998 年 8 月第一次印刷

定 价: 16.00 元

前　　言

我们正在进入信息社会，信息在国家政治、经济、科学、文化乃至整个社会发展中所起的重要作用已日益明显地表现出来。在当前我国大力推进社会信息化以及信息社会化的过程中，建设、开发和利用信息资源就具有特别重要的意义。

然而，从信息的特征我们知道，信息本身作为一种客观存在，一方面是无时不在，无处不有；一方面却是看不见、摸不着，它只能依附于一定的物质载体才能得以保存和传递。这样，我们就把那些记录有信息，并专门用于传递交流信息的物质载体称为信息媒体。因此可以说，在很大程度上，信息的采集首先是信息媒体的采集，信息资源的建设就是信息媒体资源的建设。本书没有采用“信息采集”这一比较宽泛的概念来组织书中的内容，就是基于以上考虑。与其抽象地不分采集主体而泛泛地全面介绍信息与信息的采集，不如实实在在地以信息服务机构的任务为依据，有的放矢地论述信息媒体及其采集方法，这也更符合信息服务机构建设信息资源、服务信息用户的需要。

信息媒体及其采集作为信息管理专业的一门基础课，是从原图书馆藏书和文献资源建设课程发展起来的。作为教材，本书不仅介绍了传统纸质信息媒体，而且介绍了现代各种非纸质信息媒体，特别是对电子出版物、联机数据库、网络信息资源等进行了介绍，力求能使学生对信息媒体的产生、发展及其各自的特点、功能等有比较全面的了解。本书还结合我国信息服务机构信息媒体采访的实际，全面介绍了国内外信息媒体生产、发行情况和采访的书目工具，分析了信息媒体采集的原则和依据，并对信息媒体采集工作的组织、方法与程序进行了详细论述。为了使学生能了解信息媒体采集技术的最新发展，本

书还对信息媒体自动采访系统进行了分析,对网络信息资源的采集做了介绍。因此,本书不仅可供信息管理专业学生使用,也可供各类信息服务机构的工作人员参考。

在本书写作过程中,作者力图摆脱纯粹从经验层次上对信息媒体及其采集所做的一般介绍,而是努力追求理论上一定的深度和广度,探寻信息媒体资源发展的规律。为此,本书参考和引用了许多专家学者的著作和论文。在这里,首先要向这些原作者表示衷心的感谢!此外,本书初稿自1995年作为系内教材使用后,系主任吴慰慈教授给予了热情的鼓励;赖茂生教授、岳剑波副教授在审读后提出了很好的修改意见;张锡高老师、陈文广老师协助完成了本书第九章所需信息的网上检索与整理;研究生周菁、张久珍、梁海丽等同学在资料收集和部分文稿录入方面也分别给予了帮助,在此一并表示感谢!

由于信息媒体在现代技术支持下发展变化很快,信息媒体的采集方式也在不断出新,而作者对此跟踪研究还很不够;加上作者学识和水平有限,使本书的缺点和疏漏在所难免。殷切希望使用本书的读者提出宝贵的修改意见。

刘兹恒

1998年2月于北京大学信息管理系

目 录

第一章 信息与信息媒体	(1)
1. 1 信息概述	(1)
1. 2 信息媒体	(9)
1. 3 信息媒体的类型.....	(15)
第二章 纸质信息媒体	(19)
2. 1 图书.....	(19)
2. 2 期刊.....	(23)
2. 3 特种文献资料.....	(31)
第三章 非纸质信息媒体	(42)
3. 1 缩微型信息媒体.....	(42)
3. 2 视听型信息媒体.....	(48)
3. 3 机读型信息媒体.....	(57)
3. 4 光盘型信息媒体.....	(61)
3. 5 其他电子型信息媒体.....	(68)
3. 6 电子出版物.....	(79)
第四章 信息媒体的生产发行机制	(87)
4. 1 信息媒体的生产.....	(87)
4. 2 我国信息媒体的生产机制.....	(90)
4. 3 我国信息媒体的发行机制	(100)
4. 4 国外信息媒体的生产机制	(108)
4. 5 国外信息媒体的发行渠道	(118)
第五章 信息媒体采集的书目工具	(124)
5. 1 书目工具概说	(124)

