

 山东大学出版社
Shandong University Press

现代信息检索

贺志刚 李修波 主编



现代信息检索

贺志刚 李修波 主编

山东大学出版社

图书在版编目(CIP)数据

现代信息检索/贺志刚,李修波主编. —济南:山东大学出版社,
2003.8

ISBN 7-5607-2591-0

I. 现…

II. ①贺…②李…

III. 情报检索-高等学校-教材

IV. G252.7

中国版本图书馆 CIP 数据核字(2003)第 063918 号

山东大学出版社出版发行

(山东省济南市山大南路 27 号 邮政编码:250100)

山东省新华书店经销

莱芜圣龙印务书刊有限责任公司印刷

787×980 毫米 1/16 17.25 印张 326 千字

2003 年 8 月第 1 版 2003 年 8 月第 1 次印刷

印数:1—3500 册

定价:24.80 元

版权所有,盗印必究!

凡购本书,如有缺页、倒页、脱页,由本社发行部负责调换

本书编委会

主 编 贺志刚 李修波

主 审 韩子军

编 者 (按姓氏笔画排列)

马建春 王勋鸿 李修波

宋文彬 贺志刚 程川生

序

现代教育、科研和学习所必须面对的一个事实是，虚拟现实（virtual reality）所带来的信息高速流动使地球已变成一个小小的村落。“秀才不出门，便知天下事”的神话已成为了现实，信息越来越重要地影响人们的生活，因此很多人称现代社会为信息社会。

信息社会给人们带来了浩如烟海的信息。据有关资料统计，现在世界每年发表的科技论文约 1500 万篇，出版图书约 70 万种，申请专利约 100 万件，此外还有大量的特种文献出版。全世界传递的信息量以每天近百亿信息单元的速度递增，并以 15%—20% 的加速度不断发展。近 30 年来，人类生产的信息已经超过过去 5000 年的信息量的总和。这大量的知识与信息可以历史性地满足人们的需求。

面对如此庞大的信息集合，非常现实的问题是，如何才能尽快获取自己所需要的信息呢？令人备感庆幸的是，光荣的计算机网络技术经过近半个世纪的发展，已为准确、快捷、完整地获取信息做好了漂亮的嫁衣。计算机技术的成熟与发展大大丰富了包括载体类型、传播手段、组织形式等传统的信息处理技术。网络技术的发展，则使海量信息的远距离获取成为现实（即时完成信息的传递无疑已成为信息社会最精彩、最动人的场景之一）。信息社会信息的获取更加强调“检索”和“分析”而非简单的拥有。

话已至此，我不能不谈到另外一个重要的概念：“信息素质”（Information Literacy）。它最早是由美国信息产业协会前主席保罗·泽考威基（Paul Zurkowski）1974 年提出。他认为信息素质是人们在工作中运用信息、学习信息技术、利用信息解决问题的能力。1998 年，美国学术图书馆协会和教育交流技术协会研究制定了信息素质培养标准，其中非常重要的一条就是具有良好的信息意识，并能在信息检索方面占优势。

单就信息意识而言，它是指信息用户在存在信息需求时的主观能动性，表现在信息检索能力上，信息意识强，信息获取的能力就强。因为目前科研工作的各个环节都与信息获取密切相关，只有具有较强的信息意识，才会有效提高信息检索的效率，才会在信息资源的获取过程中发现问题，逐步修正检索和分析策略，从而得出正确的结论。



同时我认为,信息检索即从各种信息源查找、评价和使用信息的能力是信息素质深蕴之义,是信息社会的成员必须掌握的技能。良好的检索能力体现在以下几个方面:能够熟练制定检索策略;掌握基本的检索技术;能够准确地评价信息;能根据实际用途组织信息;能将新信息融会到现有知识结构中,在批判性思考、解决问题和交流的过程中使用信息。

我们赞叹制定标准者的匠心独运,如果没有信息意识,信息检索就无法明确努力的方向;没有信息检索,则真正的信息意识根本就无法树立起来。

现在,已经没人再否认信息检索在科研工作中的重要性了。事实上,技术高超的信息检索工作常常会使信息需求者惊叹“原来事情的解决可以这么简单”。同时,当人们遇到一些具体的问题的时候,都开始试着问一句:我到哪儿去查这类信息?这类信息准确吗?是否还有更为科学的信息?这无疑是一种令人欣慰的现实写照。

