



计算机



献检索

主编：张怀涛

沈阳出版社

东方圣明图书

计算机文献检索

主 编:张怀涛

副主编:(以姓氏笔画为序)

王云生	王 军	王歌雯
范文兰	岳修志	胡运清
赵 颖	程伟杰	谢卫国

沈阳出版社

图书在版编目(CIP)数据

计算机文献检索/张怀涛主编. —沈阳:沈阳出版社,

2007.5

(图书馆建设研究/张怀涛主编)

ISBN 978 - 7 - 5441 - 3332 - 6

I. 计… II. 张… III. 计算机应用—情报检索 IV.

G252.7

中国版本图书馆 CIP 数据核字(2007)第 056718 号

出 版 者:沈阳出版社

(地址:沈阳市沈河区南翰林路 10 号 邮编:110011)

印 刷 者:河南中医学院印刷厂

幅面尺寸:850mm × 1168mm 1/16

印 张:17.5

字 数:690 千字

印 数:1 - 1000

出版时间:2007 年 5 月第 1 版

印刷时间:2007 年 5 月第 1 次印刷

责任编辑:萧大勇

封面设计:东方圣明

版式设计:东方圣明

责任校对:钟 仁

责任监印:杨 旭

ISBN 978 - 7 - 5441 - 3332 - 6

定 价:48.00 元

联系电话:024 - 62564958

邮购热线:024 - 62564935

E - mail: sysfax_cn@ sina. com

目 录

第一章 计算机信息检索原理	(1)
第一节 信息资源	(1)
第二节 信息检索	(7)
第三节 计算机检索的含义	(9)
第四节 计算机检索的类型	(10)
第五节 计算机信息检索的趋势	(13)
第二章 计算机信息检索系统	(18)
第一节 计算机信息检索系统的构成	(18)
第二节 计算机网络	(21)
第三节 数据库	(29)
第三章 计算机信息检索方略	(34)
第一节 计算机信息检索语言	(34)
第二节 计算机信息检索步骤	(39)
第三节 计算机信息检索策略	(45)
第四节 计算机信息检索方法	(48)
第四章 网络信息检索工具	(53)
第一节 网络信息检索工具的类型	(53)
第二节 网络目录	(55)
第三节 搜索引擎	(56)
第四节 网络信息检索工具选介	(58)
第五节 网络信息检索工具的质量	(67)
第五章 网络特殊形式的信息检索	(69)
第一节 FTP 信息检索	(69)
第二节 Telnet(远程登录)信息检索	(72)
第三节 Usenet 信息检索	(73)
第四节 Mailing List(邮件列表)检索工具	(74)
第五节 Gopher 信息检索	(76)
第六节 WAIS 信息检索	(80)
第七节 Archie 信息检索	(85)

第六章	综合性计算机检索系统	(90)
第一节	中国知网	(90)
第二节	中国高等教育文献保障系统	(95)
第三节	万方数据资源系统	(97)
第四节	国家科技图书文献中心服务系统	(103)
第五节	Dialog 系统	(107)
第六节	STN 系统	(109)
第七节	OCLC 系统	(109)
第八节	工程索引	(110)
第九节	科学引文索引	(120)
第七章	专业性计算机检索系统	(124)
第一节	新华社多媒体数据库	(124)
第二节	法律信息网	(126)
第三节	经济信息网	(131)
第四节	科技信息网	(136)
第五节	美国《教育资源信息中心》	(144)
第六节	美国《生物学文摘》	(145)
第七节	英国《科学文摘》	(145)
第八节	《化学文摘》	(151)
第八章	图书文献检索	(156)
第一节	读书网站	(156)
第二节	网上书店	(158)
第三节	中国数字图书馆	(159)
第四节	超星数字图书馆	(168)
第五节	书生之家电子书	(171)
第六节	中国年鉴数据库	(173)
第七节	Apabi 电子书	(176)
第八节	美国国会图书馆书目检索	(179)
第九节	联合国图书馆书目检索	(181)
第九章	期刊文献检索	(184)
第一节	全国报刊索引数据库	(184)
第二节	中文科技期刊全文数据库	(186)
第三节	中国科学引文数据库	(188)
第四节	人大复印资料数据库	(189)
第五节	中国社会科学引文数据库	(190)
第六节	ASTP 期刊全文数据库	(192)
第七节	Springer Link 数据库	(194)
第八节	Elsevier 数据库	(199)
第九节	EBSCO HOST 数据库	(204)

第十章 特种文献检索	(211)
第一节 专利文献的检索	(211)
第二节 标准文献的检索	(216)
第三节 会议文献的检索	(220)
第四节 学位论文的检索	(223)
第五节 科技报告的检索	(225)
第六节 政府出版物的检索	(227)
第七节 档案信息的检索	(230)
第八节 产品信息的检索	(234)
第十一章 实用信息的检索	(238)
第一节 人物信息的检索	(238)
第二节 机构信息的检索	(239)
第三节 地理信息的检索	(240)
第四节 统计信息的检索	(242)
第五节 教育信息的检索	(243)
第六节 就业信息的检索	(245)
第七节 医疗信息的检索	(249)
第八节 音像信息的检索	(250)
第十二章 计算机检索障碍与效率	(252)
第一节 信息用户需求	(252)
第二节 个人数字图书馆建设	(257)
第三节 计算机信息检索障碍	(265)
第四节 网络信息资源评价	(269)
第五节 计算机信息检索效率	(271)
后 记	(275)

第一章 计算机信息检索原理

第一节 信息资源

一、信息资源的概念

信息是一种重要的资源,这已成为人们的共识。对信息资源进行组织、管理、建设、开发、利用是人们普遍关心的问题,对信息资源管理的研究也成为国内外研究的热点之一,并已发展成为一门新兴学科。

“信息资源”的概念最早是由国外学者提出来的,时间大约在20世纪60年代末、70年代初。我国学者自20世纪80年代开始研究信息资源的概念问题,至今有多种表述。若对国内外的这些定义进行归并整理,去同存异,则目前关于信息资源的解释可概括为下列观点:

(1)信息资源=信息源。这种观点将“信息资源”与“信息源”这两个根本不同的概念混为一谈,显然是不正确的。信息源是指信息的出处或来源。信息的来源多种多样,可来自个人、单位、机构、团体,也可来源于文献、实物,还可来源于市场、会议等场所。信息源中有些可称其为信息资源,如文献,但大多数不能称其为信息资源。

(2)信息资源=文献资源。这种观点缩小了信息资源的外延。因为文献资源只是信息资源中的一部分,而非全部的信息资源。

(3)信息资源=数据。数据有广义和狭义之分。广义的数据多用于计算机和通信领域,指形成信息的原材料;狭义的数据则是与文字、图形等相对的一种高度浓缩的信息。将信息资源等同于广义的数据显然扩大了其外延;而将信息资源等同于狭义的数据则又缩小了其外延。

(4)信息资源=信息。信息有有效、无效、负效之分。作为资源必须有用,有用性是资源的基本属性。从这个意义上说,那些无效的甚至有害的信息是不能作为信息资源的,将信息资源等同于全部信息,同样扩大了信息资源的外延。

(5)信息资源=支持工具(人员、技术、设施等)。信息资源的核心是信息。信息人员是收集、开发、传播信息的主体。信息技术和信息设施则是获取、传播、开发、利用信息的方法、手段、工具、条件。舍弃信息而将信息支持工具作为信息资源无疑是一种本末倒置。

