

信息检索

Xinxi Jiansuo

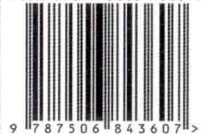
钟 萍 林泽明 著

责任编辑：徐国威 成晓春

设计制作：鸿儒文轩

信息检索

ISBN 978-7-5068-4360-7



定价：60.00元

信自检索

Xinxi Jiansuo

钟 萍 林泽明 著



中国书籍出版社
CHINA BOOK PRESS

图书在版编目 (CIP) 数据

信息检索 / 钟萍, 林泽明著. —北京: 中国书籍出版社, 2014.8

ISBN 978-7-5068-4360-7

I . ①信… II . ①钟… ②林… III . ①情报检索 IV . ① G252.7

中国版本图书馆 CIP 数据核字 (2014) 第 196346 号

信息检索

钟 萍 林泽明 著

责任编辑 徐国威 成晓春

责任印制 孙马飞 马 芝

封面设计 鸿儒文轩

出版发行 中国书籍出版社

地 址 北京市丰台区三路居路 97 号 (邮编: 100073)

电 话 (010) 52257143 (总编室) (010) 52257140 (发行部)

电子邮箱 chinabp@vip.sina.com

经 销 全国新华书店

印 刷 北京富达印务有限公司

开 本 710 毫米 × 960 毫米 1/16

字 数 430 千字

印 张 18

版 次 2014 年 8 月第 1 版 2014 年 8 月第 1 次印刷

书 号 ISBN 978-7-5068-4360-7

定 价 60.00 元

版权所有 翻印必究

前 言

现代科学技术的进步和生产力的发展,特别是信息网络技术的迅猛发展,使人类社会进入了一个新的发展阶段——信息社会。信息素质已经成为人们生存、学习和发展的基本素质,也是衡量个人综合素质的重要指标之一。因此,信息素质教育也得到了更多的关注和重视。

在高等学校,开设信息检索相关课程一直是信息素质教育的主要形式。时代在进步,学科在发展,信息检索课的内容、形式,乃至名称都发生着变化,但其核心宗旨始终如一,那就是以训练学生检索技能为核心,促进学生信息素养的提升,以便充分、有效地开发、利用丰富的信息资源,使之真正成为知识创新的活水源头。

信息检索,是基于特定信息需求的驱动,以信息检索系统为基础,以检索方法与技术为手段,以信息获得与利用为目的,以信息思维的统领为核心的智力活动及其过程。所以信息检索课的教学目标应为:熟悉信息基本概念、类型、特点、功能及分布规律,在此基础上熟悉获取信息的基本途径、方法,尤其是计算机检索的技术方法;掌握常用信息检索系统的结构和检索方法,熟悉网络信息的特点、鉴别及获取的方法和技巧;通过信息的查找、筛选、鉴别、利用过程,体会作为科学素养重要组成部分的信息素养的重要意义。

信息检索课是一门实践性很强的课程,根据编者多年的教学经验,我们在教材策划之初就将其定位为实践性教材,在理论与实践结合的基础上,更加突出其实践性。

(1) 全书以信息素质的核心内容“知源、知取、知用”为主线,全面介绍信息资源的检索与利用。

(2) 本书在第1章详细介绍了图书馆尤其是高校图书馆的资源与服务,并对图书馆资源与 Internet 资源进行对比分析,以引导学生根据信息需求特点对信息源进行恰当选择。

(3) 鉴于目前基于网络的计算机信息检索已经成为人们最普遍利用、效率最高的信息检索形式,本教材以计算机检索为基础,介绍中外文各类常用数据库、学术搜索及互联网资源,对每种资源的介绍都配有大量样图及检索实例,可操作性极强,使读者可以很容易掌握其检索及利用方法。

(4) 为了提高学生对文献信息的分析、研究与利用能力,培养其学术素养,在最后一章,以科学研究的基本流程为脉络,介绍了选题、资料搜集、整理,文献资源管理、论文写作与发表及学术规范等。附录部分设计了信息综合分析研究的具体案例,是对信息资源检索、分析与利用全过程的总结,供读者参考。