自20世纪80年代初期尝试着在学校开设信息检索课至今,已有20多年了。在此期间,数十万上过检索课的学子已经走上了工作岗位,并在各自的工作中熟练地应用着检索知识。自感在“德艺双馨”人才的培养上做了一点点事情,心里非常的宽慰。

我的几位从事信息检索工作多年的同仁将要出版一本信息检索方面的书,请我做序。一方面我要向他们表示祝贺,祝贺他们在这一领域探得新知;另一方面我也想提出一点要求,那就是希望他们积极进取,能够在今后学校信息素质教育的工作中,发挥更大的作用。

一下子说了这么多,归根结底一句话:一切为了读者。

于山东大学

2003.7.1

目 录

序.....	(1)
第一章 绪 论.....	(1)
1.1 信息与信息检索	(1)
1.1.1 信息	(1)
1.1.2 信息、知识、情报与文献	(3)
1.1.3 信息检索	(4)
1.2 学习信息检索的意义	(7)
1.2.1 可以迅速适应信息社会的发展	(7)
1.2.2 有助于增强情报意识	(8)
1.2.3 培养基本的情报检索技能.....	(10)
1.3 信息检索的学习方法.....	(10)
1.3.1 理论联系实际.....	(10)
1.3.2 比较法.....	(11)
1.3.3 例证法.....	(11)
第二章 信息检索概论	(14)
2.1 信息检索原理.....	(14)
2.2 信息源.....	(15)
2.2.1 什么是信息源.....	(15)
2.2.2 当代信息源的特点.....	(15)
2.2.3 信息源的类型.....	(16)
2.2.4 信息发表类型的识别.....	(22)
2.3 检索语言.....	(27)
2.3.1 什么是检索语言.....	(27)
2.3.2 检索语言的特征.....	(27)
2.3.3 检索语言的类型.....	(27)
2.4 检索工具.....	(34)



2.4.1	检索工具的定义	(34)
2.4.2	检索工具的基本类型	(34)
2.4.3	检索工具的一般结构	(36)
2.5	信息检索的方法、途径及评价	(37)
2.5.1	检索方法	(37)
2.5.2	检索途径	(38)
2.5.3	情报检索效果的评价	(41)
2.6	信息检索的一般步骤	(43)
2.6.1	分析研究课题	(43)
2.6.2	确定检索工具(选择检索系统)	(44)
2.6.3	选择检索方法	(44)
2.6.4	确定检索途径和检索策略	(45)
2.6.5	实施检索并调整检索策略	(45)
2.6.6	索取原始文献	(45)
第三章	传统检索工具概述	(50)
3.1	传统检索工具	(50)
3.1.1	传统检索工具的定义	(50)
3.1.2	传统检索工具的特点	(50)
3.1.3	传统检索工具的基本类型	(50)
3.1.4	传统检索工具的历史作用和现实意义	(51)
3.2	国内传统检索工具综述	(51)
3.2.1	发展历史	(51)
3.2.2	特点	(52)
3.3	国外传统检索工具综述	(52)
3.3.1	概况	(52)
3.3.2	特点	(52)
	附录:国内主要中文检索工具	(53)
第四章	传统检索工具检索及利用	(55)
4.1	国内传统检索工具的使用	(55)
4.1.1	查找整本文献的检索工具	(55)
4.1.2	查找论文的常用传统检索工具	(56)
4.2	国外传统检索工具的使用	(57)



4.2.1	美国《工程索引》	(57)
4.2.2	科学引文索引(SCI)	(58)
4.2.3	科学文摘(SA)	(61)
4.2.4	《化学文摘》(CA)	(64)
4.2.5	《金属文摘》(MA)	(67)
4.2.6	《文摘杂志》	(69)
4.2.7	《科学技术文献速报》	(72)
第五章 电子检索工具概论		(76)
5.1	概述	(76)
5.1.1	电子检索工具的定义、产生及发展	(77)
5.1.2	电子检索工具的特点及实施检索的必备要件	(77)
5.1.3	电子检索工具与传统检索工具的比较	(79)
5.1.4	电子检索工具的类型	(79)
5.2	电子检索工具的“正文”——信息数据库	(80)
5.2.1	数据库的定义和类型	(80)
5.2.2	构成	(81)
5.2.3	特点	(82)
5.3	检索技术及检索策略	(83)
5.3.1	布尔逻辑(Boolean Logical)算符	(83)
5.3.2	优先处理算符“()”或“ ”	(83)
5.3.3	位置算符	(84)
5.3.4	截词符	(84)
5.3.5	限制符与检索限制	(86)
5.3.6	加权检索	(86)
5.3.7	检索策略的制定与修订	(86)
5.4	电子信息检索的主要方式	(87)
5.4.1	从电子检索工具所依托的物理载体而言,可分为光盘检索、 网络学术资源数据库、 <u>互联网信息检索</u> 等	(87)
5.4.2	按照检索工作开展所依据的检索技术,可分为初级检索、 高级检索	(89)
第六章 电子检索工具的应用		(93)
6.1	电子检索工具应用概述	(93)



