

第1章 绪 论

学习目标

1. 了解信息的特征以及信息检索的概念、目的和意义。
2. 理解检索工具的结构及目录、索引、文摘 3 种主要检索工具的特点及作用。
3. 熟悉信息检索的原理、类型和基本工具的使用。
4. 了解信息检索的基本途径，掌握信息检索的方法和步骤。

随着信息时代的到来以及全球市场的形成，信息的瞬息万变给人类社会带来了前所未有的机遇与挑战。在这样多变的环境下，利用具有潜在价值的信息资源，及时、准确地获取和传播有价值的信息，是人类社会存在和发展的前提条件之一。

1.1 信息与文献基本知识

未来学家托夫勒把人类社会的发展过程划分为 3 个阶段：以农业经济为基础的农业社会；以工业经济为基础的工业社会；以信息经济为基础的信息社会。在信息社会中，信息资源已被人们公认为是物质、能量资源之后的人类可利用的第三类资源，是人类社会领域的基础性资源。

1.1.1 从信息到文献

在当今的信息社会中，只有将自然现象和社会现象的信息上升为对自然和社会发展规律的认识，这种再生信息才能在社会和人类生活中发挥日益重要的作用。

1. 信息(Information)

美国数学家、控制论的主要奠基人维纳在 1950 年出版的《控制论与社会》中是这样来理解信息的：“人通过感觉器官感知周围世界”，“我们支配环境的命令就是给环境的一种信息”，因此，“信息就是我们在适应外部世界，并把这种适应反作用于外部世界的过程中，同外部世界进行交换的内容的名称”。“接收信息和使用信息的过程，就是我们适应外界环境的偶然性的过程，也是我们在这个环境中有效地生活的过程”。

从哲学的观点看，“信息”是物质的一种普遍的属性，它反映不同物质所具有的不同本质、特征以及运动状态和运动规律。它是用来消除人们对客观物质不确定性认识的东西。“信息”是人们认识事物和获取新知识的唯一方式。只有掌握了事物的信息，人们才能对事物的运动规律进行调控，人们在认识世界和改造世界的过程中才能处于主动地位。

中国国家标准《情报与文献工作词汇——基本术语》(GB/T4894—1985)中信息的定义为:信息是物质存在的一种方式、形态或运动状态,也是事物的一种普遍属性,一般指数据、消息中所包含的意义,可以使消息中所描述事件的不确定性减少。

信息具备以下基本属性。

(1) 客观性。信息的客观性是指信息是事物本质特征和客观规律的表征,是客观存在的,不以人们的主观意识为转移。

(2) 传递性。信息在事物之间的相互联系必定在信息的流动中发生。信息的传递性表现在人与人之间的信息交换,人与计算机之间的信息交换,动物界及植物界之间的信号交换等方面,同时,人类进化过程中的细胞选择、遗传也是信息的传递与交换。

(3) 多态性。信息在不同的领域具有不同的特性或表现形式,如:客观事物中的各种自然属性;人工设备的技术特征;人类社会的各种社会特征;人脑中反映客观事物认识的思想、知识;人类交流信息过程中的声音、文字、图像以及用各种编码形式记录下来的数据、新闻、情报、消息等。各种形式的信息又常常以综合的方式表现事物的特征,所谓“多媒体”正是信息多样性和综合性的集中体现。

(4) 共享性。同一个信息资源可以被不同用户在不同地点、不同时间中同时利用,不需要任何限制条件,信息量也不会损失。信息不会像物质一样因为共享而减少,反而可以因为共享而衍生出更多信息。

(5) 时效性。信息的时效性是指信息的效能依赖于时间,它既表明了信息的时间价值,又表明了信息的经济价值和社会效益。客观事物不断变化而产生的信息具有很强的时效性。客观事物变化越快,信息的时效性就越强。时间的延误,就是扼杀信息的生命——导致信息价值衰减,乃至完全消失。时过境迁的信息是毫无价值的。

(6) 价值相对性。由于人们认识的能力与认知的条件不同,信息的接受者获得的信息及信息量也就不同,从这个意义上讲,信息的价值是有相对性的。

(7) 依附性。信息必须通过一定的符号,依附于某一载体才能被表现。没有载体就没有信息。

(8) 增长性。物资和能源在使用后会有消耗,而信息的使用不但不减少信息的数量,还能产生更多的信息。

2. 知识(Knowledge)

知识是人们在改造世界的实践中所获得的认识和经验的总和,是对客观世界物质形态和运动规律的认识。根据韦伯斯特(Webster)词典 1997 年的定义,知识是通过实践、研究、联系或调查获得的关于事物的事实和状态的认识,是对科学、艺术或技术的理解,是人类获得的关于真理和原理的认识的总和。

从信息的观念看,知识来源于信息,是信息的一部分。简而言之,知识是理性化、优化、系统化了的信息。从反映的内容而言,知识是客观事物的属性与联系的反映,是客观世界在人脑中的相对正确的反映。从反映的活动形式而言,知识有时表现为主体对事物的感性知觉或表象,属于感性知识,有时表现为关于事物的概念或规律,属于理性知识。

知识是人们在实践活动中获得的关于世界的最本质的认识, 是对信息的提炼、比较、挖掘、分析、概括、判断和推论。一般而言, 知识具有共享性、传递性、非损耗性(可以反复使用, 其价值不会减小)及再生性等特点。

3. 情报(Intelligence)

情报是作为交流对象的有用的知识。“情报是判断、意志、决心、行动需要的能指引方向的知识 and 智慧”; “情报是解决问题所需要的知识”; “情报是激活了的知识”。情报的基本属性是知识性、传递性和效用性。

(1) 知识性。一般而言, 情报是一种新的知识。然而称为情报的知识并不排斥时间上的“不新”。这就是说, 凡原先不知道的(不论时间多久), 而现在又迫切需要的知识, 都属于“新”的知识。

(2) 传递性。情报传递的过程包括从情报源进行收集到向用户提供情报的全过程。无论多高深、渊博、新奇的知识, 如果不记录、不传递、不交流, 都不能称为情报。

(3) 效用性(判断的标准)。情报是一种有价值、有效用的知识。情报均须“有用”, 当用户不需要时, 百分之百的知识也不能把它称之为情报。

4. 文献(Literature/Document)

国家标准《文献著录总则》(GB/T3792.1—1983)给出的定义为: “文献是记录有知识的一切载体。”国家标准化组织 ISO 在《文献情报术语国际标准》(ISO/DIS5217)给出的定义是: “为了把人类知识传播开来和继承下去, 人们用文字、图形、符号、音频、视频等手段将其记录下来, 或写在纸上, 或晒在蓝图上, 或摄制在感光片上, 或录到唱片上, 或存储在磁盘上。这种附着在各种载体上的记录统称为文献。”

从定义中不难看出文献包括 4 个要素。

(1) 知识和信息, 即文献的内容。

(2) 物质载体, 如竹筒、纸张、胶片、磁盘等, 它是文献的外在形式。

(3) 记录知识和信息的符号, 如文字、图表、声音、图像等。

(4) 记录的方式或手段, 如书写、印刷、复制、录音、录像等, 即将知识和信息固化到载体上的手段。

文献是记录有知识和信息的一切载体, 或称为固化在某种物质载体上的知识和信息, 当文献中记录的知识传递给用户, 并为用户所利用时, 文献中的知识就转化为情报。因此, 文献是记录、积累、传播和继承知识的最有效手段, 是人类社会活动中获取情报的最基本、最主要的来源, 也是交流传播情报的最基本手段。现在通常将其理解为图书、期刊等各种出版物的总和。

5. 信息、知识、情报与文献

生活中信息无处不在, 无时不有, 它们是知识产生的原料, 这些原料经过人脑接收、处理后, 成为系统化的信息, 知识由此产生。从知识管理的层面上看, 知识是指可直接用于行动的信息, 它使人们可以随时随地做出正确决策。

目前,我国图书情报学界对信息、知识、文献和情报的看法是:知识是对信息加工、吸收、提取、评价的结果,即系统化的信息成为知识,知识记录下来成为文献,文献经传递并加以应用成为情报,情报体现了人运用知识的能力。

1.1.2 信息的分类

信息的分类没有固定的标准,可以根据不同需求从多个角度进行划分。

1. 按信息的内容划分

按信息的内容划分,其可分为科学技术信息和社会科学信息。科学技术信息包括自然科学信息、工程技术信息等;社会科学信息包括政治信息、军事信息、经济信息、技术经济信息、社会生活信息等。