5.2	采集中文信息媒体的书目工具	(128)
5.3	采集国(境)外信息媒体的书目工具	(135)
第六章	信息媒体采集的原则	(146)
6.1	实用性原则	(146)
6.2	系统性原则	(154)
6.3	发展与剔除原则	(160)
6.4	特色化和协调性原则	(169)
第七章	信息媒体采集的组织与方法	(179)
7.1	信息媒体采集调查	(180)
7.2	信息媒体的选择	(185)
7.3	信息媒体采集中的复本问题	(192)
7.4	信息媒体采集的方式	(199)
7.5	信息媒体采集工作流程	(212)
第八章	信息媒体采集工作自动化	(223)
8.1	采访工作自动化的发展历史	(223)
8.2	自动采访系统及其功能分析	(227)
8.3	自动采访系统的数据与文档组织	(235)
8.4	自动采访系统的发展趋势	(242)
第九章	网络信息资源及其采集	(246)
9.1	网络信息资源	(246)
9.2	网络信息资源的采集软件	(249)
9.3	网络信息资源的采集	(256)
9.4	网上书店	(282)
参考文献		(288)

第一章 信息与信息媒体

人类社会正在进入信息社会。信息作为当今世界推动社会生产力发展的新的动力，正日益受到人们的重视。信息同能源、材料一起，被看作是人类生产与生活必不可少的三大资源。人们的衣食住行离不开信息，社会的交流沟通离不开信息，科技的发展进步离不开信息，经济的繁荣兴旺离不开信息，国家民族的强大更是离不开信息。现在，不仅人与人之间的竞争、企业与企业之间的竞争，甚至国家之间国力的竞争，在相当程度上其实质也都是信息的竞争，谁先掌握了信息，谁掌握的信息较多，谁就能立于不败之地。

1984年，邓小平同志“开发信息资源，服务四化建设”的指示，为我国信息产业的发展指明了前进的方向。目前，信息的采集、开发和利用，已成为我们发展教育科学文化、发展国民经济的一项基本手段，信息的重要性已愈来愈深入人心。

1.1 信 息 概 述

1.1.1 什么 是 信 息

信息一词对于我们来说并不陌生，我们几乎每天都能听到、看到、接触到各种各样的信息。在哲学、自然科学和社会科学等各个领域，信息都是最基本的概念之一。信息无所不有、无处不在，无始无终。但是如果我们要问什么是信息，答案恐怕会是五花八门。毫不奇怪，因为信息这个概念本身就是一个虽应用十分普遍，但内容相当丰富，含义也相当模糊的概念。即使学术界，至今也未形成一个统一完整而又得到公认的信息定义。不同学科的专家，只是从不同角度来定

义信息，例如，哲学家认为信息就是认识论；数学家认为信息就是概率论；语言学家认为信息的英文单词是 information，实为 in + formation 组成，因此拿进来并整理成章就是信息；……还有很多很多。所有这些看法都有各自的道理，但我们认为，它们都不能全面反映信息的实质。

据考证，信息一词在我国最早见于《三国志》，其中记载诸葛亮围合肥新城，城中遣士刘整出围传消息，王子俭期曰：“正数欲来，信息甚大。”一千多年前，唐代诗人李中在《碧云集》中也有这样的诗句：“梦断美人沉信息，目穿长路倚楼台。”由此可见，信息一词在我国古代就是消息音信的意思。