6.2 获取整本文献的电子检索工具	(94)
6.2.1 获取图书的电子检索工具	(94)
6.2.2 获取期刊的电子检索工具	(99)
6.2.3 UnCover 及 CALIS-UnCover	(103)
6.2.4 其他用于获取整本文献的常用检索工具	(103)
6.3 获取学术论文的电子检索工具	(104)
6.3.1 获取学术论文的文摘性电子检索工具	(104)
6.3.2 获取学术论文的全文性电子检索工具	(126)
第七章 因特网信息检索	(172)
7.1 因特网信息资源	(172)
7.1.1 因特网信息资源的概念	(172)
7.1.2 因特网信息资源的类型、特点	(173)
7.2 因特网信息资源检索工具	(174)
7.2.1 搜索引擎概述	(174)
7.2.2 搜索引擎基本类型	(174)
7.2.3 搜索引擎的检索方式及检索技术	(177)
7.2.4 搜索引擎的一般检索步骤	(178)
7.2.5 检索策略与技巧	(181)
7.2.6 一种实用的搜索引擎——Google	(183)
第八章 特种文献的检索	(187)
8.1 特种文献概述	(187)
8.2 专利文献及其检索	(187)
8.2.1 专利制度及其沿革	(187)
8.2.2 专利的种类	(188)
8.2.3 专利申请的条件	(188)
8.2.4 专利的审查制度	(189)
8.2.5 专利文献及其特点	(190)
8.2.6 国际专利分类表	(191)
8.2.7 中国专利文献的手工检索	(193)
8.2.8 中国专利信息检索系统	(194)
8.2.9 英国德温特《世界专利索引》及其利用	(195)
8.2.10 欧洲专利信息网	(196)



8.3 会议文献及其检索	(198)
8.3.1 概述	(198)
8.3.2 印刷型的检索工具	(199)
8.3.3 会议文献的计算机检索	(199)
8.4 学位论文及其检索	(203)
8.4.1 学位论文概述	(203)
8.4.2 学位论文的手工检索	(204)
8.4.3 学位论文的计算机检索	(206)
8.5 标准文献及其检索	(212)
8.5.1 概述	(212)
8.5.2 标准文献的种类	(213)
8.5.3 标准文献的特点	(214)
8.5.4 标准文献的检索	(214)
8.6 技术报告及其检索	(217)
8.6.1 概述	(217)
8.6.2 技术报告的类型	(218)
8.6.3 技术报告的特点	(218)
8.6.4 美国四大报告简介	(219)
8.6.5 技术报告的重要检索工具——《政府报告通报及索引》	(221)
8.6.6 技术报告的其他检索工具	(226)
8.6.7 技术报告的计算机检索	(227)
第九章 事实与数据检索	(230)
9.1 概述	(230)
9.2 参考工具	(230)
9.2.1 参考工具的特点	(231)
9.2.2 参考工具的选择	(231)
9.2.3 参考工具的类型	(232)
9.3 参考工具介绍与利用	(232)
9.3.1 词典	(232)
9.3.2 百科全书	(235)
9.3.3 手册、便览	(239)
9.3.4 年鉴	(240)
9.3.5 名录、指南	(242)