2. 按产生信息的客体的性质来划分

按产生信息的客体的性质来划分,其可分为自然信息(瞬时发生的声、光、热、电,形形色色的天气变化,缓慢的地壳运动,天体演化等)、生物信息(生物为繁衍生存而表现出来的各种形态和行为,如遗传信息、生物体内的信息交流、动物种群内的信息交流)、机器信息(自动控制系统)和社会信息(人类)。社会信息是指人与人之间交流的信息,既包括通过手势、身体、眼神所传达的非语义信息,也包括用语言、文字、图表等语义信息所传达的一切对人类社会运动变化状态的描述。

3. 按信息所依附的载体来划分

按信息所依附的载体来划分,其可分为文献信息、口头信息等。文献信息,就是以文献为载体的信息,以文字、符号、声像信息为编码信息,是经人们筛选、归纳和整理后记录下来的,它与人工符号本身没有必然的联系,但需要通过符号系统实现传递。口头信息是原始的没有形成文字、符号、声像的信息,不如文献信息易于流传。

1.1.3 文献的分类

文献是信息存在的一种重要形式,其优点是易识别、易保存、易传播,使人类精神文明千古流传。文献是信息的主要载体,大多规范化的重要信息都以文献形式出现,所以文献的分类非常重要。

1. 按照文献的加工层次分类

文献是信息的主要载体,根据对文献信息的加工层次可将文献分为零次文献、一次文献、二次文献和三次文献。

(1) 零次文献(Zeroth Literature),指未经正式发表或未形成正规载体的一种文献形式,如书信、手稿、文献草稿、会议记录、实验笔记、口头演讲等。零次文献具有客观性、针对性、直观性强等特点,但也有传播面窄、难以保管以及信息零散和不成熟的缺点。零次文献一般是通过口头交谈、参观展览、参加报告会等途径获取,不仅在内容上有一定的价值,而且能弥补一般公开文献从信息的客观形成到公开传播之间费时甚多的弊病。

(2) 一次文献(Primary Literature), 指人们直接以自己的生产、科研、社会活动等实践经验为依据生产出来的文献, 也常被称为原始文献(或叫一级文献), 其所记载的知识、信息比较新颖、具体、详尽, 如期刊论文、专利说明书、科技报告、会议论文、学位论文、技术标准等。这些文献具有创新性、实用性和学术性等明显特征, 是科技查新工作中进行文献对比分析的主要依据。

(3) 二次文献(Secondary Literature)又称检索性文献, 指对一次文献按照其外部特征和内容特征进行有序化加工、整理所形成的文献形式, 如目录、题录、索引、文摘等。二次文献在内容上不具有原创性, 只是一次文献特征的汇总, 提供检索一次文献的线索。二次文献内容相对集中、系统性强, 便于管理和传播, 是查新工作中检索文献所利用的主要工具。

(4) 三次文献(Tertiary Literature)又称参考性文献, 指围绕某个专题, 在一、二次文献的基础上, 经过筛选、综合、分析和浓缩等深度加工而形成的文献, 如专题评述、动态综述、进展报告以及手册、百科全书、年鉴、大全、词典等工具书。三次文献是对现有成果加以评论、综述并预测其发展趋势的文献, 具有系统性、综合性、知识性和概括性等特点。在查新工作中, 利用三次文献可以在短时间内了解其研究历史、发展动态及水平等, 以便能更准确地掌握待查项目的技术背景, 把握查新点。

总之, 从零次文献、一次文献、二次文献到三次文献, 是一个由分散到集中, 由无序到有序, 由多到精的对知识信息进行不同层次加工的过程。它们所含信息的质和量是不同的, 对于改善人们的知识结构所起到的作用也不同。零次文献和一次文献是最基本的信息源, 是文献信息检索和利用的主要对象; 二次文献是一次文献的集中提炼和有序化, 它是文献信息检索的工具; 三次文献是把分散的零次文献、一次文献和二次文献, 按照专题或知识的门类进行综合分析加工而成的成果, 是高度浓缩的文献信息, 它既是文献信息检索和利用的对象, 又可作为检索文献信息的工具。

2. 按照文献的载体分类

随着信息记录与存取技术的发展, 文献载体形式呈现多样化, 如音像磁带、缩微胶卷、光盘等, 这些非纸型文献的出现使文献的范围进一步扩大; 使文献的生产和传递更加迅速; 使知识、信息的存储和利用更加便捷。目前文献主要有印刷型、缩微型、电子型(机读型)、音像型 4 种。

(1) 印刷型文献(Paper Literature), 指以手写、打印、印刷等为记录手段, 将信息记载在纸张上形成的文献。它是传统的文献形式, 便于阅读和流传, 但存储密度小、体积大, 不便于管理和长期保存。

(2) 缩微型文献(Microform), 指以感光材料为载体、用缩微照相技术制成的文献复制品, 如缩微胶卷、缩微平片。其特点是存储密度大、体积小, 便于保存和传递, 但必须借助专门的设备才能阅读。世界上许多文献信息服务机构都将欲长期收藏的文献制成缩微品加以保存。

(3) 电子型文献(Electronic Literature), 指以数字代码方式将图、文、声、像等信息存储到磁、光、电介质上, 通过计算机或类似设备阅读使用的文献, 也称机读型文献, 如各种电子图书、电子期刊、联机数据库、网络数据库、网络新闻、光盘数据库等。其特点是信息存储量大, 出版周期短, 易更新, 传递信息迅速, 存取速度快, 可以融文本、图像、声音等多媒体信息于一体, 信息共享性好, 易复制。

(4) 音像型文献(Audio-Visual Literature), 是采用录音、录像、摄影、摄像等手段, 将声音、图像等多媒体信息记录在光学材料、磁性材料上形成的文献, 又称声像资料、视听资料、音像制品, 如音像磁带、唱片、幻灯片、激光视盘等。音像文献脱离了传统的文字记录形式, 直接记录声音和图像, 给人以直观的感觉, 又称直感型资料。其特点是形象、直观, 尤其适于记录用文字、符号难以描述的复杂信息和自然现象, 但其制作、阅读需要利用专门设备。

3. 按照文献的出版形式分类

按照文献的出版形式和内容, 文献可以分为图书、期刊、报纸、特种文献(学位论文、会议论文、专利文献、标准文献、科技报告、政府出版物、产品样本资料等)。

(1) 图书。联合国教科文组织对图书的定义是: 凡由出版社(商)出版的不包括封面和封底在内 49 页以上的, 具有特定的书名和著者名, 编有国际标准书号(ISBN), 有定价并取得版权保护的出版物称为图书。

图书是以传播知识为目的, 用文字或其他信息符号记录于一定形式的材料之上的著作物; 图书是人类社会实践的产物, 是一种特定的不断发展着的知识传播工具。其包括: 专著、教科书、词典、丛书、工具书、百科全书等。

知识窗: ISBN 号

国际标准书号(International Standard Book Number, ISBN)是国际通用的图书或独立的出版物代码。一个国际标准书号只有 1 个或 1 份相应的出版物与之对应。国际标准书号由 13 位数字组成。前 3 位数字代表图书, 中间的 9 个数字分为 3 组, 分别表示国家代码、出版社代码和书序码, 最后一个数字是校验码, 从 0 到 9 或 X。

格式举例: 《科技信息检索》, ISBN 为 978-7-030-26151-9。

(2) 期刊, 也称杂志, 是由多位作者撰写的不同题材的作品构成的定期出版物。期刊有固定刊名, 以期、卷、号或年、月为序。期刊是指定期或不定期连续出版的印刷读物, 每期的内容不重复。期刊出版单位出版期刊, 必须经新闻出版总署批准, 持有国内统一连续出版物编号(ISSN)。

根据期刊的出版周期, 期刊可以分为: 旬刊、半月刊、月刊、双月刊、季刊、半年刊、年刊。期刊按用途不同可以分为科普类期刊和学术类期刊两大类。学术类期刊按主管单位的不同, 可以分为省级、国家级、科技核心期刊(统计源期刊)、中文核心期刊(北大中文核心)、中文社会科学引文索引(CSSCI)、中国科学引文数据库(CSCD)、双核心期刊等。

知识窗：ISSN 号

国际标准连续出版物编号(International Standard Serial Number, ISSN)是根据国际标准 ISO3297 制定的连续出版物国际标准编码,是连续出版物的唯一代码标识。该编号是以 ISSN 为前缀,由 8 位数字组成。8 位数字分为前后两组,每组 4 位,中间以连接号相连。

格式举例:《中国经济问题》,ISSN 为 1000-4181。

(3) 报纸,也是连续出版物的一种,是以刊载新闻和时事评论为主的定期向公众发行的印刷出版物。其是大众传播的重要载体,具有反映和引导社会舆论的功能。根据出版周期,报纸可分为日报、早报、晚报、双日报、周报、旬报等。

(4) 学位论文,是指学生为了获得所修学位,按要求所撰写的论文。学位论文是学术论文的一种形式,有严格的格式要求,一般不公开出版。学位论文分为学士论文、硕士论文、博士论文 3 种。