(6)信息资源=信息+支持工具。将支持工具列入信息资源的范畴,悖于常理,显得牵强附会。因为人们在说××资源时是不将其开发人员、开发条件包括在内的。如一个国家的石油资源,往往是指该国家的石油储量,并不包括其石油工人、开采技术和开采设施。可见,将支持工具包括在信息资源的范畴中同样扩大了信息资源的外延。

(7)信息资源=部分信息的集合。这种观点比较有道理,但关键在于是哪部分信息的集合。美国一些学者认为“是一系列已经制度化了的、为一个或多个用户集团反复使用的信息”,而我国一些学者则认为是“可以利用的信息的集合”。应该说,这些表述已接近信息资源的本质,但还需要进一步地推敲和修正。

从上述观点中可以看出:

(1)信息资源是信息的集合。俗话说:“独木难成林”。一棵树构不成森林资源,一滴水构不成水资源,一滴原油也构不成石油资源。同样,一条信息或几条信息也构不成信息资源。只有当信息达到一定丰度和凝聚度时,才能成为信息资源。从这个意义上说,信息资源应是多种多样信息的总和或集合。

(2)信息资源是经过人类选择、获取的有用信息的集合。信息资源并不是所有信息的简单相加和堆砌,而是经过人类选择的、对人类有用或能满足人类需求的那部分信息的总和或集合。有用性是一切资源的本质属性,信息资源也不能例外。从信息海洋中挑选出有用信息,并将之与无用信息区分开来,正是信息管理人员的基本任务之一。

(3)信息资源是经过人类组织序化的信息的集合。与非信息资源相比,信息资源最显著的特征就是有序性。对水资源、石油资源、矿产资源等自然资源来说,无所谓有序、无序,只要具备一定的丰度和凝聚度,值得人们开采、获取即可。信息资源却不然,无序的信息不仅无法利用,还会造成信息通道的“栓塞”,阻碍信息的传播、交流、开发和利用。因此,组织、序化的信息才能成为信息资源,而没有控制的、未经组织的信息将不再成为一种资源。

我们认为,信息资源是经过人类选取、组织、序化的有用信息的集合。

二、信息资源的属性

信息资源的属性是什么?很多专家进行过探讨。由于对信息资源内涵与外延的不同认识,必然导致对其属性认识的差异。我们认为信息资源具有如下几个属性:

(1)社会性。信息普遍存在于自然界和人类社会,而信息资源则是人类社会的产物,它通过特定的符号表述和传递,可以在特定的社会条件下广泛地往复地为人类所利用。信息资源不仅是社会生产力的反映,而且任何一类信息资源,都几乎不是孤零零存在的,而是与诸多的社会条件密切联系。在现代社会,要求人们不仅要注重自然科学信息资源的开发与利用,而且要注重社会科学、人文科学信息资源的开发与利用,善于在各类信息资源的相互影响和渗透中发现、挖掘信息资源的巨大社会价值。

(2)人工性。信息作为物质或事物运行的状态与方式,无论人类是否感知它,它都自然客观存在的。信息的资源化离不开人类的参与。信息资源的生产、形成乃至组织、建设、开发、利用,无不打上人类加工的烙印。信息资源的人工性特点正是我们建设、开发、利用信息资源的理论依据。

(3)稀缺性。信息是无限的,有物质的地方就一定有信息存在。只要物质不灭,信息就会像物质一样永恒地存在。物质取之不尽,信息就用之不竭。信息资源仅是信息中的一部分,是经过人类选择的有用的那部分信息,是有限的。信息犹如无边无际、无始无终的宇宙,是无限的,信息资源好似太阳或地球,是有限的。这是就其存在来说的。另一方面,从人类对信息资源的需求来看,是无限的,与这种无限的需求相比,人类拥有的信息资源则永远是有限的。

(4)有序性。信息浩如烟海,且杂乱无章,处于一种混沌无序状态,面对浩瀚无边的信息海洋,人们常常发出无可奈何的感叹和“信息爆炸”、“信息污染”、“信息垃圾”的惊呼。大量无序的信息,常常造成信息通道的“拥塞”,使信息的传递发生迟滞性干扰,人类也无法利用。而信息资源则是人类按照一定次序组织起来的信息,具有序列性。就既定的信息资源而言,它必定是不同内容的信息集合,集合中的每一信息都具有独特的性质。

(5)再生性。信息资源在满足社会需求和利用的同时,不仅不会被消耗掉,还会生产出新的信息资源,而且信息资源利用得越多越广,其效用发挥得就越充分,创造出的新信息就越多。发明家利用专利信息设计出新的产品,学者查考大量书刊资料创作新的专著,均说明信息资源利用的结果是再生新信息。因此应当鼓励消费、利用信息资源。

(6)积累性。信息资源是有用的信息的总和或集合。一条信息构不成信息资源。只有经过一定时间积累使信息达到一定的丰度和凝聚度,才能成为信息资源。正是这种积累性,才使不断流散在空间和时间中的信息,能够汇集到信息机构,跨越时空限制从不同角度、不同方向满足人们特定的信息需求。每一代信息生产者都不是白手起家,他们在继承前人成果的基础上开展自己的工作,他们的产品和前人的成果一起又构成后人生产的基础和条件。

(7)价值性。传统的物质经济活动主要依赖于物质原料、劳动工具、劳动力等物质资源和能源资源的投入,现代的信息经济活动则主要依赖于信息、信息技术、信息劳动力等信息资源的投入。人类之所以把信息资源当做是一种生产要素,主要是因为不仅可以将信息本身看做是一种重要的生产要素,可以全部或者部分取代物质原料等非信息投入要素,而且可以通过与这些非信息要素的相互作用,使之增值。正确而有效地利用信息,可以在同样的条件下创造出更多更好的物质财富,可以开发或节约更多的能量,可以节省更多的时间。

(8)多用途性。信息资源与经济活动相结合,使信息资源具有很强的渗透性,可以广泛渗透到经济活动的方方面面。同一信息资源可以作用于不同的作用对象上,并产生不同的作用效果。经济活动行为可以根据这些不同作用对象所产生的不同作用效果,对信息资源的使用方向进行选择,这样就产生了信息资源的有效配置问题。

(9)不可分性。首先,信息资源在生产过程中是不可分的。信息生产者为某个用户生产一组信息与为许多用户生产同一组信息比较起来,所花费的努力(如费用、难度等)几乎是一样的,如个人计算机应用软件包,当其被开发出来以后,生产一个拷贝与生产多个拷贝在工作量以及费用上的差别是微不足道的。其次,信息资源在使用过程中也是不可分的。信息资源不能像多少吨水泥那样任意计量。即使信息资源在交换中是可分的,某一组信息资源的一部分亦具有市场价值。但对于特定目的而言,如果整个信息资源集合都是必需的,不能任意缺少,则只有整个信息资源集合都付诸使用,其使用价值才能得到最直接的发挥。