而信息一词被引入科学领域，则是从 1948 年美国数学家、通信工程师申农(C. E. Shannon)创立信息论开始。申农在他的《通讯的数学理论》一书中，给信息下了如下定义：“信息是二次不定性之差。不定性就是对事物认识不清楚，不知道。信息就是消除人们认识上的不定性。”在这之后，美国数学家、控制论的创始人维纳(N. Wiener)也给信息下了一个定义。他在《控制论和社会》一书中指出：信息是“我们用于适应外部世界，并且使这种适应为外部世界所感到的过程中，同外部世界进行交换的内容的名称。”专家们对信息的解释，有其一定的合理性，但受到他们各自学科的局限，仅从功能的角度来认识信息，似乎也不够完善。

尽管在理论上人们还在对信息的涵义进行艰辛的探讨，但在现实生活中，信息一词的使用已十分普遍。在这里，我们不必过多地纠缠信息的科学定义，而只需从信息管理的角度来了解什么是信息，以便能更好地理解信息媒体的概念。

我们认为，信息的概念可以做广义和狭义的区分。广义的信息是指，对各种事物的存在方式、运动状态和相互联系特征的一种表达和陈述，是自然界、人类社会和人类思维活动普遍存在的一切物质和事物的属性。这种广义的认识可以使我们看到信息存在的普遍性，有利

于我们扩展认识信息和收集信息的范围，并注意对尚未被开发的一些潜在信息资源进行开发利用。狭义的信息则是指，具有新内容或新知识的消息，即对接收者来说预先不知道的报导。这种狭义的认识，可以帮助我们把目光集中到那些与现实发展密切相关的信息的采集上，使信息的采集更有效益性。

这里，要注意信息与消息、信息与知识等不同概念的区别。一般来说，消息的含义比信息要广。虽然信息是消息的内核，但信息只是能够给人们带来新知识的消息。一则消息可能包含非常丰富的信息，也可能只包含很少的信息。同样，信息能够增长人的知识，但这并不意味着信息就是知识。知识是一种特定的人类信息，信息只有经过科学的系统的加工，才能上升为知识。知识是经过浓缩、系统化和优化了的信息。

1.1.2 信息的特征

尽管人们对信息的定义还在进行不懈的探索，但人们对信息本质特征的认识却在逐渐趋于一致。尤其是信息基本特征的揭示，使我们能够对信息概念作出全面的了解。

信息具有以下一些基本特征：

1. 真实性

信息是事物存在、运动、变化特征与规律的客观描述，具有客观真实性。正因为如此，信息必须反映真实情况才能供使用者利用，用户只有依据可靠的信息才能做出正确的决策。不反映客观实际的信息是虚假信息，不但没有价值，反而会导致相反的效果。因此，真实、准确是信息的首要特征。

2. 知识性

对于信息用户来说，信息的内容一定要是未知的，否则就不能算是信息。信息之所以能消除人们认识上的未知性或不定性，正是由于它具有知识的特性。人们对客观事物发展的规律不清楚，就是对该事

物缺乏必要的知识。信息能够改变用户的知识状态,由不知到知,或由知之较少变为知之较多,从而使自己的认识由不清楚、不确定向清楚、确定转化。可以说,一切知识都是信息赋予的,人们获得了信息,也就是获取了某种知识。

3. 指向性

信息是一种客观存在,其数量是无穷的,但并不是所有的信息对每一个人都是有用的,其效用如何,与信息用户的需要有密切联系。人们的社会分工不同,对信息的需要也不会相同。涉及到分工范围内的信息,对他就是有用信息,否则是无用的信息。因此,信息总是具有一定的指向性,即某一信息往往向一定的使用者流通,表现为人们对信息进行有目的的选择利用;信息向着需要的用户进行有目的的定向传播。人类的分工越细,有效的信息也越专业化;而信息数量越多,人们选择利用信息的难度也越大。

4. 共享性

信息作为一种资源,可以供全世界全人类共同享用。它不是某个个人、某个社会、某个国家的私有品和专利品。也就是说,当某一信息被某一用户利用后,仍然可以为其他用户所使用。即信息可以同时被多个用户占有和使用,这一特征与物质和能量的唯用性不同。物质、能量被某人占有和利用后,其他人就没有了;如果占有者将自己拥有的物质和能量转让给他人,那么他就失去了这些被转让的物和能。而信息却不会因为利用的人数增加而丝毫减少。如一篇文章,一个人看和十个人看,提供的信息量都是一样的,但看的人越多,信息实现的价值就越大。