(5) 会议论文,是指在会议等正式场合宣读的首次发表的论文。会议论文属于公开发表的论文,一般正式的学术交流会议都会出版会议论文集。会议论文集不是期刊,但是有的期刊为会议论文出增刊。

(6) 专利文献,是包含已经申请或被确认为发现、发明、实用新型和工业品外观设计的研究、设计、开发和试验成果的有关资料,以及保护发明人、专利所有人及工业品外观设计和实用新型注册证书持有人权利的有关资料的已出版或未出版的文件(或其摘要)的总称。

(7) 标准文献,是经公认权威机构(主管机关)批准的一整套在特定范围(领域)内必须执行的规格、规则、技术要求等规范性文献,简称标准。

(8) 科技报告,是记录某一科研项目调查、实验、研究的成果或进展情况的报告,又称研究报告、报告文献。每份报告自成一册,通常载有主持单位、报告撰写者、密级、报告号、研究项目号和合同号等。其按内容可分为报告书、论文、通报、札记、技术译文、备忘录、特种出版物。

(9) 政府出版物,是指由政府机构制作出版或由政府机构编辑并授权指定出版商出版的文献。常见的政府出版物有报告、公报、通报、通讯、文件汇编、会议录、统计资料、图表、地名词典、官员名录、国家机关指南、工作手册、地图集以及传统的图书、期刊,也包括缩微、视听等其他载体的非书资料。

(10) 产品样本资料,是指厂商或贸易机构为宣传和推销其产品而印发的资料,如产品目录、产品说明书、产品总览、产品手册等。

1.2 信息检索基础

信息时代一个突出的特点是信息爆炸。人们每天都面临着来自四面八方的庞大信息量,快速准确地找出需要的信息,提高信息的利用率,掌握有效的信息检索方法是很有必要的。

1.2.1 信息检索的原理

1. 信息检索的概念

检索“Retrieval”即“查找”之意。信息检索的定义有广义和狭义之分，广义的信息检索“Information Retrieval”是指将信息按照一定的方式组织和存储起来，并根据信息用户的需要揭示、查找、传递相关信息的过程，包括了信息的存储过程和查找过程。狭义的信息检索仅指信息的查找，是指从信息集合中找出用户所需要的有关信息的过程。狭义的信息检索包括3个方面的含义：了解用户的信息需求；掌握信息检索的技术或方法；满足信息用户的需求。

国内外有关专家关于信息检索给予了不同的解释，较有代表性的观点主要有以下几种。

(1) 信息检索的范围较大。动态信息、静态信息、声频信息、视频信息及各种数值信息均属信息检索范围。如果将信息检索作为一门学科，它应该包括矩阵记数法、概率论、最优化理论、模式识别及系统分析技术等各学科领域的内容。

(2) 信息检索主要是文献检索。信息检索是从大量的文献中查找出与情报提问所指定的课题(对象)有关的文献，或者是包含用户所需事实与消息的文献的过程。这里谈到的文献，不仅指文献线索，也包括文献的片断，如章、节、段落以及与事实有关的直接情报等。

(3) 信息检索是指将信息按一定的方式组织起来，并根据用户需求找出相关信息的过程。这是指信息的存储与检索，是针对信息工作者和用户来定义的，如果仅针对用户，信息检索是指在信息集合中找出所需信息的过程。

2. 信息检索的原理

信息检索的目的就是在信息用户与信息源之间充当媒介。信息检索经历了手工检索、计算机检索、网络检索和智能化检索等多个发展阶段，其基本原理就是在对信息进行整理排序形成检索工具的基础上，按照用户的要求利用检索工具或检索系统，将用户检索提问标识(检索词)与已形成的或存储在系统中的信息的存储标识(文献特征标识、标引词)进行匹配比较，若取得一致，则为匹配，即达到了用户的检索需求，如图 1.1 所示。

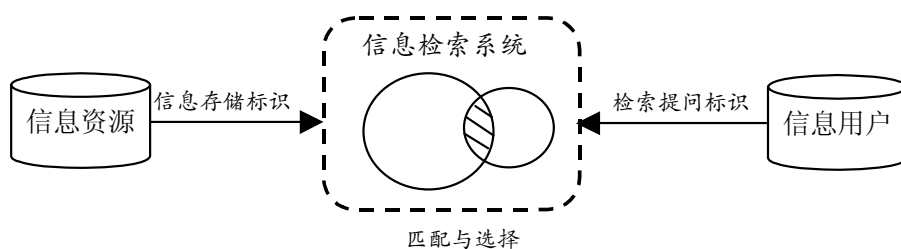


图 1.1 信息检索原理图

由信息检索原理可知，信息的存储是实现信息检索的基础。这里要存储的信息不仅包括原始文档数据，还包括图片、视频和音频等，首先要将这些原始信息进行计算机语言的转换，并将其存储在数据库中，否则无法进行机器识别。待用户根据意图输入查询请求后，

检索系统根据用户的查询请求在数据库中搜索与查询相关的信息,通过一定的匹配机制计算出信息的相似度大小,并按从大到小的顺序将信息转换输出。

1.2.2 信息检索的类型

1. 按信息存储和检索的内容分类

信息检索按存储和检索的内容可划分为:文献检索、数据检索和事实检索。

(1) 文献检索。文献检索通常指的是在检索系统中检索以二次文献为对象(目录、索引、文摘)的信息,它们是文献信息的外部特征与内容特征的描述集合体。信息用户通过检索获取的是原文的“替代物”。

(2) 数据检索。数据检索指在检索系统中检索的数值型数据,如科学技术常数、各种统计数据、人口数据、气象数据、市场行情数据、企业财政数据等,即事物的绝对值和相对值的数字。检索系统提供一定的运算推导能力,例如外推、内插、填补空缺数据,甚至列出曲线图或进行各种分析等功能。信息用户可用通过检索获得的经过核实、整理的数值信息再做定量分析。

(3) 事实检索。事实检索指在检索系统中的原始文献中检索关于某一事物(事件、事实)发生的时间、地点和过程(情况)等方面的信息。它包括数值信息和系统数据信息。一般从系统中检索出所需信息后,再加以逻辑推理才能给出结论。这类信息主要是用于管理决策。

以上3种信息检索类型的主要区别在于:数据检索和事实检索是要检索出包含在文献中的信息本身,而文献检索则检索出包含所需要信息的文献即可。

2. 按照检索系统中信息的组织方式分类

信息检索按系统中信息的组织方式可划分为全文检索和多媒体检索。

(1) 全文检索。全文检索是指检索系统中存储的是整篇文章乃至整本书。检索时,用户可以根据需要从中获取有关的章、段、句、节等信息,还可以进行各种频率统计和内容分析。随着计算机容量与运算速度的增大和提高,全文检索正迅速扩大到各学科、专业。

(2) 多媒体检索。多媒体检索是对超文本检索的补充。其存储对象超出了文本范畴,融入了静、动态图像(形)以及声音等多种媒体信息。信息的存储结构从单维发展到多维,存储空间范围在不断扩大。

3. 按照信息存储的载体和实现查找的技术手段分类

信息检索按照信息存储的载体和实现查找的技术手段可划分为以下3种:

- (1) 手工检索;
- (2) 机械检索;
- (3) 计算机检索。

其中发展比较迅速的计算机检索是“网络信息检索”,是指互联网用户在网络终端,通过特定的网络搜索工具或是通过浏览的方式,查找并获取信息的行为。

1.2.3 信息检索的工具

信息检索工具是指为检索需要而组织的文献检索系统,它是收集、加工、标引、组织文献的结果,是检索工作赖以生存的手段和条件,包含目录、索引、文摘等。

典型的检索工具通常由3部分组成:主体、辅助索引和使用说明。

1. 主体

主体即检索工具的主要部分,由所收录文献的著录部分组成,是文献储存与检索的实体。其包含文献的内部特征与文献的外部特征。内部特征包含分类系统、主题系统和文摘等。外部特征包含书名系统、著者系统、编码系统等。

2. 辅助索引

辅助索引是在文献检索工作中,从主体部分的文献款目编排系统以外的角度增加检索途径、扩大检索效果的辅助手段,它不能离开检索工具的主体单独存在。

3. 使用说明

使用说明向用户介绍本检索工具的性质、内容范围、收录年限、读者对象、著录方法、分类体系、检索方法等,帮助用户利用检索工具。

有的检索工具还包含附录,其附在检索工具后面,一般包括出版物对照表、代码表等。

信息检索工具具有以下基本特点:

- (1) 详细描述文献的内容特征和外部特征;
- (2) 每条文献记录必须有检索标识;
- (3) 文献条目按一定顺序形成一个有机整体,能够提供多种检索途径。

1.2.3.1 目录

1. 目录的定义

目录是图书以及其他单独出版物的外部特征的解释和系统化记载,是将一批文献款目按照一定的次序或规则编排而成的一种揭示与报道的文献工具。

小提示: 图书著录格式

书名: 信息检索导论

责任者: 叶继元主编

出版项: 北京: 电子工业出版社, 2002

载体项: 412页: 20cm

价格: CNY33.00

ISBN: 7-5053-8694-8

索书号: G252.7/Y621

主题词: 情报检索—高等学校—教材

2. 目录的著录款目

目录的著录包括如下款目。

- (1) 书名(刊名)项。
- (2) 著者(编者)项。
- (3) 出版项: 包括出版地点、出版者、出版期、版次等。
- (4) 稽核项: 包括页数、图表、开本、装订形式、定价等。
- (5) 附注项: 对上述各项的补充或说明。
- (6) 提要项: 内容简介。
- (7) 业务注记: 指图书馆或保存单位在目录上做的业务记载, 包括索取号、登录号、分类号、主题词、存储地点等。

3. 目录的排检

1) 书名目录

- (1) 中文图书或期刊, 按中文字顺排检。方法有汉语拼音法、笔画法、部首法等。
- (2) 外文图书或期刊, 西文书名或篇名按字顺排检。日文书名或篇名按汉字或假名排检, 如按汉字排检, 用笔划法; 如按假名排检, 利用假名顺序。俄文按俄文字母顺序排检。

2) 著者目录

中文著者按中文字顺排检。外国著者按先姓后名排检。欧美著者一般名在前姓在后, 编排时需将其倒置。

3) 分类目录

分类目录是按分类法规定的图书、资料所属的学科、专业内容对应的分类号编目, 并按分类号的顺序排检。分类号可以由阿拉伯数字、字母或两者结合组成。

4) 主题目录

主题目录是按图书、资料的主题内容编排的目录。主题目录使用主题词的字顺排检。

4. 目录的作用

目录能反映一定历史时期科学文化发展的概貌, 是人们对浩如烟海的文献加以控制的有效手段, 是查阅和利用文献必不可少的工具。目录的具体作用概括起来主要有几个方面。

(1) 推荐作用。针对特定读者, 按知识的连续性和发展的阶段性编纂和推荐文献, 指导阅读, 成为治学的途径。因此有人将目录比作在海洋中航行不可缺少的导航图。

(2) 检索工具。其能够帮助读者从特定角度去查找资料, 例如题名、著者、主题等。理想的目录应具有分析的功能, 能满足读者检索某一著作中最小的特殊单元的要求。专科目录根据不同的研究需要选编不同学科内容、不同水平和各文种的资料, 分门别类加以编排以供查阅, 为专业研究提供方便。

(3) 文献工作。其能够用以核对各个著录项目, 如著者、题名、版本和日期等。

(4) 提供出处。其有助于文献资料的收集选择、编目加工、报道推广和翻译评介。

1.2.3.2 索引

1. 索引的定义

索引是将报刊书籍中某些重要的或有意义的信息,如书名、刊名、篇名、主题、人名、地名等分别摘录出来,按一定方式编排,并注明出处,以供检索的工具。它包括4个基本要素:索引源、索引款目、编排方法和出处指引系统。

2. 索引的作用

索引可以将文献所包含的若干信息分析摘录出来,作为排检标识,这种标识可以根据需要,在一定的范围内进行任意深度的选择和标引,从而满足更深层次或更多途径的查找要求。索引可以方便用户更快地达到检索目标,是依附在目录或文摘等检索工具之上的辅助性工具,不能独立存在。

3. 常用索引

1) 主题索引

主题索引是将文献中具有实质意义的词语或能揭示文献主体概念的词语抽出来,经过规范化处理后,再按字顺排列起来形成标识系统。通常在各主题词下面给出副标题词,或在各主题词下面给出篇名性的说明语、关键词性的说明语,在说明语后列出文摘号,以这种方式编制的索引称为主题索引。

主题索引提供了一个按揭示内容的主要词汇进行查找文献的检索途径,能迅速查找到专指度高的文献内容,越来越受到文献工作者的青睐。

2) 分类索引

分类索引是将文献内容所属的分类号按隶属关系排列起来,并列出分类号对应的类名。以这种方式列出的索引称为分类索引。分类索引提供了一个按学科分类角度进行查找文献的检索途径,可以迅速查到泛指度较高的文献。

3) 著者索引

著者索引是以文献著者的姓名进行标引的索引,按姓名字顺排列。著者索引在著者名字后面列出该著者所著文献的文摘号,读者根据文摘号可查阅文摘或原文。著者索引提供了一个按著者姓名进行查找文献的检索途径。

常用的索引还有著者所在单位索引(或团体著者索引)、化学分子式索引、引用刊物索引、出版单位索引等。

1.2.3.3 文摘

1. 文摘定义

文摘是以提供文献内容梗概为目的,简明、确切地描述文献主要内容的短文。我国国家标准 GB3468—83 定义文摘为“除题录部分外还对文献内容作实质性描述的文摘条目称文摘”。文摘客观、如实地反映了原文献的内容,但又比原文文字简洁。作为一种检索工具,文摘将论文或书籍中摘录出来的主要论点、数据等按一定方式编排,供读者检索和阅读。

目录和文摘的共同点都是以论文、文章、报告等为主要报道对象，区别主要是以报道文献的外表特征为主还是以报道内容特征为主，在形式上是看有无摘录或评价原文的文字材料。文摘提供文献的外表特征和内容特征，目录仅提供文献的外表特征。

2. 文摘作用

文摘的准确作用概括起来有以下4点。

1) 帮助读者快速判断文献内容

通过目录查找文献仅能查阅文献题目，通过文摘查找文献不仅能查阅文献题目，还能查阅文献内容。文献题目虽然在一定程度上表达了文献内容，但与文摘相比相差甚远，利用文摘可以减少误检和漏检。

2) 节省阅读时间

由于文摘是对原文献主要内容的描述，每一条文摘实际上是一篇高度浓缩、信息完整的文献，因此，读者在不需要阅读原文的情况下，可真实地了解该文献的内容，明确其基本要点，并可直接引用，从而节约阅读时间。

3) 减少语言障碍

文摘能帮助读者获得因语言障碍无法得到的科学文献。在目前出版的科学文献中，有50%以上的文献是以科学家没有掌握的语言出版的，这妨碍了国与国之间学术信息的传递、交流和利用。用本国语言或通用语言翻译的文摘，可以帮助读者克服语言障碍，了解国外有关领域的发展水平和趋势。

4) 替代原文

由于文摘所摘录的是经过筛选的某一学科或某一专业领域最新、最有学术价值的文献，并把相关论题集中一处，因而通过一组文摘，能获得该学科专业领域的学术概况和最新进展。在难以获得原文或不需获得原文的情况下，阅读文摘就显得非常重要了。

1.2.3.4 检索工具类型

检索工具按照不同的标准可以划分为不同的种类。

1. 按信息加工的手段或设备划分

按信息加工的手段或设备划分，检索工具可分为以下几类：

- (1) 手工检索工具；
- (2) 机械检索工具；
- (3) 计算机检索工具。

2. 按编制方法划分

按编制方法划分，检索工具可分为以下几类：

- (1) 目录型检索工具；
- (2) 题录型检索工具；
- (3) 文摘型检索工具；
- (4) 索引型检索工具。

3. 按信息载体形态划分

按信息载体形态划分,检索工具可分为以下几类:

- (1) 书本式检索工具,包括期刊式、单卷式和附录式;
- (2) 卡片式检索工具;
- (3) 缩微式检索工具;
- (4) 磁性材料式检索工具。

4. 按收录范围划分

按收录范围划分,检索工具可分为以下几类:

- (1) 综合性检索工具;
- (2) 专科性检索工具;
- (3) 专题性检索工具;
- (4) 全面性检索工具;
- (5) 单一性检索工具。

5. 按时间范围划分

按时间范围划分,检索工具可分为以下几类:

- (1) 预告性检索工具;
- (2) 现期通报性检索工具;
- (3) 回溯性检索工具。

1.2.4 信息检索的意义

随着信息技术的飞快发展,信息、物质和能源已经成为人类社会的三大财富,信息对于经济、社会的发展,科学、文化的进步都起着重要的作用。有效、快速、准确地在信息海洋中找到所需要的信息,并有效地利用信息,已经成为信息社会人才的必备素质,因此,掌握信息检索的方法和技能有着重要的意义。