(10)非消耗性。物质资源和能源资源是可消耗的,在消费和使用中最终消灭其独立的物体形式和使用价值,因此物质资源和能源资源不会在使用中再生,也不会表现为任何方式的积累。而信息具有共享性,信息资源则具有非消耗性。当物质资源或能源资源一定时,甲乙双方在资源利用上就存在着竞争关系,即若甲多则乙少,反之亦然。物质经济以大量消耗物质、能源为代价,建立在信息资源基础之上的信息经济则完全不同,它使人们的眼光转向了信息。在投入现代经济的劳动、资本、能源、物质和信息等成分中,只有信息资源是不可消耗的,而且是可以再生的,因此在物质经济和物质产品中不断增加信息成分已成为当代经济发展必须考虑的重大课题。

三、信息资源的类型

根据信息资源的概念与特征可知,信息资源是按照特定目的和规则,通过人们一系列采集、加工等劳动和创造过程之后以各种形式储存在特定载体上可供利用的信息集合。信息资源涵盖了传统沿用的文献、情报、知识、数据等概念。从信息表达的内容看,信息资源既包括了科技信息、经济信息、教育信息、军事信息、文化信息等各种社会人文信息资源,也包括空间地理信息、气象信息、动植物物种信息以及人类遗传信息等各种自然信息资源。从信息内容的表达方式看,信息资源具有文字、数值、声音、图像、多媒体等多种存在形式。

(一)按信息资源的表现形式分类

1. 文献型信息资源 文字是人们实现信息交流、通信联系而创造的一种形象符号,文献型信息源是以文字形式存储于各种不同载体上的信息,是目前内容最丰富、使用频率最高的信息源。

2. 数据型信息资源 以数值数据形式存储于各种不同载体上的信息,如统计数据、测量数据、理化数据等。

3. 声像型信息资源 以声音或图像形式出现的信息源,它比文字直观,易于理解。如电影、电视、CA 光盘、VCD 光盘等。

4. 多媒体信息资源 “多媒体”一词译自英文“Multimedia”。媒体(Medium)原有两重含义,一是指存储信息的实体,如磁盘、光盘、磁带、半导体存储器等,中文常译为介质;二是指传递信息的载体,如数字、文字、声音、图形等,多媒体信息源是集声音、文字、图像、数据等多种通信媒介为一体的信息,一般以网络形式或光盘出现。

(二)按信息资源的载体形式分类

可以分为手写型信息资源、印刷型信息资源、缩微型信息资源、声像型信息资源与电子型信息资源。

1. 手写型信息资源 是指在印刷术发明之前以手写记录的文献形式,其中包括古埃及的泥板、羊皮纸、甲骨、卜辞、碑铭(石刻与青铜器上)、竹简缣帛,以及后来的笔记、手稿和会议记录等。

2. 印刷型信息资源 是采用纸质作为载体,通过印刷(包括油印、版印、胶印、铅印)作为记录手段而产生的文献形式。印刷型文献的优点是便于携带、传递与阅览;缺点是卷帙繁多、笨重易污、存储内容密度较低、占据空间范围较大。图书、期刊、报告、会议录、政府文献、学位论文、标准、产品样本、专利、档案构成印刷型文献的主体。目前,印刷型文献已经不利于快速传输信息、高效查阅和高密存储,于是其他载体的文献类型应运而生。

3. 缩微型信息资源 是采用感光材料作为存储知识的物质载体,以缩微照相技术为记录手段而产生的文献形式。缩微型文献的主要优点是体积小、质量轻和信息储量大。此外,缩微型文献具有成本低、保存期长、不易散乱,易于检索、复制、放大,并且能够转换成为其他类型文献的优点。因此,缩微型文献已在图书馆、档案馆得到广泛应用。图书、报纸、研究报告、卡片目录、学位论文、电话本、地图都有缩微件。缩微型文献按照外形可以分为卷片型和平板型(条片、缩微平片、缩微卡片等),需用透射式阅读机阅读;按照其穿透力可以分为透明体和不透明体。缩微印刷品,以不透明感光纸印制而成,用反射式阅读机阅读。普通缩微平片比纸质文献节省空间,每卷长 30.5m 的 16mm、35mm 宽的胶卷可以存储几千页甚至上万页的资料。超缩微平片(Ultrafiche)一般每张可以存储 1000 ~ 3200 页的资料内容。全息照相存储技术的发展促使更高密度的存储载体问世。缩微型文献的主要缺点是必须借助缩微放大机才能阅读,因阅读时眼睛容易疲劳,人们一般不愿意使用。

4. 声像型信息资源 是采用感光材料或磁性材料作为载体,以光学感光或电磁转换为记录手段,利用专门的电子机械装置记录来显示声音、图像而产生的文献形式等。视听型文献完全脱离传统的文字记录形式,直接记录声音和图像,给人以直观真切的感觉,由于可以随时涂抹更新,所以又称直感型资料。与纸质文献相比,视听型文献除了能够如实记录、存储声音和图像信息以外,还能超越时空反复播放,并使信息接受者多通道地摄入信息,有助于理解知识、加深印象,获得长久记忆。视听型文献还可以使信息按照需要展示,例如放慢瞬时动作以供分析,放大肉眼看不见的微观现象等,从而提高教学研究的效果。

5. 电子型信息资源 采用数字、电子等高新技术手段,把信息存储在磁盘、磁带或光盘等载体中,通过计算机对电子格式的信息进行存取、处理、传播。电子型文献内容丰富,类型多,包括电子图书、电子期刊、电子报纸、电子地图等。

(三)按信息资源的出版形式分类

这是一种最常见的分类方法,包括图书、期刊、报纸、档案、标准、图谱、研究报告、会议资料、学位论文、专利说明书、产品说明书、政府出版物等。

1. 图书

狭义的图书,包括印本书和写本书两种基本形态,其历史悠久,流传广泛,数量庞大,使用方便,影响深远,是迄今为止最主要的信息资源。国家标准《情报与文献工作词汇·传统文献》(GB 131434-91)对图书的解释是:一般不少于 49 页并构成一个书目单元的文献。根据联合国教科文组织(UNESCO)和国际标准化组织(ISO)规定:49 页

不包括封面与扉页。如为 48 页或少于 48 页则称为小册子。图书是最古老、最重要的文献类型。科技图书大多是对已发表的科学技术成果、生产技术知识和经验经过著者的选择、鉴别、核对、组织而成的,论述比较系统,全面可靠,查阅方便(有目次表、索引),但出版周期较长,知识的新颖性不够。图书一般属于三次文献,但有的专著往往包含著者的新观点,或使用新的方法、新的材料,往往又具有一次文献的意义。

图书种类较多,包括专著(Monograph)、教科书(Textbook)、词典(Dictionary)、手册(Handbook)、百科全书(Encyclopedia)等各种阅读型图书和参考书。

2. 连续出版物

连续出版物(Serial),是具有统一名称、固定版式、统一开本、连续编号,汇集多位著者的多篇著述,定期或者不定期编辑发行的出版物。通常,连续出版物又可分为:期刊(杂志)、报纸、年度出版物(年鉴、指南等)、报告丛刊、会议录丛刊等类型,其中以期刊和报纸流行最广、影响最大。

期刊又称杂志(Journal, Magazine),一般是指具有固定题名,定期或不定期出版的连续出版物。其特点是出版周期短,报道文献速度快,内容新颖,发行及影响面广,能及时反映科学技术中新成果、新水平、新动向。期刊发表的论文大多数是原始文献,许多新的成果、新的观点、新的方法往往首先在期刊上刊登。科学技术的研究人员应熟悉本专业有关的期刊。常常阅读期刊可以了解动态,掌握进展,开阔思路,吸收新的成果。期刊论文是文献的主要类型,是检索工具报道的主要对象。