9.3.6 其他类型参考工具选介	(244)
9.4 事实与数据检索的一般步骤与实例	(245)
9.4.1 检索步骤	(245)
9.4.2 检索实例	(246)
第十章 信息的整理、分析及利用	(249)
10.1 信息的搜集和整理	(249)
10.1.1 信息的搜集	(249)
10.1.2 信息搜集的方法和途径	(251)
10.1.3 信息的整理	(251)
10.2 信息的分析与研究	(251)
10.2.1 信息分析与研究工作的目的与作用	(252)
10.2.2 信息分析研究的性质	(252)
10.2.3 信息分析研究工作的必要性	(252)
10.3 信息分析与研究的特点	(253)
10.3.1 针对性	(253)
10.3.2 预测性	(253)
10.3.3 科学性	(254)
10.3.4 综合性	(254)
10.4 信息分析和研究工作的类型	(254)
10.4.1 竞争对手的发展趋势信息分析研究	(255)
10.4.2 专项课题的信息分析	(255)
10.4.3 市场信息的分析与研究	(255)
10.4.4 管理信息分析和研究	(256)
10.4.5 政策信息分析和研究	(256)
10.5 信息分析与研究的主要环节	(257)
10.5.1 选题	(257)
10.5.2 制订计划	(258)
10.5.3 信息的搜集	(259)
10.5.4 信息的鉴别和整理	(260)
10.5.5 分析与研究	(261)
10.5.6 撰写研究报告	(262)
后 记	(263)

第一章 绪 论

学习指南

- 信息、文献、情报、知识的基本概念
- 信息的分类、特点
- 信息检索的概念、意义
- 信息检索与文献检索的区别
- 信息检索课的学习方法

1.1 信息与信息检索

1.1.1 信息

1. 什么是信息

人类社会与信息紧密相连。人类不仅需要对生于斯的世界有所了解，而且不同人群之间需要协作与沟通，因此人类不仅时刻需要获得自然界信息，也需加强相互之间的通信和交流。近年来，赛伯空间（cyberspace，又称为网络空间）和虚拟现实（virtual reality）的出现已使“秀才不出门，便知天下事”的神话成为了现实，信息高速流动使地球已变成一个小小的村落。信息越来越重要地影响着人们的生活，准确全面地了解一下信息的含义非常必要。

“信息”一词最早出自南唐诗人李中《暮春怀故人》中的“梦断美人沈信息，目穿长路倚楼台”^①，原意指消息。英文 information 一词可常译作情报、资料、消息、报道、知识的意思。因此，很多人长期以来就把信息看作是消息的同义语，简单地把信息定义为能够带来新内容、新知识的消息。但是后来发现信息的

^① 见南唐李中《暮春怀故人》：“池馆寂寥三月尽，落花重叠盖莓苔。惜春眷恋不忍扫，感物心情无计开。梦断美人沈信息，目穿长路倚楼台。琅环绣段安可得，流水浮云共不回。”引自《全唐诗》748卷，北京：中华书局1960版，第8519页。



含义要比消息的含义广泛得多,不仅消息、情报是信息,指令、代码、符号语言、文字等,一切含有内容的信号都可称为信息。

在信息概念的发展过程中,许多知名学者都试图给出信息的准确含义。如信息论的奠基人、美国数学家申农(C. E. Shannon, 1916—2001)在《在噪声中的通信》一文中从通信学的角度将信息定义为:信息是不确定性的减少或消除。随后控制论的创始人维纳(Norbert Wiener)从控制论的角度提出了信息定义:“信息这个名称的内容是我们对外界进行调节并使我们的调节为外界所了解时而与外界交换的东西”、“信息是有秩序的量度”。一些知名词典的编纂者也试着给出了信息的定义,如《辞海》的注释是:“信息是指对消息接受者来说预先不知道的报道。”《简明自然辩证法词典》的解释是:“信息一般泛指我们听说的消息、情报、指令、数据、信号等有关周围环境的知识。”

总结以上信息的定义,我们认为,信息是事物或事物之间不确定性的量度,即负熵。掌握的信息越多,所需认识的事情的不确定性就会越少。

2. 信息的特点

信息特征说法众多,总结如下:

(1) 客观真实性。客观、真实是信息的最重要的本质特征,这也是信息的生命所在。

(2) 针对性。信息的作用和价值可以随接受者的不同而不同。

(3) 传递性。传递是信息的基本要素和明显特征。信息只有借助于一定的载体(媒介),经过传递才能为人们所感知和接受。没有传递就没有信息,更谈不上信息的效用。

(4) 时效性。信息的功能、作用、效益都是随着时间的延续而改变的,这种性能即信息的时效性。时效性是时间与效能的统一性,它既表明信息的时间价值,也表明了信息的经济价值。一个信息如果超过了其价值的实用期就会贬值,甚至毫无用处。如金融信息,在需要知道的时候,会非常有价值,但过了这一时刻,这一信息就会毫无价值。又如战争时的信息,敌方的信息在某一时刻有非常重要的价值,可以决定战争或战役的胜负,但过了这一时刻,这一信息就变得毫无用处。所以说,信息有非常强的时效性。