1. 掌握获取知识的捷径,提高信息利用的效率

现在,社会信息丰富多样,知识的存储过于庞大和无序,人们在“信息的海洋”中面临着3种挑战:无限的文献资料对有限的阅读时间的挑战;急涌而至的文献对人们接受能力的挑战;大量新知识的出现对人们理解能力的挑战。信息检索正是从大量无序知识中搜索有用的、准确的知识的技能,是快速获取知识的捷径,同时又能帮助人们提高信息利用的效率。信息检索已经构成知识体系中一个不可缺少的部分。

2. 提高信息素质,培养终身学习的能力

信息素质是一种能够发现信息需求,查询、分析判断、加工、筛选、综合利用、创造信息的各方面能力的总和。提高了人们的信息素质就培养了人的独立自主学习的态度和方法,使之善于从瞬息万变的事物中捕捉信息、创造新信息。信息素质已成为每个社会成员的一种基本生存能力,更是在学习型社会中终身学习的必备素质。信息素质的培养日益成

为世界各国教育界乃至社会各界所关注的理论与实践的重大课题。

3. 科学研究的向导，科技查新的基础

信息检索在信息用户与信息源之间充当媒介的作用。信息检索的任务就是将用户与信息源经济地、有效地结合在一起。文献检索是科技查新的基础，查新是以通过检出文献的客观事实来对项目的新颖性做出结论，帮助研究人员继承和借鉴前人成果，避免重复研究或走弯路，节省研究人员查找文献的时间。

1.3 信息检索的途径

信息检索工具是把众多的信息资源进行分析加工后，按照一定的特征标识排检组织而形成的信息集合体。信息检索就是根据用户的检索需求，分析信息的某些特征标识，然后利用检索工具进行匹配比较的过程。所谓信息检索途径就是利用信息的特征来实现信息的查找。

根据信息的基本特征，可以把信息检索途径分为外部特征检索途径(题名途径、著者途径、引文途径等)和内容特征检索途径(分类途径、主题途径等)两大类。

1.3.1 外部特征途径

信息源的外部特征是指在文献载体的外表上标记的可见特征，如：题名(刊名、书名、篇名等)、责任者(作者、编者、译者、专利权人、出版机构等)、号码(标准号、专利号、报告号、索取号等)。

1. 题名途径

题名途径指通过文献的题名来查找文献的途径。题名包括文献的篇名、书名、刊名、专利名称、标准名称、数据库名等，检索时可以利用检索工具的书名索引、刊名索引、会议论文索引等进行。

2. 著者途径

著者途径是指根据文献责任者名称来查找文献的途径，包括个人著者和团体著者。许多检索系统备有著者索引、机构(机构著者或著者所在机构)索引，专利文献检索系统有专利权人索引，利用这些索引按著者、编者、译者、专利权人的姓名或机关团体名称字顺进行检索的途径统称为著者途径。

3. 引文途径

文献所附参考文献或引用文献，是文献的外表特征之一。利用这种引文而编制的索引系统，称为引文索引系统，它提供了从被引论文去检索引用论文的一种途径，称为引文途径。

4. 其他途径

有些文献有特定的序号,如专利号、报告号、合同号、标准号、国际标准书号和刊号等。文献序号对于识别一定的文献,具有明确、简短、唯一性等特点,依此编成的各种序号索引可以提供按序号自身顺序检索文献信息的途径。

总之,以外部特征途径进行检索的最大优点是,它的排列与检索方法以字顺或数字为准,不易错检和漏检,查准率高。以文献外部特征作为检索途径适宜查找已知文献题名、作者姓名或序号的文献。

1.3.2 内容特征途径

信息源的内容特征是文献所载的知识信息中隐含的、潜在的特征,如分类、主题等。以文献内容特征作为检索途径更适合检索未知线索的文献。

1. 分类途径

分类途径是指按照文献所属学科(专业)属性(类别)进行检索的途径。分类途径是以课题的学科属性为出发点,按照学科分类体系,利用学科分类表、分类目录或分类索引等进行检索。按分类途径检索文献便于从学科体系的角度获得较系统的文献线索,即有族性检索功能,便于从学科所属范围来查找文献资料。

分类途径可把同一学科的文献信息集中检索出来,但一些新兴学科、边缘学科的文献难以给出确切的类别,易造成误检和漏检。因此,从分类途径查找文献,一定要掌握学科的分类体系及有关规则。

目前,中国文献分类方法主要有5种:《中国图书馆图书分类法》、《中国图书资料分类法》、《中国科学院图书馆图书分类法》、《中国人民大学图书馆图书分类法》和《国际图书集成分类法》。国外比较重要的分类法有《杜威十进分类法》、《国际十进分类法》、《美国国会图书馆图书分类法》等。

知识窗:《中国图书馆图书分类法》类目中的基本大类

- | | |
|---------------------|----------------|
| A. 马克思主义、列宁主义、毛泽东思想 | N. 自然科学总论 |
| B. 哲学 | O. 数理科学和化学 |
| C. 社会科学总论 | P. 天文学、地球科学 |
| D. 政治、法律 | Q. 生物科学 |
| E. 军事 | R. 医药、卫生 |
| F. 经济 | S. 农业科学 |
| G. 文化、科学、教育、体育 | T. 工业技术 |
| H. 语言、文字 | U. 交通运输 |
| I. 文学 | V. 航空、航天 |
| J. 艺术 | X. 环境科学、劳动保护科学 |
| K. 历史、地理 | Z. 综合性图书 |

2. 主题途径

用规范化词语来表达文献信息的内容特征的词汇叫主题词。主题途径是按照文献信息的主题内容进行检索的途径,利用能代表文献内容的主题词、关键词、叙词,并按字顺排列实现检索。

利用主题途径检索文献关键在于分析检索目标的核心、提炼主题概念,并能运用合适的词语来表达主题概念。主题途径是一种主要的检索途径,表达概念灵活、准确,能集中反映一个主题的各方面文献资料,所获得的检索信息专指性强,便于读者对某一问题、某一事物和对象做全面系统的专题性研究。

3. 其他途径

利用事物的某种代码编成的索引,如分子结构式索引、环系索引等,可以按特定代码顺序进行检索。还有某些专门项目途径,如按文献信息所包含的或有关的名词术语、地名、人名、机构名、商品名、生物属名、年代等的特定顺序进行检索,可以解决某些特别的问题。

1.4 信息检索的步骤和方法

1.4.1 信息检索的步骤

无论是手工检索还是计算机检索,都是一个经过仔细思考并通过实践逐步完善查找方法的过程。检索过程通常要包含以下几个步骤:分析课题、选择检索工具、确定检索途径及检索式、进行检索和获取原文,如图 1.2 所示。

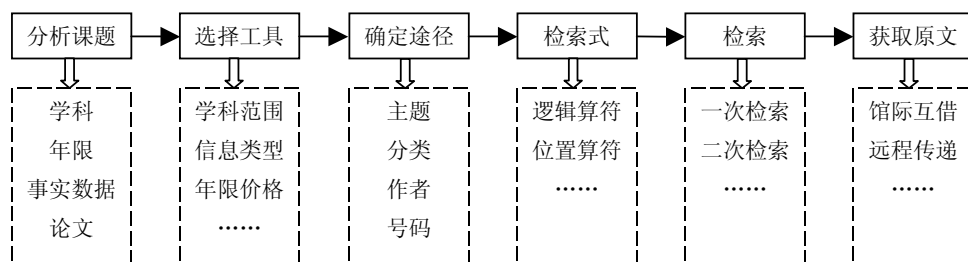


图 1.2 信息检索的步骤

1.4.1.1 分析课题

在确定研究课题后,研究者应知道所需信息的学科性质、特点和水平层次。

1. 明确课题的学科属性及专业范围

首先考虑检索课题是单一学科、多学科还是跨学科。当课题涉及多学科时,应以主要学科为检索重点,次要学科为补充。明确检索的学科范围——越具体越有利于检索。如果问题属于多学科或交叉学科,则分别列出多学科或交叉学科的相关部分,确定所需信息在复杂的学科体系中的位置。

2. 明确课题的信息检索类型

弄清检索课题是属于文献检索、数据检索还是事实检索。文献类课题主要以图书、期刊、学位论文、会议论文、专利等文献为检索对象。数据类课题的检索是要根据用户的需要,从某种数据集合中检索出准确数据。事实检索是一种确定性检索,一般要将检索到的信息进行逻辑推理、分析后才能得出结论。