根据内容的性质和用途,期刊大致可以分为以下类型。

(1) 学术性期刊 主要是指刊载学术论文、实验报告和学术性报告等学术文献的期刊。

(2) 政论性期刊 是指以报道和评论国际、国内时事和重大事件,发布消息,宣传政策,制造舆论作为主要目的的政治性刊物。

(3) 检索性期刊 是指根据某一特定范围全面搜集有关文献,著录有关文献项目,按照一定方法编排组织,以期刊形式加以报道,以供读者查找原始文献线索的检索刊物。

(4) 评述性期刊 就是反映某一学科或专业领域的科技水平、动态和发展趋势的刊物,这类期刊登载评述性和综述性文章。

(5) 通讯性期刊 又称快报,它是以简要的文字报道最新科研成果的消息及动态的刊物。

(6) 数据性期刊 是以数据为主要内容的刊物。这类期刊只刊登一些实验数据、产品参数、技术规范、法令条例、社会指标和统计资料等。

(7) 行业性期刊 是为本系统、本行业有关厂商推销产品,开拓市场所做的新技术、新产品、新工艺、新设备、新材料等简短报道的刊物。

(8) 通俗性期刊 包括科普性、知识性和趣味性的刊物。

(9) 文艺刊物 指以文学作品为主,附以评论、消息报道等内容的刊物。

报纸,也是一种连续出版物。报纸具有固定名称,是出版周期最短的定期连续出版物,通常以报道新闻为主,也对一些重大问题发表评论。报纸的基本特点是内容新颖、涉及面广、读者最多,是影响面较广的文献信息源。对社会科学特别是对广泛的社会研究和企业经营来说,报纸是非常重要的信息源。

3. 特种文献

特种文献通常是指具有特定内容、特定用途、特定读者、出版形式比较特殊的文献资料。特种文献通常介于图书与期刊之间,似书非书,似刊非刊,其内容广泛新颖,类型复杂多样,涉及科学技术、生产生活的各个领域,出版发行无统一规律,一般单独成册,限制发行,难以获得。特种文献报道迅速、数量庞大、内容新颖、论述专深,多为原始文献,是一种非常重要的情报来源。

通常,特种文献主要包括:科技报告、专利文献、标准文献、会议文献、学位论文、政府出版物、产品资料、其他资料(如档案资料、地图、乐谱等零散文献)等几种类型。

特种文献主要包括以下几种类型。

(1) 政府出版物 政府出版物是各国政府部门及其所属的专门机构发表、出版的或是通过指定出版商出版的文献。政府出版物的主要特点:一是内容广泛——包括基础科学、应用科学和社会科学等各个领域;二是形式多样——常见的形式有报告、公报、通报、通讯、文件汇编、统计资料、图表、官员名录、国家机关指南、工作手册;三是权威可靠——由于政府的国家权威性,因而政府出版物是反映一个国家的方针、政策、经济和科技现状的权威性信息来源,对于了解有关一国方针政策和科技发展状况具有重要的参考价值。

通常,政府出版物可分为两类:一类是行政性文献(包括宪法、司法文献),主要涉及政府政策、法律、国会和议会记录、议案、决议、司法资料、听证记录、调查统计资料等。另一类是科技文献,主要指政府部门出版的科技报告、

标准、专利文献、科技政策文件,公开后的科技档案、经济规划、气象资料等。

(2) 学位论文 学位论文是高等学校、科研机构的学生为获得学位,在导师指导下,通过完成科学研究后撰写的学术论文。有价值的学位论文是在占有大量资料的基础上提交的研究成果和实验创新,具有独创性、新颖性、科学性的特点,内容质量须经学位或学术委员会的把关考核。学位论文通常需要比较详细地总结前人的工作和当前的研究水平,进行选题论证,开展系统的实验研究及理论分析,提出自己的观点。学位论文探讨的问题往往比较专一,有一定的创造性,是一种重要的文献来源。

(3) 科技报告 科技报告是研究人员围绕某一专题从事研究取得成果以后撰写的正式报告,或者是某项研究过程中的阶段性进展的真实记录,及在研究过程中每一个阶段的进展情况的实际记录。其特点是内容详尽专深。报告的类型有技术报告(Technical Reports)、札记(Notes)、论文(Papers)、备忘录(Memorandum)、通报(Bulletin)、可行性报告(Feasibility Report)、市场预测报告(Market Prediction Report)等。报告一般单独成册,有具体的篇名、机构名称和统一的连续编号(即报告号)。

报告一般划分为:保密(Classified)、解密(Declassified)及公开(Unclassified)等几种密级。保密的报告经过一定时间后往往会转为解密报告。非密资料中,又分为非密控制发行和非密公开发行。科技报告出版发行周期较短,报道及时,其主要特点为:内容专深,叙述详尽,数据可靠,推理严谨。许多科研成果首先反映在科技报告中。因此,它能较好地反映一个国家某项研究课题的发展水平。

(4) 专利文献 专利是发明人创造发明了某种新技术,经政府专利局审批后,即获得一定年限的垄断权。专利权可以作为商品买卖。专利文献,主要是指专利说明书,分发明专利和实用新型专利两种。这两种说明书是专利申请人向专利局递送的说明其发明创造的文件。在说明书中,发明人常常论述其发明解决了什么特殊问题,解决的方法,对旧有产品的改进及其他用途等。它同时对企业引进技术和设备,以及保护企业自身利益的技术有着非常重要的作用。因此,专利文献已成为情报的一个重要来源。

专利文献有如下特点。一是内容广泛。凡在工业上直接应用的一切发明、改进、外观设计,乃至别针、纽扣等日常生活用品到尖端技术,以及与此相关的生产工艺、设备、材料、方法等都可以在专利文献中得到反映。二是详尽具体。各国专利法对于专利说明书内容叙述的详尽程度都有明确规定。一般要求叙述详尽具体,以所属技术领域的人员能够利用发明成果为准。这一特点促使有的企业通过了解竞争对手的专利即可大致了解其产品门类的变化和技术改造,从而制定自己的发展对策。三是有法律性,即专利权受到国家法律保护。专利说明书既是技术文件又是法律文件。法律性要求专利说明书用法律语言概括、准确地说明申请专利权的范围和技术细节,这一性质使它成为保护知识产权的主要依据。通常,90%以上的新技术首先发表在专利文献上,所以专利文献可以反映一个国家的先进技术水平。

(5) 标准文献 标准文献是一种规范性的技术文件,它是在生产或科学研究活动中对产品、工程或其他技术项目的质量品种,检验方法及技术要求所做的统一规定,供人们遵守和使用。

技术标准按使用范围可分为:国际标准、区域性标准、国家标准、专业标准和企业标准五大类型。每一种技术标准都有统一的代号和编号。独自构成一个体系。技术标准是生产技术活动中经常利用的一种情报信息源。

标准文献有如下主要特征。一是有约束力,由于经过权威部门有计划、有组织、有步骤地工作及多次修改制定成为生产法规,因而具有法律性质。二是时效性强,所反映的发展水平只能是当时技术所能达到的,随着经济发展和科技水平的不断提高,标准需要进行修订、补充、替代或废止,所以标准反映着一个国家的生产水平和工艺水平。三是针对性,一个标准一般只解决一个问题,不同种类和级别的标准只能在不同的范围内贯彻执行。