(5) 有用性(或称目的性)。信息是为人类服务的。它是人类社会的重要资源,人类利用它认识和改造客观世界。信息可以多次利用且无损耗。有时不仅不会损耗,而且增值,可以产生出新的信息。

(6) 可处理性。这一特征包括多方面内容,如信息的可拓展、可引申、可浓缩等。这一特征使信息得以增值或便于传递、利用。同时信息可以廉价复制,可以广泛传播。信息的复制不像物体的复制,一条信息复制成100万条信息费用十



分低廉。尽管信息的创造可能需要很大的投入,但复制只需要载体的成本,可以大量地复制,广泛地传播。

(7) 可共享性。信息与一般物质资源不同,它不属于特定的占有对象,可以为众多的人们共同享用。实物转赠之后,就不再属于原主;而双方交流信息,两者都有得无失。这一特性通常以信息的多方位传递来实现。

(8) 不灭性。信息不像物体和能量,物质是不灭的,能量也是不灭的,其形式可以转化,但信息的不灭性同它们不一样。一个杯子被打碎了,构成杯子的原子、分子没有变,但已不成为一个杯子。又如能量,我们可以把电能变成热能,但变成热能后电能已经没有了。而信息的不灭性是一条信息产生后,其载体可以变换、可以被毁掉(如一本书、一张光盘),但信息本身并没有被消灭。

1.1.2 信息、知识、情报与文献

1. 知识

人类习惯于把在利用自然、改造自然的社会实践中所搜集和利用的信息加以概括与总结,由此取得新知。他们在实践经验的基础上,发挥主观能动性,对所掌握的信息加以系统化与完备化,就成了有一定模式的知识。因此,知识是人类对自然界、人类社会以及思维方式与运动规律的认识,是人的大脑通过思维重新组合的信息的集合,是系统化完备化的信息。可见,知识是信息的一部分。知识是人类改造和利用自然的经验总结,是文明发展的动力,人类利用知识就会产生出改造客观世界的无穷无尽的物质力量。

2. 情报

情报一词是外来语。它来自日语。其一种意义相当于英语的 intelligence,比如美国的中央情报局,英文就叫 Central Intelligence Agency;另一种意义就是指英语的 information。我国 1989 年版《辞海》定义情报为:一是“以侦察手段或其他方法获得有关敌人的军事、政治、经济等各方面的情况,以及对这些情况进行分析研究的成果,是军事行动的重要依据之一”;二是“泛指一切最新的情况报道”。

前苏联情报学家米哈依洛夫认为“情报是作为存储/传递和传播对象的知识”。这种定义更加接近我国情报工作的实践。因此我们认为“经过加工分析、提炼并加以传递的知识,才是情报”。

另外,需要注意的是,只有参与交流或传递的知识才能成为情报。情报具有特定的知识内容,且是以运动形式存在的,是在特定条件下对特定目的有用的那一部分知识。知识性、传递性和效用性是情报的基本属性。情报的知识性是指情报的本质是知识。情报的传递性是指无论多么重要的知识,人们不知道其存在就



不能成为情报。知识要变成情报，还必须经过运动。情报的效用性是指只有那些能满足特定需要的知识才可称之为情报。

3. 文献

我国 1983 年颁布的国家标准《文献著录总则》中，将文献定义为“记录有知识的一切载体”。文献由知识内容、信息符号和载体材料三个不可分割的基本要素构成。知识必须依附于一定的载体，如存在于人的大脑中或固化在一定的物质载体上。人类在漫长的社会生活、生产劳动和科学实践中积累了大量的知识，要借助于一定的载体（如泥块、甲骨、竹简、绢帛、纸张、胶片、磁带、磁盘、光盘、网络等）进行记录与传播。这种以文字、图形、符号、声频、视频、电子等手段记录和传播知识的载体就是文献。