3. 明确课题的检索时间要求

对于学科发展较快的领域,应该缩短检索时间。而对于一些特殊的研究课题(如历史类课题研究),则要考虑特定的检索时间。

1.4.1.2 选择检索工具

检索工具成千上万,除了要知道哪类问题该用哪类工具解决,还要熟悉一些具体检索工具的内容、范围、特点与编排结构。选择检索工具的主要依据包括以下几个方面。

(1) 检索工具所包含的情报信息是否广泛全面、准确可靠,这有利于查全、查准。

(2) 其检索工具中文献报道量是否充足,这是衡量检索工具存储功能的主要标准。

(3) 著录款目是否准确、详尽,这是衡量程度的主要标准。

(4) 信息报道时差是否短,这是衡量检索工具报道功能的主要标准。报道时差是指文献从第一次发表到被收录于检索工具中的时间差。报道时差短,它提供的文献内容新,信息及时。

(5) 检索途径是否完善,这是衡量检索工具检索功能的主要标准。检索工具中,辅助索引种类越多,检索时也就越方便。

确定了检索工具类型之后,还要根据编撰目的、材料收录范围和编排方法,选择有希望获得所需信息的检索工具。检索工具收录的学科范围、地区范围、语种范围、时间范围和文献类型,在每一种检索工具书的用户指南(或导言、前言、编例、User's Guide、Helps for the reader等)中都有说明,这是首次使用某一检索工具前必读的内容。

1.4.1.3 确定检索途径及检索式

检索工具确定以后,应确定检索途径找到所需信息。检索途径往往不止一种,使用者应根据“已知”信息特征确定检索途径。所有文献的特征可分为两大类:外部特征(如著者、题名、文献类型、文种、发表时间、语种、文献号等)和内容特征(如分类号、主题、检索词、代码等)。

检索时,如果已知文献的外部特征,尽量使用该外部特征进行检索,以便达到信息检索更高的专指度。在不知文献外部特征的情况下,通常采用内容特征进行检索。在采用内容特征进行检索时,如果希望检索词泛指度高,则通常选择分类途径。如果希望检索词专指度高,则可采用主题途径。有些检索工具仅采用一种方式进行标引,用户只能根据已有的标引方式进行检索。

检索式是计算机检索中用来表达检索提问的一种逻辑运算式,又称检索表达式或检索提问式。它由检索词和检索系统允许使用的各种运算符组合而成,是检索策略的具体体现。

构造检索式就是把已经确定的检索词和分析检索课题时确定的检索要求用检索系统所支持的各种运算符联结起来,形成检索式。

1.4.1.4 进行检索

在检索系统中将检索标识与系统中存储的文献标识进行匹配,查出相关文献,并对所获结果进行分析,看其是否符合需要。如果试检结果满意,可进行正式检索;否则,要分析原因,修改、调整检索策略。调整检索策略包括修改检索式、调整检索词、重新选择检索系统等。

在实际研究课题的检索过程中,往往进行一次检索是达不到目的,一般都要反复调整检索策略或在一次检索结果中进行二次检索。

1.4.1.5 获取原文

从检索工具上获得所需文献线索后,利用馆藏图书目录或报刊目录获取原文。由于馆藏目录或联合目录一般提供了文献的索取号和收藏地址,获取原文时只要履行借阅手续或馆际互借手续即可,有时可从著者处获取原文。现在许多网络检索工具都提供原文的链接,只要用户支付相关费用就可以方便地获取原文。如果不能在网络上获取原文,应收集以下信息,通过这些信息获取原文。

(1) 识别文献类型。不同类型的文献收藏地点不同、外表特征不同,在索取原文时首先就要区别文献的类型。

(2) 充分利用刊名。检索工具通常采用刊名缩写,并附有引用出版物一览表,供查对收录的出版物全称之用。

(3) 论文著者的地址。获取会议论文、学位论文、公司报告以及一些尚未公开发表的文章的原文,必须获得论文著者(包括团体著者)的详细地址。大多检索刊物的著录项中附有作者的工作单位名称缩写,可以据此查阅机构名录全称。

(4) 识别语系。在检索工具中,俄文、中文、日文等的文献作者、出版物名称通常采用音译法转换成英文进行著录。故索取原文前,要将这些音译的人名、出版物名称还原成原来的语种。

(5) 其他外部特征。文献的其他外部特征包括文献篇名、编号(索书号和专利号等)、著者姓名、出版时间、分类号等。

了解文献的上述特征后采用由近及远的方式进行查找,可通过以下途径获取原文。

其一,从本单位图书情报部门获取原文。这是获取原文最方便、最经济的途径。首先应立足于本馆(或本所),其次是附近的图书馆或信息中心。如果读者目录有缺省,应利用公务目录、典藏目录或各院系分馆的目录这类保存较好的、能完整反映馆藏的图书期刊目录,或通过馆际协作获取原文。

其二,从著者处获取原文。给国外的论文著者写信索取复制件已是国际上通用的学术交流方式,美国有一些情报机构开展这方面的服务,如研究图书馆中心(CRL)、国家技术情报服务中心(NTIS)、美国专利商标局(PTO)等;亦可以从学术团体指南这一类工具书中获

得更详尽的资料。

其三,从检索刊物出版机构获取原文。国外一些著名检索机构如美国化学文摘社(CAS)、情报科学研究所(ISI)等都可向用户提供原文。ISI 设有原迹论文服务,可使用户看到彩色版图、照片及复杂的图像资料。有些非营利性出版物是商业书目所不收录的,要获取这些文献可以通过学术机构或信息服务名录获得有关信息,然后通过函索方式获取出版物。

其四,利用国际联机检索终端向国外订购原文。如果读者急需原文,而一般订购渠道又很慢,则可以采用这个途径。订购时需填写“联机订购原文申请单”,注明文档名称及代号、原文题目、著者和出处等。这是一种较快的订购办法,一般需要半个月左右,但费用昂贵。

1.4.2 信息检索的方法

信息检索方法就是为实现某个检索计划所采用的具体操作方法,其目的在于寻找一种以最少的时间、最佳的途径,获得最满意的检索效果的方法。下面是常用的几种检索方法。

1. 顺查法

顺查法指按照时间的顺序,由远及近地利用检索系统进行文献信息检索的方法。这种方法能收集到某一课题的系统文献,它适用于较大课题的文献检索。例如,已知某课题的起始年代,现在需要了解其发展的全过程,就可以用顺查法从最初的年代开始,逐渐向近期查找。该方法的优点是漏检率、误检率比较低,缺点是工作量大。

2. 逆查法

逆查法是由近及远地查找,逆着时间的顺序利用检索工具进行文献信息检索的方法。此方法的重点是放在近期文献,只需查到基本满足需要时就可以了。使用这种方法可以最快地获得新资料,而且近期的资料总是既概括了前期的成果,又反映了最新水平和动向。这种方法工作量较小,但是漏检率较高,主要用于新课题立项前的调研。

3. 抽查法

抽查法是针对检索课题的特点,选择有关该课题的文献信息最可能出现或最多出现的时间段,利用检索工具进行重点检索的方法。它适用于检索某一领域研究高潮很明显的、某一学科的发展阶段很清晰的、某一事物出现频率在某一阶段很突出的课题。该方法是一种费时较少而又能查到较多有效文献的检索方法。

4. 追溯法(引文法)

追溯法是指利用已经掌握的文献末尾所列的参考文献,进行逐一追溯查找“引文”的一种最简便的扩大情报来源的方法。它还可以从查到的“引文”中再追溯查找“引文”,像滚雪球一样,依据文献间的引用关系,获得越来越多的内容相关的文献。这些内容相关的文献反映着某一课题的立论依据和背景,也在某种程度上反映着某课题或其中的某一观点、某种发现的发展过程。

追溯法的缺点是:原文著者引用的参考文献是有限的,不可能引出全部有关文献,而且有的著者引用某一文献只是为了说明一下经过情况,与原文的内容关系不大。因此,用

追溯法查找文献出现漏检索和误检索的可能性均较大,同时也比较麻烦。由于一般科研课题需要追溯的主要是最新文献登载的最新研究成果,而不是旧文献登载的过时的研究成果,显然单纯使用此法查找文献是很不够的,它具有很大的局限性。

5. 综合法

综合法又称为循环法,这是把上述方法加以综合运用的方法。综合法既要利用检索工具进行常规检索,又要利用文献后所附参考文献进行追溯检索,分期分段地交替使用几种方法。即先利用检索工具(系统)检索到一批文献,再以这些文献末尾的参考目标为线索逆行查找,如此循环进行,直到满足要求为止。