本书共分九章。第1、5、6、7、8、9章由钟萍撰写;第2、3、4章由林泽明撰写。

本书在撰写过程中参考了大量专家学者的论著,在此一并表示感谢。

信息检索理论与实践尚有许多问题值得研究和探讨,因学识及时间所限,书中难免有疏漏及错误之处,敬请批评指正。

编著者

2014年7月

目 录 ▶▶

CONTENTS

第 1 章 信息检索的基本理论

1.1 信息检索	001
1.2 信息检索的基本原理	005
1.3 信息检索策略	010

第 2 章 计算机信息检索

2.1 计算机检索的基本方法	023
2.2 计算机检索主要技术	032
2.3 搜索引擎	037

第 3 章 中文文献检索平台及其数据库检索

3.1 中国知网数据资源平台	048
3.2 维普期刊数据整合服务平台	065
3.3 万方数据知识服务平台	073

第4章 中文电子图书数据库的检索

4.1 超星数字图书馆	081
4.2 书生之家数字图书馆	087
4.3 读秀检索平台	091
4.4 常见英文电子书数据库	099

第5章 常用外文检索平台及其数据库检索

5.1 EBSCOhost 平台	101
5.2 Elsevier ScienceDirect 平台	107
5.3 SpringerLink 平台	118
5.4 外文专业性全文数据库	123

第6章 特种文献数据库检索

6.1 会议文献检索	149
6.2 专利文献检索	161
6.3 学位论文检索	178
6.4 标准文献检索	182
6.5 科技报告检索	186

第7章 引文索引及其检索

7.1 引文索引概述	191
7.2 SCI 与 SSCI 简介	193
7.3 引文索引网络数据库的检索	196

第 8 章 文献资源管理、分析与利用

8.1 信息调研与资料的收集、鉴别与整理	211
8.2 个人信息资源管理	230
8.3 信息分析方法	241
8.4 信息调研报告的撰写	243

第 9 章 科技文献写作

9.1 科技文献概述	249
9.2 科技综述与科技述评的写作方法	250
9.3 学术论文的写作方法	256
9.4 学术论文投稿	272

参考文献	275
------------	-----

第1章 信息检索的基本理论

1.1 信息检索

人类已经进入信息时代，在生活、学习和工作中，人们无时不在感知、接受、理解、应用、传递和交流信息。信息的收集获取、分析综合、去伪存真、开发利用、增值创新已成为当代个人、组织乃至国家求生存、谋发展的不可或缺的能力和素质。正是在这种与信息互动中，人们加速创造着物质财富和精神财富。及时、准确、全面地获取信息是当代人的渴求。人们对信息检索的需求从未像现在这样迫切。

1.1.1 信息检索的概念

信息检索（Information Retrieval）起源于图书馆的参考咨询和文摘索引工作，从19世纪下半叶开始，至20世纪40年代，索引和检索已成图书馆独立的工具和服务项目。通俗地说，信息检索就是信息用户为处理各种问题而查找、识别、获取相关的事实、数据、知识的活动。

20世纪中叶以前，信息的存储和传播主要以纸作为介质，信息检索活动主要围绕着相关文献的获取和利用而展开，因此“文献检索”成为信息检索的同义词被更为广泛地使用。随着1946年世界上第一台电子计算机问

世, 计算机技术逐步走进信息检索领域, 并与信息检索理论紧密结合起来; 脱机批量情报检索系统、联机实时情报检索系统相继研制成功并商业化, 20 世纪 60 年代到 80 年代, 在信息处理技术、通信技术、计算机和数据库技术的推动下, 信息检索在教育、军事和商业等各领域高速发展, 得到了广泛的应用。Dialog 国际联机情报检索系统是这一时期的信息检索领域的代表, 至今仍是世界上最著名的系统之一。

信息检索作为一种实践活动由来已久, 但作为一个比较规范、正式的学术术语, 是由美国信息科学的先锋 Calvin Northrup Mooers 首先提出的。他认为“信息检索是一种时间性的通信形式”, “此种通信是时间性的, 在时间上从一个时刻通往较晚的时刻, 而在空间上可能还在同一地点”。也就是说, 通过信息检索得到了一些文献, 从而使得信息发布者与信息用户之间建立起了一种通信。这一观点旨在说明, 信息传递是一种特殊的通信形式, 正是这种通信促使了人类科学、文化、思想等领域的交流。

而目前我们所说的信息检索有广义和狭义之分。广义的信息检索全称为“信息存储与检索”, 是指将信息按一定的方式组织和存储起来, 并根据用户的需要找出有关信息的过程。即包括信息的“存”和“取”两方面。狭义的信息检索为“信息存储与检索”的后半部分, 通常称为“信息查找”或“信息搜索”, 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、信息检索的技术或方法、满足信息用户的需求。

1.1.2 信息检索的类型

1. 按存储与检索对象划分

按存储与检索对象划分, 信息检索可以分为文献检索 (Document Retrieval)、数据检索 (Data Retrieval) 和事实检索 (Fact Retrieval)。