从信息、知识、情报与文献的定义可以看出，信息、知识、情报是一个范畴，它们直接可用于消除事物的不确定性，是一个内部概念。三者之间存在着大小与从属关系，即信息的概念领域最广，知识次之，情报最狭，知识与情报最后的落脚点都是信息。英国著名情报学家布鲁克斯（B. C. Brooks）用知识方程形象地说明了知识与情报之间的关系， $K(S) + I = K(S + S_1)$ ，即“情报是使原有的知识结构发生变化的那一小部分知识”。文献则属于另外一个范畴的概念，它与事情不确定性没有任何关系，是一个外部概念。

1.1.3 信息检索

1. 何谓信息检索

信息检索（Information Retrieval）是指迅速、准确、无重大遗漏地从浩如烟海的信息集合中查找出自己所需情报的过程。广义的信息检索包括信息存储，两者又往往合并称为“信息存储与检索”（Information Storage and Retrieval）。

根据所要解决的问题不同，信息检索分为三种类型即专题检索、数据检索和事实检索。

专题检索是指根据某一特定课题查找一批各类型的出版物，从而获取所需情报的过程。专题检索主要通过查找文献线索来实现。因此，凡属查找某一主题、某一时代、某一单位、某一著作的有关文献及其出处的，均属专题检索的范畴，例如：要查找有关“碳纤维复合材料的机械性能”有什么文献，这些文献发表在什么地方。

数据检索（Data Retrieval）是以数据为对象的检索，因其是以数值或图表形式表示的数据为检索对象，又称“数值检索”。例如：要查某数学公式，某电子元器件的参数，某材料的成分、性能等。

事实检索（Fact Retrieval）是以从信息中抽取的事项为检索内容的信息检



索, 又称“事项检索”。其检索对象既包括事实、概念、思想、知识等非数值信息, 也包括一些数据信息, 但要针对查询要求, 由检索系统进行分析、推理后, 再输出最终结果。

这三种针对性有异的检索模式中, 专题检索与科研工作联系得最为紧密, 是信息检索工作的重点。很多检索教材将专题检索称为文献检索, 我们放弃了这种说法, 是因为该分类在概念的统一性上存在较大的问题, 同时也以这个概念来表明我们的观点, 那就是信息检索的真实目的是为了查找情报, 而非仅仅是载体。

信息检索是信息处理工作的一个非常重要的环节。图 1-1 所示的是情报获取的过程。

信息检索介于信源与信宿之间, 信息检索水平的高低直接决定信源与信宿之间关系是否得以良好契合, 直接决定情报需要者的情报满足程度。

专题检索按照所利用的检索工具的不同可以分为传统检索和电子检索。传统检索主要通过索引、目录和文摘等纸质检索工具来开展, 主要靠手工操作的方法, 对书刊资料进行查

检获取情报的线索, 所以又称手工检索, 是图书馆最早采用的检索方法。这种检索全文文献的获取需要异地进行, 我国 1978 年以前的资料, 除少部分专题制成光盘, 其他资料还得用这种方法来查找。这种检索工具记录的著录格式一般如下:

101107 Making money from the information industry Databases of large companies were investigated to examine growth rate and the future prospects for different segments of the information... in English

Brown, D, J. *Inst. Britisch Telecom. Engineer.* V 1 n 1 January/March 2000 p45-51

电子检索一般是指通过电子数据库所进行的检索。这类数据库包括文摘性数据资源及全文数据资源, 题录性数据资源非常少见 (很多所谓的题录数据库都是文摘型的, 如中国学术期刊题录数据库)。电子检索按照存储电子数据介质的不同可以分为光盘检索、网络数据库和镜像站点检索、联机检索和网络检索。

光盘检索是利用计算机技术, 把检索软件及文献刻录在光盘上, 由计算机来检索。随着镜像站点服务技术的进一步提高, 光盘检索有萎缩的趋势。网络数据库检索即是通过互联网来检索出版商、集成商等的专业数据库。这种数据库的数

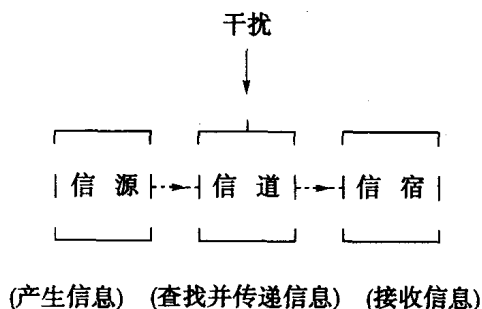


图 1-1 信息的传递与获取的渠道