1.5 计算机信息检索基础

随着计算机技术、通信技术和高密度存储技术的迅猛发展,利用计算机进行信息检索已成为人们获取文献信息的重要手段之一。

1.5.1 计算机信息检索概述

计算机信息检索是指利用计算机存储和检索信息,就是指人们在计算机或计算机检索网络的终端上,使用特定的检索指令,从计算机检索系统的数据库中检索出所需信息的过程。

1. 计算机信息检索的特点

与手工检索相比,计算机检索具有如下特点。

(1) 速度快。仅几分钟就可以从成千上万条记录中找出所需信息,特别是现在许多搜索引擎采用更加有效的算法以及宽带网络在通信上的应用使信息检索的速度更快。

(2) 检索范围广。其几乎覆盖互联网上所有的网络资源。

(3) 检索不受时空的限制。只要拥有相应的软件和硬件设备,用户就可以在任意地方借助光盘和通信网络查询所需信息。

(4) 及时。许多专业数据库每周更新一次信息,管理好的网站更新速度很快,新闻信息当天可以看到,甚至快于广播和电视。由于数据更新快,用户可以及时获得最新信息。

(5) 界面友好。检索软件可采用菜单驱动,几乎所有检索系统都有查询服务或提供操作演示盘,界面友好的程度在不断提高,检索结果的输出方式丰富多样,并可以按要求做排序、统计、绘图等加工。Web 信息检索工具界面丰富友好,使用户的检索相当方便,易于操作。

(6) 超文本检索。互联网检索的结果是完全可以直接阅读和利用的 Web 页面文献全文。而且只需要简单地用鼠标点击,即可随意浏览。在多媒体超文本中,不但有文字、图片,还可以有声音、动画、影视等形式的信息内容。

2. 计算机信息检索分类

1) 脱机检索

脱机检索(Off-line Searching)又常常被称为离线检索。作为在线(On-line)检索或联机检索的补充,用户可以针对某一台计算机,某一光盘或某一存储器,在脱离互联网的条件下进行检索。脱机检索是单机分批处理用户提问的检索方式。典型的脱机检索方式是光盘检索。由于各大文献存储机构定期销售光盘,光盘检索有较大的使用空间。这种检索方法针对性强、检索速度快、内容可靠,是不能进行互联网检索时的一种有效的检索方法。

2) 联机检索

联机检索(Online Searching)是计算机终端通过计算机网络系统与系统中的某一主机连接,将用户需求的信息检索出来的检索形式。

最初在计算机网络尚未在全球普及的时期,大多数用户通过世界上比较著名的联机检索系统进行联机检索。这些系统包括:欧洲共同体9国所属的欧洲科技信息联机检索网络EURONET、欧洲空间组织的ESA/IRS系统、美国洛克希德公司的DIALOG系统、美国系统开发公司的ORBIT系统、美国医学图书馆的MEDLINE系统、日本科技信息中心的JICST系统等。互联网技术的发展使得这些系统更好地与网络结合,成为网络检索的一部分。

3) 网络检索

网络检索(Network Searching)是指通过网络接口软件,用户可在任意一个终端查询各地上网的信息资源。网络检索是一种广义的联机检索,如使用远程登录(Telnet),通过Internet连接用户所指定的远程计算机,共享该主机上的资源,这个过程称为联机。

只要主机在互联网上将信息共享,该信息就可被其他用户搜寻使用。现代通信技术的发展给联机检索提供了越来越广阔的空间,人们的检索目标不再是局限于与某台主机连接进行联机检索。现在的电子商务、电子政务、政府上网工程、图书馆、信息服务机构和科研机构等政府和企业或个人信息的网络化使得网络信息成为现代信息获取的主要渠道。这些上网信息构成极其丰富的网络信息资源。互联网的迅速发展和广泛应用,改变了计算机信息检索的方式和方法,将信息检索拓展到一个更广阔的领域。

1.5.2 计算机信息检索技术

计算机信息检索过程实际上是将检索提问词与文献记录标引词进行对比匹配的过程。为了提高检索效率,计算机检索系统经常采用一些运算方法,从概念相关性、位置相关性等方面对检索提问进行技术处理。

1. 检索词的选择

无论手工检索还是计算机检索,检索词的选择都非常重要。检索式由检索词和检索算符组成。检索词分为两类:受控词和非受控词。受控词是事先规范化的检索语言,取自主题词表、叙词表、分类表等,如果数据库对数据采用了受控词汇标引,并有机读式或印本式主题词表时,应优先选用其中的受控词。在一些专业性较强的检索刊物中大多使用受控词汇,以减少标引词汇量。

例如,《工程索引》采用标题词表,是用受控词汇对文献进行标引。受控词的检索效率高,一旦选定宽度适当的概念,系统就能检出这一概念的全部内容,而且,由于标引人员已事先解决了自然语言中的同义、近义关系,使检索相对容易,但由于其先组(根据主题词表进行标引)性质,受控词不能及时反映新事物的发展,概念数量有限,结构复杂,不能适应数量众多、要求各异的信息用户,不易为非专业人员所掌握。

2. 布尔逻辑检索

确定检索途径和检索词后,就可以在检索式输入窗口中输入检索表达式。检索表达式可以是一个检索词或一个逻辑表达式。将多个检索词用布尔运算符(简称检索算符)连接在一起来进行检索,称之为布尔逻辑检索。

将选定的检索词用系统规定的检索算符连接起来,就成为一个体现检索策略的检索表达式。系统使用的检索算符一般有以下几种。

1) 逻辑与

逻辑与用“AND”或“*”来表达。检索式写作“A AND B”或“A*B”,表明数据库中同时有检索词“A”和“B”的记录才为命中记录。逻辑与可增强检索的专指性,缩小检索范围。

例如,用户想在题名中检索有关计算机新闻的文献,如果只输入“新闻”或“计算机”,则命中文献太多,且有许多不是自己所要的。如果输入“新闻*计算机”,则检索出题名中同时含有“新闻”和“计算机”的文献,检索结果更加准确。使用“逻辑与”能防止误检。

2) 逻辑或

逻辑或用“OR”或“+”来表达。检索式写作“A OR B”或“A+B”,表明数据库中凡有检索词“A”或者“B”,或同时有“A”和“B”的记录均为命中记录。使用逻辑或可连接同一检索组面的多个同义词、近义词和相关词,扩大检索范围,有助于提高查全率。

例如,某用户想查找研究杜甫的文献,检索途径选择题名后,输入“杜甫”,命中540篇。但考虑到研究杜甫的文献题名中未必都出现“杜甫”两字,也可能出现“杜诗”、“李杜”,于是改用“杜甫+杜诗+李杜”表达式,结果命中608篇。

3) 逻辑非

逻辑非用“NOT”或“-”来表达。检索式写作“A NOT B”或“A-B”,表明数据库中凡有检索词“A”而不含检索词“B”的记录才为命中记录。使用逻辑非可以排除不希望出现的概念,增强检索的准确性。

4) 算符优先级

当多个布尔运算符在一个检索式中同时出现时,它们的级别是不同的。优先执行顺序通常是“NOT”、“AND”、“OR”,在有括号的情况下,先执行括号内的逻辑运算,在有多层括号时,先执行最内层括号中的运算。布尔检索比较容易掌握,但使用不当会造成大量漏检和误检,特别是非运算符的运用应特别小心,否则会把有用的文献排除了。

检索中逻辑算符使用是最频繁的,对逻辑算符使用的技巧决定对检索结果的满意程度。用布尔逻辑表达检索要求,除了要掌握检索课题的相关因素外,还应注意布尔算符对检索

结果的影响。另外,对同一个布尔逻辑提问式来说,不同的运算次序会有不同的检索结果。布尔算符使用正确但不能达到应有检索效果的情况是很多的。

5) 精确匹配和模糊检索

不同的数据库,检索途径设定的检索规则有所不同,有的检索途径允许用户用“任意一致”的方式检索,有的只允许用“完全一致”或“前方一致”的方式检索。

不同的数据库的模糊检索符号不同。例如,专利检索数据库使用“%”作为模糊检索符号。检索时,可单字段查询,也可多个检索字段组配查询。如果输入的检索内容不完整,可在输入字符的前面或后面加上模糊符号“%”进行模糊检索。

有些数据库不遵循此规律,因此,在检索前必须看使用说明。

1.5.3 光盘检索

光盘信息检索系统是单机信息检索系统的一种,它解决了单机检索系统数据存储量少的缺点,是目前应用较为广泛的一种计算机信息检索系统,它使数据库检索大众化,革新了信息传播和服务方式。在信息检索领域应用的光盘主要是只读光盘。

1. 光盘检索的特点

光盘检索的特点主要有以下几个。

(1) 检索途径多。其可提供内容、著者、关键词、标题、著者单位、年代、出处等多种检索途径。

(2) 检索系统配置简单。利用微机系统建立光盘检索系统只要一台光盘驱动器和所需要的盘片,不需经过通信线路。

(3) 检索费用低廉。由于光盘的费用是一次性投入,可多次任意使用,利用率越高,分摊的成本越低。整个检索过程不涉及远程通信网络问题,仅这一项就可节省很大一部分开支,对于通信网络尚不发达,联机检索费用昂贵或没有国际终端的地区和单位,光盘检索的优势较大。