5. 无限性

信息作为关于事物运动状态和方式的知识,是永不枯竭的。只要事物在运动,就有信息存在;只要人类认识和改造世界的活动不停止,大量的信息就会源源不断地产生。信息是一种取之不尽、用之不竭的资源,不会像材料和能源那样发生资源短缺的危机。另一方面,

信息的无限性还表现在它的可扩充性。人们对信息占有越多，使用越多，则信息的用处也越大。绝大多数信息在应用过程中（如科学研究、技术转让等），可以不断得到扩充。所以信息的无限性表现在两个方面：一是客体产生信息具有无限性，二是主体利用信息的能力以及信息所产生的作用具有无限性。

6. 时效性

信息的价值与作用是体现在一定的时空范围之内的，它不但取决于信息内容本身，还取决于该信息是否能被人们及时获得。过时的信息是贬值的，甚至毫无价值。可见信息具有鲜明的时间效应。信息只有在得到及时利用的情况下才会有理想的使用价值。

7. 依附性

载体是信息存在的必要条件。信息的存储、传递和交流必须依附于一定的物质载体。信息本身是看不见、摸不着的，它只能附着在载体上，以一定的形式表现出来。因此，人们要获得信息，首先要获得携有信息的载体，通过对载体的利用，才能解析出其中的信息内容。这也正是我们研究信息媒体、采集信息媒体的意义所在。

1.1.3 信息的类型

为了更好地掌握与利用信息，我们有必要对信息进行分类。但由于信息定义具有多样性，使得信息分类标准也具有多样性。在实际工作中，人们常采用不同的标准把信息划分为不同类型。

(1) 按信息的产生和作用机制划分，信息可分为自然信息和社会信息两大类。自然信息是指宇宙间、自然界客观存在或随机发生的各种事物、事件的信息，是物质运动和生物生存活动的产物。它随着事物的存在而存在，随着事物的变化而变化，是自然界的客观反映。自然信息包括生命信息，动植物“活动”信息及物质物理信息等。人们通过对自然信息的利用，可以达到认识世界、改造世界的目的。社会信息则是指在人类或社会维系生存、生产和发展过程中所产生、传

递、利用的信息，是人和人之间发生交往的产物，也是人类从事一切活动的基础。社会信息包括经济信息、政治信息、法律信息、文化信息、军事信息、科技信息、社会生活信息等。人类越进步、社会越发达，所产生的社会信息就越多。一般来说，目前人们较多感知或认识的信息是社会信息，而对自然信息的感知深度则是随着科学技术的发展而逐渐加大的。

(2) 按信息的功能划分，信息可分为功能性信息和非功能性信息两种。功能性信息主要由科学信息、技术信息和生产信息所组成，是人们从事社会活动和工作所必需的，对社会的生存与发展具有重要意义。非功能性信息一般指文学、艺术、电影、音乐等文化信息；衣、食、住、行等生活信息。它也是社会生活中不可缺少的信息。

(3) 按信息的运动状态划分，信息可分为动态信息和静态信息。那些时间性很强、处于不断变化的信息如新闻、市场信息、金融信息等就是动态信息；而那些相对稳定、固化的信息如存在于各种文献中的信息，则是静态信息。

(4) 按信息是否被人感知来划分，有显性信息和隐性信息两类。显性信息是指人们已经感知到了的信息，如一条新闻、一则广告，或对某事物特征、状况的报告和描述等，都是显性信息。只有显性信息才能直接为人类所利用。隐性信息是指人类还未能感知到的事物所包含的信息。这种信息尽管也是客观存在，但就目前来说，还不能够直接为人类所利用。