(1) 文献检索: 文献检索是以文献作为检索对象, 查找含有用户所需信息内容的文献。文献检索是一种相关性检索而非确定性检索。系统不直

接解答用户所提出的问题本身,而是只提供与之相关的文献或文献的属性信息与来源指示供用户参考和取舍。所以其检索对象是包含特定信息的各类文献。

(2) 数据检索:数据检索是将经过选择、整理和评价(鉴定)的数据存入某种载体中,并根据用户需要从某种数据集合中检索出能回答问题的准确数据过程或技术。数据检索是一种确定性检索,即直接提供用户所需的确切的数据,而且检索结果一般也是确定性的。有些数据检索系统不仅能查出数据,还提供一定的运算和推导,用户获得各种经过整理、计算的量化信息,从而为定量分析提供依据。

(3) 事实检索:广义的事实检索既包括数值数据的检索、算术运算、比较和数学推导,也包括非数值数据(如事实、概念、思想、知识等)的检索、比较、演绎和逻辑推理。它要求检索系统不仅能够从数据(事实)集合中查出原来存入的数据或事实,还能够从已有的基本数据或事实中推导、演绎出新的数据或事实。事实检索是信息检索中最复杂的一种。它要求系统中的数据和事实以自然语言或接近于自然语言的方式存储。不仅要存入各种数据或事实单元,还要存入各单元之间的语义关系、句法关系以及各种有关的背景知识。允许用户用自然语言提问,并能用自然语言作答。更重要的是,系统必须具有一定的逻辑推理能力和自然语言理解功能。事实检索是一个相当复杂的过程,目前通常还是依靠人工来完成。

以上三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. 按存储介质和实现查找的技术手段划分

按存储的载体和实现查找的技术手段为标准,可将信息检索划分为手工检索、机械检索和计算机检索。

(1) 手工检索:手工检索是一种传统的检索方法,即以手工翻检的方式,利用工具书(包括图书、期刊、目录卡片等)来检索信息的一种检索手段。

手工检索不需要特殊的设备,用户根据所检索的对象,利用相关的检索工具就可进行。手工检索的优点是比较简单、灵活,容易掌握,并便于控制检索的准确性;缺点是费时、费力,特别是进行专题检索和回溯性检索时,需要翻检大量的检索工具反复查询,花费大量的人力和时间,而且很容易造成误检和漏检。

(2) 机械检索:机械检索是 20 世纪 50 年代开始的用各种机械装置进行情报检索的方式,是手工检索向计算机检索的过渡阶段。机械检索系统利用当时先进的机械装置改进了信息的存储和检索方式,通过控制机械动作,借助机械信息处理机的数据识别功能代替部分人脑,促进了信息检索的自动化。例如,现代情报学创始人,美国的 V. Bush 博士与美国农业图书馆馆员拉尔夫·肖共同制造了一台快速检索机——布什·肖检索机。它利用光电原理,对复制在胶卷上的文档进行检索,胶卷的边缘上用黑白点作为编码,当遇到检索内容时就停下来。

机械检索并没有发展出信息检索语言,只是采用单一的方法对固定的存储形式进行检索,而且过分依赖于设备,检索复杂,成本较高,检索效率和质量都不理想。

(3) 计算机检索:计算机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。按计算机检索的处理方式分类,有脱机检索和联机检索;按存储方式分类,有光盘检索和网络检索等。计算机检索具有检索方便快捷,检索功能强大,获得信息类型多,检索范围广泛等优点,日益成为信息检索的主流方式。

但计算机检索相对于手检系统浏览功能相对不足,因而不利于满足用户的潜在检索需求。同时,计算机检索系统多使用自然检索语言,缺乏词间关系控制,要达到理想的查全率和查准率,需要用户自己把握同义词、近义词、上下位词、相关词等,因而检索负担较重。另外,计算机检索数据库所覆盖的信息源的时间范围有限,所以对于时间跨度较大的回溯检索除使用计算机检索外,还需要辅之以手检系统。总之,计算机检索是检索

发展的趋势和主流,手检也因其特性,还有它存在的空间。

1.2 信息检索的基本原理

有了对信息检索的初步了解,本节将详细描述信息检索的基本原理和基本理论。

1.2.1 信息检索的基本原理

信息检索的基本原理是:通过对大量的、分散无序的信息进行搜集、加工、组织、存储,建立各种各样的检索系统,并通过一定的方法和手段使存储和检索这两个过程所采用的特征标识达成一致,以便有效地获得和利用信息源。其中存储是为了检索,而检索必须先进行存储。信息的存储和检索的全过程如图 1-1 所示。