标准文献的性质和特征,是任何其他文献所没有的,因而决定其作为生产技术活动依据的实用价值。标准对于企业产品的更新换代、改进工艺水平、提高产品质量、加强市场竞争力可以起很好的借鉴作用。此外,在国际贸易竞争的激烈时代,靠关税壁垒阻止一国商品进入另一国的办法已经行不通了,但以产品不符国际或某国标准而提高市场准入难度是许多国家的常用手法。因此标准文献成为企业了解一个国家经济技术政策、科技和生产水平的重要参考资料,是企业制定出口战略和组织生产活动的依据。

(6) 档案文献 是指国家机关、企事业单位和集团、企业的机构在技术和开发、运行、操作及活动过程中形成的、具有保存价值的历史记录,并且是按一定的原则和方法保管起来的文件材料。档案文献的内容广泛、形式多样、来源庞杂。经过整理后分别成为文书档案、人事档案、会计档案、科研档案、产品档案、工程档案等。还包括有信函、日记、账簿、报告、照片、图样、协议书、备忘录、会议记录、契约、布告、通知、履历表等。

档案具有原始性、稀有性、客观性和可靠性,致使部分档案在一定的时间内是受到保护的,在利用上有特殊的要求和价值。档案的价值首先体现在其凭证作用。由于档案是实践活动保存下来的原始记录,所以成为查考和处理事务的真凭实据。此外,档案的价值还在于可以提供专门信息,例如学术部门利用历史档案研究历史发展中的

有关问题;企业利用技术档案了解技术开发应用、新技术、新工艺、新材料的引进情况;经济学专业人员利用会计档案了解产品生产成本、原材料的价格变化,进而通过归纳、分析,预测企业活动的发展趋势及市场影响。

档案大多由各系统、各单位分散收藏,一般具有保密和内部使用的特点。它是各种社会活动的实录,是真实可靠的历史信息情报,具有很高的参考价值。

(7)会议文献 会议文献是指在国际、国内等各级重要学术会议上发表的论文和报告。此类文献一般都要经过学术机构的严格挑选,代表某学科领域的最新成就,反映该学科领域的最新水平和发展趋势。所以会议文献是了解国际及各国的科技水平、动态及发展趋势的重要情报来源。会议的类型很多,归纳起来可分为国际会议、全国会议、地区性会议三种。会议文献大致可分为会前文献和会后文献两类。会前文献主要指论文预印本(Preprint)和论文摘要;会后文献主要指会议结束后出版的论文汇编——会议录(Proceedings)。它一般是指在学术会议上宣读和交流的学术报告、学术论文和会议交流的书面材料,以及会前会后散发的有关资料。

会议文献大多是论文作者自己创造的科研成果,专业性强、内容新颖、学术性强、出版较快。一些新的发现、新的理论往往首先发表在学术会议上,所以会议文献能够反映一门学科、一个专业的最新发展水平。对于及时了解国内外最新科技发展水平、动态和发展趋势很有帮助,可以说是一个很重要的情报源。

(8)产品样本 又称产品说明书,是国内外生产厂商或经销商为推销产品而印发的企业出版物,用来介绍产品的品种、特点、性能、结构、原理、用途和维修方法、价格等。查阅、分析产品样本,有助于了解产品的水平、现状和发展动向,获得有关设计、制造、使用中所需的数据和方法,对于产品的选购、设计、制造、使用等有着较大的参考价值。

由于产品样本是已经生产的产品说明,在技术上比较成熟,数据比较可靠,对产品的具体结构、使用方法、操作规程、产品规格都有较具体的说明,并常常附有外观照片和结构图。专利产品还注有专利号(根据专利号可查找专利说明书),对于新产品的的设计、试制都有较大的实际参考价值。

通常,产品样本包括产品目录、单项产品样本、产品说明书、企业介绍和广告性厂刊等,内容一般包括产品的性能、规格、构造原理、用途、使用方法和操作规程等。

产品样本内容具体、图文并茂、出版迅速。其特征主要为:第一是可靠性强。大多样本介绍的是已经投产或正在行销的产品,工艺成熟。这与专利文献所介绍的产品多属未定型和未成熟的产品区别很大。就直观信息而言,专利文献中的产品附图只是一种技术原理示意图或产品结构原理示意图,可靠性不及产品样本的外观造型图和内部结构图。第二是技术信息较为完整。作为产品宣传材料,厂家要使客户全面了解产品性能、特征、参数、型号,就要在样本上尽可能地全面公布有关技术情报和商业贸易信息,如研制背景、产品说明、特性、操作、维修、售后服务,甚至还与其他同类产品的加以比较。第三是新颖及时。产品样本大多是在投产之际或进入市场以前印制的,且为最快地送到客户手中往往采用现场散发的形式。由于上述特点,产品样本就成为商贸信息源和竞争情报源。

4. 按信息资源的加工深度分类

按信息资源的加工程度对信息进行划分,信息资源可分为零次信息、一次信息、二次信息、三次信息。

零次文献,是指未向社会公开,只为个人或某个团体所使用的、非正式出版的文献,例如,文章草稿、私人笔记,未经正式发表的名人手迹、会议记录、征求意见稿,甚至包括口头言论。零次文献没有正式出版的原因很多,诸如内容尚不成熟、未经审批、缺乏经费等。零次文献处于动态之中,可以随时补充整理。

一次信息也称原始信息,凡是依据作者本人研究成果而发表的原始作品,不论是何种出版形式,都称一次文献。专著、报纸、期刊、专利文献、标准文献、会议文献均属一次文献。一次文献是人们对自然和社会信息进行首次加工、固化而成的文字记载,数量庞大,是文献信息源的主要部分,由于在内容上是分散的、无系统的,因此不便于管理和传播。

二次信息也称检索工具,是指对一次信息加工、整理后形成的各种检索工具,如目录、题录、文摘等,它不对一次信息提供评论,仅提供一次信息的检索线索,也就是说,二次信息把一次信息从分散、无序状态变成集中、有序化的状态,供人们迅速、准确地查询所需信息。为了克服一次文献的分散性、无序性,必须对其进行加工、整理、提炼、压缩,并按其外部特征(题名、作者、文献物理特征)和内容特征(学科内容或专题内容)进行组织,使之成为集中有序的二次文献。二次文献是一次文献内外特征的有序汇集,以便查找一次文献或了解一次文献的主要内容。

三次信息,是对一、二次信息按知识门类或专题进行综合加工的产物,是围绕某一课题利用二次文献全面搜集一次文献,采用情报研究方法,对文献内容分析综合、组织编写而成的。如综述、述评、专题报告、专题汇编、数据手册、百科全书、年鉴等。三次文献具有系统性、综合性、知识性和概括性的特点,它从一次文献中汲取重要内容供给人们,便于人们高效率地了解某一领域的动态、发展趋势和有关情况。

第二节 信息检索

一、信息检索内涵

信息利用的关键在于信息获取,而信息获取的核心则是信息检索。信息用户需求的具体信息,大多埋藏在浩如烟海的信息资源之中,这就需要采用一定的工具和方法,通过有效的途径和步骤,把它们检索出来。