(4) 系统操作简单。光盘数据库一般都具备类似联机检索那样齐全的检索和输出功能,同时提供人机对话,系统界面友好。没有检索经验的用户在系统提供的菜单式“帮助”指令下可无师自通。光盘系统的服务方式是非限制性检索,用户可以充分考虑和修正自己的检索策略,直到检索结果满意为止,而没有按时间收费的压力。

2. 光盘检索的局限性

光盘检索的局限性主要体现在以下几个方面。

(1) 容量有限。光盘检索一次只能检索一个数据库,而联机检索可同时检索几个数据库,且数据量大。

(2) 更新速度慢。联机检索系统可实时更新,而光盘基本上是定期更新,更新频率快则一个月,慢则一年,这在一定程度上影响了数据的适时性,不适用于那些动态性极强的商情信息检索。

1.5.4 网络检索

作为用户最多、影响最大的计算机网络，互联网堪称世界上资源最丰富的信息库和文档资料库，几乎能满足全球用户对任何信息的需求。

1. 网络检索特点

网络检索具有如下特点。

- (1) 检索速度快。手工检索需数周的课题计算机检索只需数小时或数分钟。
- (2) 检索途径多。除手工检索工具的分类、主题、著者等检索途径之外，网络检索往往还提供更多更灵活的检索途径，如文本词途径。
- (3) 更新快。网上信息很多为日更新。
- (4) 范围广。网上数据库的种类更多，有些可免费进行检索。
- (5) 检索用词的选择及其组配更为灵活。检索策略的制定更为灵活和复杂，可利用各种计算机检索技术随意扩大或缩小检索范围，以满足用户各种不同的信息检索需求。
- (6) 检索结果可直接打印或下载，可操作性强。数据输出时，可任意选择若干记录和若干字段，还可直接获取文献线索甚至在线订购原文，并以 E-mail 等方式直接发送。

2. 数据库的选择

1) 数据库的类型是否满足检索需要

对于专业性强的，特别是科技信息检索，使用专业数据库更能反映专业发展全貌，且使用专业数据库信息较为真实可靠。一个系统往往由上百个数据库，几千个文档和几百万、上千万条记录组成，而查找主要是在文档这个层次上进行。与市场信息有关的数据库种类远多于科技文献数据库，经济或商情类数据库群中缺乏像美国《化学文摘》(CA)、英国《科学文摘》(INSPEC)那样能满足相当大一部分用户的需求的综合数据库和“核心库”。这就要求检索者针对用户的需求仔细了解数据库的内容，做出正确的选择。

2) 数据库的学科专业范围是否与检索课题的学科专业相吻合

检索专业数据库应充分了解提问的学科范围、主题内容、文献类型等信息，综合考虑用户的需要，还应注意需求的文献语种、年限、文献类型、打印格式、费用要求和用户查阅过哪些资料、已经掌握了哪些信息等一些细节问题和背景情况。

3) 数据库描述文献的质量

数据库描述文献的质量，包括对原文的表达程度、标引深度、专指度如何等，以及是否按标准化著录。

4) 检索费用

各种检索系统的收费存在着很大的差异，比如，联机信息检索系统通常收费相当昂贵，而光盘信息检索系统的收费就要低得多，互联网上的搜索引擎甚至是完全免费的。因此在利用计算机信息检索时必须考虑用户的费用承受能力，选择相应的系统进行查询。

3. 联机、光盘、互联网三大计算机信息检索系统比较

联机信息检索、光盘信息检索以及互联网信息检索的检索系统各具特色，相互补充，

多元并存。联机信息检索主要面向专业人员,检索需要掌握诸多复杂的指令,检索费用较高,但检索质量很可靠,检索效率高,能够满足各项要求都较高的检索需求;互联网上的信息检索主要面向一般用户,检索方便、容易,但查找的效率较低,检索的全面性、准确性都较低,适合于信息浏览和要求不太严格的检索需求;光盘信息检索通常用菜单驱动的方式,操作简便易学,且光盘可随意携带,非常适合专题图文信息或数据信息的检索。

本章小结

被称为词典之父的英国学者 S. 约翰逊曾说过:“知识分两类,一类是我们所知道的科学知识,另一类是关于哪儿可以获得这些知识的知识。《信息检索》是“关于哪儿可以获得这些知识的知识”的一门课程。本章通过对信息、文献、信息检索、检索工具、计算机信息检索基础等相关知识的阐述,给读者讲明信息的特征、信息检索的概念、信息检索的类型、信息检索的目的、信息检索方法、信息检索步骤等相关知识,明确信息检索就是在信息用户与信息源之间充当媒介的作用,为后续章节的学习奠定基础。

【关键术语】

信息 知识 文献 目录 索引 文摘 信息检索 检索途径

【综合练习】

一、填空题

1. 文献是信息的主要载体,根据对文献信息的加工层次可将文献分为_____文献、_____文献、_____文献和_____文献。
2. 追溯法是指利用已经掌握的文献末尾所列的_____,进行逐一追溯查找_____的一种最简便的扩大情报来源的方法。
3. 用规范化词语来表达文献信息_____的词汇叫主题词。主题途径是按照文献信息的主题内容进行检索的途径,利用能代表文献内容的主题词、关键词、叙词,并按字顺排列实现检索。
4. 计算机信息检索过程实际上是将_____与_____进行对比匹配的过程。
5. 无论是手工检索还是计算机检索,都是一个经过仔细思考并通过实践逐步完善查找方法的过程。检索过程通常包含以下几个步骤:_____、_____、_____、_____、_____。
6. 检索工具按信息加工的手段可以分为_____、_____、_____。
7. 《中国图书馆图书分类法》共分_____个基本部类,下分_____个大类。
8. 索引包括4个基本要素:索引源、_____、_____和出处指引系统。

二、判断题

1. 在检索信息时,使用逻辑符“AND”可以缩小检索范围。()

2. 逆查法是由近及远地查找, 顺着时间的顺序利用检索工具进行文献信息检索的方法。()
3. 按编制方法划分, 信息检索工具可以分为: 手工检索工具、机械检索工具、计算机检索工具。()
4. 请判断下面图书的国际标准书号的格式是否正确。ISBN: 978-7-030-26151-X。()
5. 文献的专利号、报告号、合同号、标准号、索取号、国际标准书号、刊号属于文献的内部特征。()
6. 二次检索是指在第一次检索结果不符合要求时, 重新选择检索条件再次进行检索。()

三、选择题(单选或多选)

1. 目前, 常用的文献分类方法有()。
(A) 《中图法》 (B) 《科图法》 (C) 《杜威十进分类法》 (D) 《人大法》
2. 信息检索方法包括()。
(A) 顺查法 (B) 逆查法 (C) 抽查法 (D) 追溯法
3. 所有文献的特征可分两大类: 外部特征和内容特征。请指出下面属于内容特征的是:()。
(A) 发表时间 (B) 分类号 (C) 文献类型 (D) 著者
4. 常用的索引包括:()。
(A) 主题索引 (B) 分类索引 (C) 著者索引 (D) 化学分子式索引
5. 检索工具的结构与检索工具的功能和效率密切相关。典型的检索工具通常由3部分组成:()。
(A) 索引 (B) 主体 (C) 辅助索引 (D) 使用说明

四、简答题

1. 请解释信息、知识、文献的概念, 并简述三者之间的关系。
2. 什么是信息检索? 信息检索的目的和意义是什么?
3. 什么是检索工具? 检索工具的结构是什么?
4. 什么是目录? 目录在检索中起什么作用?
5. 什么是索引? 索引在检索中起什么作用?
6. 什么是文摘? 文摘在检索中起什么作用?
7. 国内外主要使用的文献分类法有哪些?
8. 计算机检索的特点是什么?
9. 常用的信息检索方法有哪些?

【检索实训】

姓名: _____ 检索时间: _____

课题 1.1: 利用《中国图书馆图书分类法》找出分类号。

检索目的: 掌握《中国图书馆图书分类法》的基本构成特点。

检索要求：按照基本大类，在《中国图书馆图书分类法》中查找以下相关类目编号。

- (1) 计算机技术：国际互联网。
- (2) 交通运输：运输线路优选。
- (3) 环境科学：环境污染的控制及其排除。
- (4) 食品工业：淀粉加工工艺。
- (5) 文化、科学：图书馆工作者。

检索结果：

课题 1.2：查找图书的著录格式和期刊论文的著录格式。

检索目的：了解图书的著录格式和期刊论文的著录格式，掌握国际标准书号 ISBN 和国际标准连续出版物编号 ISSN 著录格式。

检索要求：

- (1) 检索图书，写出其著录格式，并说明其 ISBN 含义；
- (2) 检索期刊论文，写出其著录格式，并说明其 ISSN 含义。

检索结果：