(5) 按信息的存在和接收情况划分，有直接信息和间接信息两类。事物的存在状态和运动形式本身在人们头脑中的直接反映（如观察到的现象等）是直接信息；关于事物存在状态与运动形式的描述（如经过加工整理后的数据、资料、理论、观点等）则是间接信息。生物能利用直接信息。比如生物为了生存，就不得不从它所处的环境中获取与它有关的信息，加以识别和处理，并做出适于外界环境的反应。只有人类才能利用间接信息。人类为了表达各种各样的信息，创造了

语言,创造了文字,创造了能够储存与传递信息的载体。这使得人类开发与利用信息的能力大大提高。

(6) 按信息的处理、加工程度来划分,还可把信息分为一次信息、二次信息、三次信息。一次信息是指未经加工或略微加工的原始信息,如会议记录、论文、专著、统计报表等;二次信息是在原始信息基础上加工整理而成的供检索用的信息,如文摘、书目、索引等;三次信息是根据二次信息提供的线索,查找、使用一次信息和其它材料,进行浓缩、整合后的信息,如研究报告、综述、述评等。

除了上述分类以外,还可按信息的传播范围把信息分为公开信息、半公开信息和秘密信息;按信息的符号表达形式把信息分为语言信息和非语言信息;按信息的外化结果把信息分为有记录信息和无记录信息;按信息的使用价值把信息分为有用信息、无用信息和有害信息,等等。

1.1.4 信息的作用

随着人类社会的发展,人们对信息的生产、传递、处理和利用的能力都有了很大提高,信息的作用也因此发生了质的变化,正在成为社会发展的主导。有人为信息的作用归纳了八大功能,即生存的资源、知识的源泉、决策的依据、控制的灵魂、思维的材料、实践的准绳、管理的基础、组织的保证。我们认为,这些评价丝毫不过分。具体来看,信息在人类活动中的作用主要表现在以下几方面:

1. 信息是科学决策的依据

决策是管理的核心内容。任何企业事业单位要实行科学的管理,都必须保证决策的科学性。影响决策的因素很多,如决策人员、决策方法、决策环境与背景、决策所依据的信息等。其中最重要的就是信息。由于信息反映了客观事物的属性,因而是科学决策的前提和根据。在决策过程中,从确立目标到拟定方案、分析方案,再到选择方案、组织实施与监测,每一个环节都离不开信息的支持,只有以完整、

准确、可靠的信息为依据，才能保证决策的科学性。那种凭“长官意志”，“头脑发热”的决策方法，只能造成决策的失误。

2. 信息是提高经济效益的重要因素

任何经济单位都要以提高经济效益为自己的目标。根据价值工程理论，我们知道产品的价值是由产品的功能和产品的成本决定的，即

$$V(\text{价值}) = \frac{F(\text{功能或技术指标})}{C(\text{成本或经济代价})}$$

设未利用信息的情况下， $V_1 = \frac{F_1}{C_1}$ ；

利用信息的情况下， $V_2 = \frac{F_2}{C_2}$

由于信息本身具有价值性，因而利用信息能够使产品成本下降。故当 $F_1=F_2$ 时，即产品的技术指标一定时，因 $C_1>C_2$ ，则有 $V_2>V_1$ 。这说明利用信息所生产产品的价值较之未利用信息所生产产品的价值要大。这也充分说明了信息开发和利用的过程，同时也是信息价值转移和实现的过程。在实际生活中，诸如一条信息拯救一家企业，一条信息为国家挽回重大经济损失的事例不胜枚举。

3. 信息是推动科技进步的手段

我们所处的世界是竞争的世界。个人之间的竞争、单位之间的竞争、国家之间的竞争，关键是科技方面的竞争。我们要想在激烈的竞争中始终处于不败之地，就必须下大力气推进科技进步。而促进科技进步最简单、最有效的手段就是广泛收集、引进、消化和吸收国内外的科学技术信息。科技信息不仅能帮助我们及时掌握国内外最新的研究动态，避免重复研究别人已有的研究成果，还能够帮助我们直接采用别人的先进技术，使我们的科技进步始终处于高起点。