信息检索,是指从有序的信息资源中查寻所需信息的过程。广义的信息检索包括两部分,一是存贮,即把大量杂乱无序的信息加以科学地组织;二是检索,即从已经组织好的信息集合体中查检和索取出所需的信息。狭义的信息检索仅指广义中的检索部分。

信息的存贮过程主要包括从信息资源中选择一定范围和水平的信息,对被选定的信息的内容主题及其它有检索意义的特征进行选择、分析、描述并赋予检索标识,进而对这些信息的内容及物质形态的特征进行记录,以形成款目,再对款目进行有序地排列,使之成为一个能随时存取的检索系统。

信息的检索过程则是存贮的逆过程,就是信息用户将自己的需求加以明确化,形成检索提问,对检索提问进行概念分析并用检索系统所用词表中的词加以表达,从而构成包含有检索词以及各检索词之间逻辑关系的检索策略,然后有序地在检索系统中查寻同检索词相一致的检索标识,并要求其各检索词的逻辑组合符合检索式所指条件。这里,检索过程实质上是信息用户的检索提问与检索系统中的检索标识相匹配而决定取舍的过程。

在信息的检索过程中,检索的目的是包含在文献信息资源或网络信息资源中的信息,检索的对象是包含着信息的文献信息资源或网络信息资源。

二、信息检索类型

由于用户的信息需求多种多样,信息检索技术也在不断发展变化,进而产生了多种类型的信息检索。

(一)按检索对象划分

按检索对象划分,即按信息检索结果的内容划分,信息检索分为文献信息检索、数据、事实检索。

1. 文献信息检索 文献信息检索(Document Information Retrieval)是以获得各种类型文献信息为目的,包括文献信息线索检索和文献信息全文检索。文献信息线索检索是指利用书目、索引、文摘等检索系统和工具进行文献信息查找,其结果只是获得文献信息的线索,要获得相关文献信息还需要进一步进行信息检索;文献信息全文检索是以搜索文献全文为目的的检索,其结果是获得全文信息。

2. 数据检索 数据检索(Data Information Retrieval)是指从各种数值数据库和统计数据库储存的数据中查找用户所需的数据信息,又称“数值检索”。其检索结果包括各种参数、调查数据、统计数据、特性数据等,也包括各种图表、图谱、市场行情、化学分子式、物质的各种特性等非数字数据,并提供一定的运算推导能力。例如:查找某一企业的年销售额、某一国家的人口数、物质的属性数据等。

3. 事实检索 事实检索(Fact Information Retrieval)是以事实作为检索对象,针对特定的事件或事实的检索,包括检索事物的性质、定义、原理及发生的时间、地点和因果关系等信息,又称为“事项检索”。例如:查找某一人物的生平事迹;查找某一企业的名称、地址、业务经营范围等信息。

文献查找的观点认为,信息依附于文献而存在,用户通过获得文献而得到所需的信息,信息检索是查找出含有用户所需信息的文献的过程。由于用户所需的数据或事实不能脱离其载体(文献)而单独存在。因此,数据检索或事实检索是以文献为依托的,也就是说,数据检索或事实检索可以被视为文献信息检索的特例。它们与文献信息检索的差异主要在于:对于数据或事实检索,检索人员只需把有关的数据或事实提供给用户,就能直接满足用户的信息需求;而对于文献信息检索,检索人员通常是把查找出来的相关文献提供给用户,由用户阅读、理解、吸收和利用文献的信息内容,从而满足其信息需求。

(二)按信息检索的方式划分

按照检索的操作方式,信息检索分为手工检索和计算机机器检索。

1. 手工检索 手工检索(Hand Retrieval)以手工操作的方式,利用检索工具书进行信息检索。手工检索是信息检索的传统方式,主要依靠检索者手翻、眼看、大脑判断进行,是检索者与检索工具直接对话,其优点是便于控制检索的准确性,缺点是检索速度慢、漏检现象比较严重、工作量较大。

2. 机器检索 (Machine Retrieval)以机械、机电或电子化的方式,利用检索系统进行信息检索。机器信息检索是从20世纪40年代以后逐渐发展起来。电子计算机诞生之后,计算机检索(Computer Retrieval)以其强大的储存

能力、不断提高的处理性能以及同步降低的价格,很快便成为机器信息检索的主流和代表。因此机器信息检索主要就是指计算机信息检索。目前,计算机检索迅速发展,已经具有全文信息检索、自动信息文摘、自动信息分类等功能,其优点是检索速度快、能够多元化检索、检索的全面性较高;相对手工信息检索而言,其缺点主要是需要借助相应的设备进行检索。

(三)按检索要求划分

按照用户对检索的要求,信息检索分为强相关检索和弱相关检索。

1. 强相关检索 强调检索的准确性,向用户提供高度对口信息的检索,也称为特性检索。这种检索注重查准,只要检索得到的文献信息能够满足用户的需求即可,通常对于检索结果的数量多少不做要求。

2. 弱相关检索 强调检索的全面性,向用户提供系统完整信息的检索,也称为族性检索。这种检索注重查全,要求检索出一段时间期限内有关特定主题的所有信息,为了尽可能避免漏检相关信息,一般对于检索的准确性要求较低。

一般来说,用户的检索要求介于两者之间,既要求查找对口的信息,又希望得到所有的信息。用户的检索要求可分为两种情况:一是用户对查准率和查全率的要求都较低,要完成这样的简单检索(Brief Search)比较容易;二是用户不仅对查准率和查全率都有要求,而且要求都较高,要完成这样的复杂检索(Complex Search)相当困难。因为在信息检索中,查准和查全常常不能兼顾。因此,用户为了获得比较满意的检索结果,在掌握检索语言的基础上常常需要把特性检索与族性检索相结合,交替运用,以获得满意的查准率和查全率。

(四)按检索性质划分

按照检索的运行性质,信息检索分为定题检索和回溯检索。

1. 定题检索 查找有关特定主题最新信息的检索,又称为SDI检索。其特点是只检索最新的信息,时间跨度小。定题检索需要对检索主题进行随时跟踪检索,即在文献信息库更新或加入新的文献信息,就运行一次定题检索,从而查找出特定主题的最新信息。这种检索非常适合于信息跟踪,以便及时了解有关主题领域的最新发展动态。如果用户一旦向检索服务机构订购定题检索,一般就会在较长时间内多次运行,由检索服务机构持续地向用户提供最新信息。

2. 回溯检索 查找过去特定时期内有关主题信息的检索,也叫追溯检索。其特点是既可以查找过去某一特定时间的特定主题信息,也可以查找最近的特定主题信息。与每个定题检索需要多次检索有所不同,回溯检索一般只从已有的文献信息库中检索一次,查找出特定主题的信息并提供给用户就算完成检索任务。

目前,用户利用最多的是回溯检索,大多数的检索课题都属于回溯检索。同时,随着现代化信息检索技术的迅猛发展,计算机定题检索也发展得很快,应用越来越多并深受用户欢迎。

(五)按检索的信息形式划分

按照检索的信息形式即按信息检索技术划分,信息检索分为文本检索和多媒体检索。

1. 文本检索 传统文本检索已发展了几十年,它是以前文本特别是二次文献为检索信息源的。查找含有特定信息的文本文献检索,其结果是以文本形式反映特定信息的文献。目前,文本检索在信息检索中依然占据着主要地位。

2. 多媒体检索 是指能够支持两种以上媒体的数据库检索,查找含有特定信息的多媒体文献的检索,其结果是以多媒体形式反映特定信息的文献,如文字、图像、声音、动画、影片等,是在网络环境下发展起来的全新检索类型。目前,互联网上有大量的多媒体信息,这些信息的标引、著录和有序化编排等处理方法与传统的文本文献有很大区别,用户需要学习全新的信息检索方式和信息检索理论,才能高效率地检索所需的多媒体信息,这就向信息检索工作者和用户提出了新的挑战。

3. 超文本检索 超文本中正文信息以节点而不是以字符串为信息的基本单元,节点间以链连接。在检索时,节点间的各种链接关系可以动态地选择激发,通过链从一个节点跳到另一个节点,实现联想式检索。其形式有的在网页的文字处有下划线,或以图标方式标志,用户点击这些标志便能进入与此信息相关的下一页。在该页面上通过超文本链接进入下一个网页,超文本起信息导向作用。这样,用户在从一个页面转向另一个页面的过程中就可以获取自己所需要的信息了。网络信息检索是由网络站点、网页浏览器和搜索引擎以及网络支撑组成的检索系统。其中的核心部分是网络浏览器和具有收集、检索功能的搜索引擎。众多站点、网页上的信息是网络信息的基本组成部分。在网络发展初期,浏览器和简单的搜索引擎即可帮助人们检索所需的文献信息。浏览器相当于提供了一个信息总目,方便用户对各个网站进行直接点击、浏览,通过超文本链接,选择自己所需的信息。浏览虽然方法简易、直接,但随机性强,耗时费用较高。因此,更科学的方法是借助搜索引擎。搜索引擎是网络信息的检索工具。搜索引擎使用网络信息资源自动采集机器人程序(也称网络蜘蛛、爬虫软件),动态访问各站点、收集信息、建立索引,并自动生成有关资源的简单描述,存入数据库中供检索。它可以帮助用户快速搜索所需信息及其相关资料,但查准率有待提高。

第三节 计算机检索的含义

一、计算机信息检索基础

(一) 计算机信息检索概论

1. 计算机信息检索的含义

传统的信息检索是通过手工方式对书本式的检索工具进行检索,这种检索过程是采用手工操作配合人脑的判断进行的,因此,这种检索所进行的匹配与选择主要是靠人脑来进行思考、比较和选择的。

现代科学技术的发展使得以缩微品、声像制品、磁盘、光盘等载体形式记录的非纸信息量急速上升。面对信息爆炸,仅靠纸质载体的信息已不能满足科技人员对信息的完整性、及时性的要求,而且传统的手工检索靠“手翻、眼看、大脑判断”的检索方式已难以全面适应当今信息时代的发展。计算机在信息检索领域的应用克服了手工检索的弊端,使信息检索不仅能跨越时空,在短时间内查阅大型数据库,还能快速地对几十年前的文献资料进行回溯检索,而且大多数联机检索或网络检索系统的数据库更新速度非常快,用户通过计算机信息检索可以得到更新的信息。

计算机信息检索是指人们在计算机或计算机检索网络的终端机上,使用特定的检索指令、检索词和检索策略,从计算机检索系统的数据库中检索出需要的信息,继而再由终端设备显示或打印的过程。由此可见,为了实现信息检索,必须事先将大量的原始信息进行加工处理,存储在计算机中待用,所以从广义的角度来讲,计算机信息检索包括信息的存储和检索两个方面。

信息的存储是按照既定的标准和方针,从信息源中选择合适的信息,并对这些信息的内容进行概念分析,用系统规定的系统语言进行标引,形成信息的特征标识,进行整理与排序,构成可供检索的数据库。主要包括文献的自动标引、自动编写文摘、自动生成机读索引文档和数据库的建立。信息检索作为信息存储的逆过程,主要指用户对检索课题加以分析,明确检索范围,弄清主题概念,然后用系统语言来表示主题概念,形成检索标识以及检索策略,输入到计算机中进行查找。查找的过程实际上是一个比较、匹配的过程,检索标识、检索策略只要与数据库中的信息的特征标识及其逻辑组配关系相一致,则属于“检索命中”,即找到了符合要求的信息。检索结果可以按一定格式显示,也可以联机或脱机打印输出。

因此,与手工检索相比,计算机信息检索是针对数据库进行的检索过程,是在人和机器的协同作用下完成的,匹配由机器执行,而人则是整个检索方案的设计者和操纵者。这极大地减轻了信息用户的负担,并且在一系列系统的检索技术、检索策略的保证下能达到较为理想的查全率和查准率,计算机信息检索系统提供的数据库资源丰富,更新速度快,经过严格的加工、整理,质量较高。作为获取信息的一种能力,科技人员必须要掌握计算机信息检索的方法,从而适应社会的要求。

二、计算机信息检索基本原理及检索技术

(一) 计算机信息检索原理

人类的信息检索行为总是随特定的信息需求的产生而开始,并在特定的环境和信息检索系统中完成。信息检索的环境主要指产生需求的环境、信息检索系统运行的环境以及其他制约因素。特定的信息检索系统包括完成检索过程所需要的一定的设施和工具,可以是纸制检索工具或数据库。

人类的信息检索需求千差万别,获取信息的方法也各种各样,但信息检索的基本原理却是相同的。其本质的部分就是对信息集合和需求集合的匹配和选择,如图 1-1 所示。

匹配和选择是一种机制,它负责把需求集合与信息集合进行相似性比较,然后根据一定的标准选出符合需要的信息。要想进行有效的匹配和选择,首先必须对大量的信息进行收集和加工处理,使之从无序到有序,使信息获得特征性描述,让原来隐含的、不易识别的特征显性化。另一方面,对用户提出的信息需求也要做类似的加工处理,即分析需求的内容,提取出主题概念和其他属性,并利用与信息集合相同的标识系统来表示需求中包含的概念和属性。

信息检索过程中要储存大量的数据,要对这些数据进行各种组合,有大量的排序和对比操作。为了提高信息处理的速度、准确性和可靠性,人们将计算机应用到信息检索中。计算机信息检索中检索的本质没有变,但信息的表示方式、存储结构和匹配方法变化了。要用计算机可以识别的代码来表示信息,用便于计算机快速存取的方式

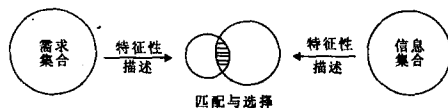


图 1-1 信息检索原理

存储信息,匹配的方法亦由人工比较变为机械匹配,匹配标准由隐式变为显式。在这种机械匹配的过程中,原先表达概念的词语符号变为没有内涵的字符串。检索过程就是字符串匹配和逻辑运算的过程。若两者一致或部分一致,并符合给定的逻辑运算条件即为命中,然后将命中的结果输出给用户。因此,计算机信息检索的实质就是由计算机将输入的检索策略与系统中存储的文献的特征标识及其逻辑组配关系进行类比、匹配的过程。

第四节 计算机检索的类型

一、计算机信息检索的类型

计算机信息检索的类型很多,根据检索系统的工作方式可分为以下几种:

(一)脱机信息检索

这是一种早期应用的系统,是利用单台计算机的输入输出装置进行检索的系统,用磁带作为存储介质,并且为连续的顺序检索方式,适合大批量的定题信息检索,因此,也称成批检索或定题服务。

在脱机检索中,必须事先把众多用户的各种信息要求编成“用户提问档”格式,以机读形式存储在磁带上,定期地(一个月或半个月)检索数据库新增加的内容,然后把命中的文献信息发给用户。

脱机检索由专职检索员统一处理,而且不提供任何实际浏览的可能性,用户不接触机器,所以它是一项“机会检索”。这就要求检索者必须制定完备的检索策略,才能保证较好的检索效果。

(二)联机信息检索

由于计算机软、硬件技术的不断提高,出现了一台主机带多个终端的联机信息检索系统。这种系统具有分时的操作能力,能够使许多相互独立的终端同时进行检索,计算机对用户的提问能及时处理、即刻回答。用户还可以浏览有关信息,随时修改提问,直至得到满意的结果。随着现代通信技术的发展,利用公用通信网或专用通信网,联机信息检索已经超出一个地区、一个国家的范围,进入国际信息空间,出现了像 DIALOG、ORBIT 这样的国际联机系统。

(三)光盘信息检索

光盘是一种用激光记录和读取信息的盘片,具有容量大、保存时间长、成本低等优点。光盘检索系统中目前使用最为成熟的是只读光盘(CD-ROM),数据库的检索软件及数据库存储在盘片上,利用计算机和光盘驱动器进行检索,现在已出现了光盘检索网络。

(四)网络化信息检索

网络化信息检索同联机检索的主机和用户终端的主从关系不同,客户和服务器是对等关系。Internet 就是该系统的典型,Internet 上的主机既可以作为用户访问别的主机里的信息,又可以作为信息源被其他终端访问。

二、联机信息检索

联机信息检索一般是指信息用户利用终端设备,通过通信网络与世界各地的信息检索系统联机,进行人-机对话,从检索系统的数据库中查找出用户所需信息的过程。

联机检索起源于 20 世纪 60 年代的美国。目前,联机检索业已形成了覆盖全球的信息检索系统,如 DIALOG、OCLC 等。我国从 20 世纪 80 年代开始从事国际联机检索,经过 20 余年的发展已建立起了自己的联机信息检索系统,如 ISTIC、MEIRS 等。

联机信息检索系统的优点是:

- 检索速度快;
- 检索范围广而全面;
- 检索途径多、质量高;
- 检索内容新、实时性强;
- 检索辅助功能完善、使用方便,检索结果输出方式灵活、实用。正因为如此,联机信息检索目前仍是计算机信息检索的重要方式之一。但是,联机信息检索也存在自身的缺陷:
 - 主机负担重,一旦出现故障,则整个网络都将瘫痪;
 - 信息组织方式以线性为主,不够灵活。

因此,在 Internet 网络信息检索的冲击下,传统联机信息检索业纷纷采取改进措施,将自己的系统安装在 Internet 服务器上,成为 Internet 的一个有机部分。如 DIALOG、American online 等世界著名联机系统都建立了自己的服

务器,使用超文本技术,增加服务项目,改善用户界面。

1. 联机信息检索系统的结构

联机信息检索系统是由检索服务机构、国际通信网络及终端三部分构成。

(1) 检索服务机构

它由中央主机、数据库及其他外部设备组成。中央主机是检索系统的主体部分,其主要功能是进行信息的存储、处理、检索以及整个系统的运行和管理。一般联机系统拥有多台中央主机,可以随时切换,以保证检索的正常进行。数据库是存储在磁带或磁盘上的文献或数据记录的集合,是联机检索的对象,它相当于手工检索的检索工具书,其实有些数据库就是手工检索工具的机读版。因此,掌握各种数据库的性能是联机检索的前提。其它外部设备主要有主机操作台、高速打印机等设备,以解答用户在检索过程中提出的各种问题,处理用户脱机打印要求。

(2) 国际通信网络

它由通信线路、调制解调器、自动呼叫应答器、多路复用器、通信控制器组成。

(3) 终端

目前采用 IBM PC 系列或兼容机作为终端,利用计算机的键盘输入检索指令,也可预先存入计算机,待接入联机系统后由计算机直接发送,系统对检索指令的响应显示到计算机的显示屏上。

2. 联机信息检索的技术原理

联机信息检索系统是一个典型的计算机信息系统,能完成数据收集、分析、加工处理、存储、传递通信和检索的全过程。在信息存储的过程中,由系统按一定的规律对信息进行加工处理,并赋予特征标识;在信息检索的过程中,由用户通过系统提供的检索指令,向系统提交含有需求特征的检索表达式。计算机信息检索系统接收到正确的指令后,自动地将相关信息集合的特征标识与用户提交的检索特征进行“匹配”。这种匹配完全是一种字符串的类比运算。匹配结束,系统自动给出存储信息的特征与检索提问的特征相符的记录篇数,即命中数量。用户通过显示命中记录的内容,判断检索是否成功,这就是联机信息检索技术的基本原理。

3. 联机信息检索的服务方式

联机信息检索的服务方式主要有以下几种:

(1) 定题信息提供(SDI, Selective Dissemination of Information)

这种服务是由检索系统工作人员将用户信息需求转换成一定的检索提问式,并将此提问式存入计算机中,信息检索系统定期从新的文献信息中为用户检索,为用户加以编排和打印。利用 SDI 服务,用户可定期获得所需要的最新信息,及时掌握同类专题的动态和进展。

(2) 专题回溯检索(RS, Retrospective Search)

这是用户对检索系统中累积多年文献资料的数据库进行检索,查找特定时间范围以内或特定时间以前的文献,通常采用联机检索方式进行。RS 服务的结果一般要求切题,但又无大的遗漏,尽量做到省时、省费用。通过 RS 进行专题查询或情报调研时,可全面系统地了解有关文献的线索。

(3) 联机订购原文

联机检索的结果通常是一些文摘或题录形式的二次文献形式。用户通过阅读这些二次文献了解大致的内容,然后根据这些文献线索查找全文。

(4) 电子邮件(E-mail, Electronic Mail)

联机系统开展此项业务,以满足用户与系统之间,用户与各机构之间,用户与用户之间发送、接收、存储各种信息的需要。每个系统用户都拥有一个 E-mail 地址,联网系统有参加电子邮政的名单。输入接受者的 E-mail 地址和通信内容,接受者在数秒钟内便可接收到本需耗时几天的信件。

三、光盘信息检索

光盘是继纸张、缩微胶片、磁存储器之后的一种用激光束记录和再现信息的存储载体。用于检索和阅读的光盘通常为只读光盘(CD-ROM, Compact Disc Read Only Memory)。它是一种信息载体,而要对其中的信息进行检索和利用则需要计算机的配合。光盘产品自 20 世纪 70 年代出现以来,最初只用于娱乐,直到 1985 年人们才研制出第一种专用于信息服务的光盘。自此,以光盘为载体的数据库产品层出不穷,为信息产业的发展注入了新的生命力,特别是光盘与计算机的结合,使得信息检索模式发生了革命性的变化。

(1) 光盘信息检索特点

· 使用光盘检索系统,可免除联机检索系统所必须使用的电信设备,节省了电信费和联机系统使用费,还可免除由通信线路传输过程中所造成的失误;