4. 信息是市场竞争的关键

在市场经济的条件下，企业的发展无时无刻不与市场保持着联

系。市场既是企业生产的出发点,又是企业生产的归宿,脱离了市场,企业的发展就会受到制约,甚至威胁到生存。为此,企业不得不把市场调查和商品销售分别看作是生产过程的前奏和延伸,并把它们看成是企业经营活动的主要内容。实质上,市场调研的过程就是对各类市场信息进行搜集、获取、加工、整理和分析的过程;商品销售的过程也包含着将经过加工、处理的信息以广告、宣传和实物输出的形式向市场传递的过程。显然,市场竞争,在很大程度上是信息竞争,市场信息已经日益成为当今世界经济结构的核心。

上面所举的只是信息在当代社会中的一部分作用,远未包括信息的全部功能。随着信息业的发展,信息的应用还会越来越广。

1.2 信 息 媒 体

1.2.1 信息媒体的含义

从信息的特征我们已经知道,信息本身不是物质,也不具有能量。因此,任何信息,其存在和传递都离不开一定的物质载体和能量。信息与信息载体是密不可分的,没有载体,信息就无法存在。换句话说,信息总是以一定类型或形式的载体作为其存在和传递的物质基础的。

信息载体根据其功能的不同,可以分为信息载体和载体的载体两大类。所谓信息载体,指信息内容赖以存在的语言文字、符号、声波、光波、电波等用以记录信息的载体,称为第一载体或软载体;所谓载体的载体,即承载信息的物质载体,指信息载体赖以存在和传递的纸张、胶片、磁带、光盘等物质材料,称为第二载体或硬载体。只有第一载体和第二载体有机地结合,才能保证信息内容得到完整的保存和有效的传递使用。

为了便于说明问题,我们先把各种承载了信息的物质载体看成是信息源。在人们的社会生活中,常用的信息源主要有三类:

1. 口头信息源

口头信息源是指以人脑为信息承载体的人与人之间的对话、交谈、讨论、演讲等。社会生活中的每一个人，每天都在进行信息的口头交流，不断地发出信息又不断地接收信息。口头信息源在利用上有许多优势，表现在：① 获取信息快，交流及时。这是其他信息源所不具备的；② 具有高度的选择性和针对性。例如从研究同一课题的同行处获取有关的口头信息，较之查找大量散见于各种文献中的信息，显然既方便、又对口；③ 信息反馈及时、准确。在直接交谈时，对于所涉及的问题双方可以立即询问和澄清，必要时还可以互相补充和修正，避免理解上的错误；④ 根据人们接触时的气氛、语气、手势、暗示等，可以领会语言文字所不能表达出来的东西，因而有助于全面了解对方所谈内容，对所获得的信息进行恰当的评价；⑤ 可以提供其他信息不能包含的许多细节，而这些细节往往更具有启发和借鉴作用。但是，口头信息源的缺点也是显而易见的。由于人们直接通过口头交流来获取信息的机会总是有限的，机会分布也不均衡，所获信息的满意度和完整性极不稳定；由于难于对口头交流实行有效的社会监督，故其可靠性也不易检验，加之难于积累和整理，信息会随时间推移而失真和泯灭，所以人们在获取、传递和利用口头信息时，常常将它转换成文献信息。

2. 实物信息源

实物信息源是指蕴含着信息的实物载体，包括自然实物信息源和人造实物信息源两大类。自然实物信息源如植物、动物、河流、山川、宇宙、矿产、土地等，它们是未经人们加工的自然存在，能直接反映事物的原始状态，人们可以从中获得有关自然界的信息。如气象的变化、矿产的蕴藏，动植物的繁衍等。人造实物信息源是经过人类有目的地加工而成的物体，如产品、样品、样机等。它是人类活动的产物，具有内隐性。其内涵为加工工艺、化学成分、物质参数、设计指标、外观状态等。对这类信息源的利用，需要通过较为复杂的分析研究，