信息存储过程包括:①文献信息采集,即根据一定的原则收集文献。②文献标引、著录,即对文献的信息特征,包括内容特征和形式(外部)特征进行揭示和描述。其中,对文献内容特征的揭示,是按照系统所采用的信息检索语言(分类表、主题词表等)对文献主题进行标引,为文献的内容特征加上标识。另外,根据需要,对文献形式特征中有检索意义的项目,如著者、文献题名等也可以做出标引,用作标识。③建立检索系统或编制检索工具。即按标引用语的顺序,将著录的大量文献款目有机地组织成一个排检系统,形成有序的、系统化的检索工具或数据库检索系统。

信息检索的前提是信息存储有序化。无论是手工检索工具、机械检索工具,还是计算机检索工具都会根据自身系统特性,在一定专业范围内进行信息收集和选择,对收集的信息进行分析、选择、标引、描述以及组织加工转化,形成信息数据库,以使用户检索使用。检索的过程是信息存储的逆过程,即用户对检索课题进行分析,形成检索提问信息,选取合适的

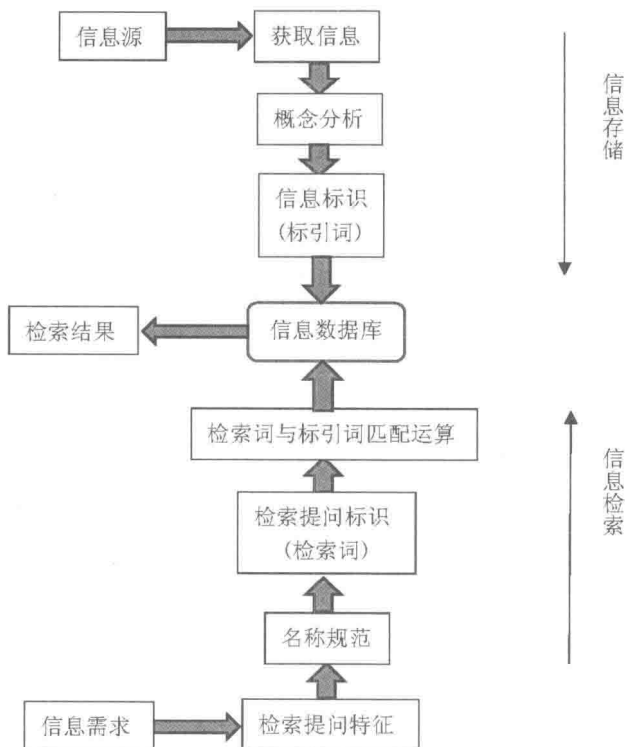


图 1-1 信息检索基本原理示意图

检索用语，利用检索工具或检索系统查出相关信息。简单地说，检索就是查找，查找的过程实际上是一个逻辑匹配的过程，即确定检索用语并将检索用语与标引用语作相符性比较，检索用语与标引用语一致，就算找到了符合要求的文献信息。

1.2.2 信息检索的效果评价

1. 检索效果的评价指标体系

检索效果指的是利用检索系统或工具开展检索服务时的有效性。它直

接反映着检索系统的性能,影响系统在信息市场上的竞争能力和用户的利益。根据 F.W.Lancaster 的阐述,判定一个检索系统的优劣,主要从质量、费用和时间三方面来衡量。质量标准主要通过查全率 (Pertinence Factor) 与查准率 (Omission Factor) 进行评价。费用标准即检索费用是指用户为检索课题所投入的费用。时间标准是指花费时间,包括检索准备时间、检索过程时间、获取文献时间等。查全率和查准率是判定检索效果的主要标准,而后两者相对来说要次要些。

查全率是指系统在进行某一检索时,检出的相关文献量与系统文献库中相关文献总量的比率,它反映该系统文献库中实有的相关文献量在多大程度上被检索出来。其数学表达式如下:

$$\text{查全率} = [\text{检出相关文献量} / \text{文献库内相关文献总量}] \times 100\%$$

查准率是指系统在进行某一检索时,检出的相关文献量与检出文献总量的比率,它反映每次从该系统文献库中实际检出的全部文献中有多少是相关的。其数学表达式如下:

$$\text{查准率} = [\text{检出相关文献量} / \text{检出文献总量}] \times 100\%$$

查全率和查准率与文献的存储和信息检索两个方面是直接相关的,也就是说,与系统的收录范围、索引语言、标引工作和检索工作等有着非常密切的关系。一般来说,查全率和查准率之间存在互逆关系,即当某一系统的查全率和查准率处于最佳比例关系时,继续提高查全率,检出的相关文献量会增加,但同时由于检出文献中不相关文献的增加会导致查准率降低;继续提高查准率,由于增加检索限定,就会造成查全率降低。

2. 影响查全率和查准率的主要因素

(1) 从存储角度看,影响查全率的因素主要有文献库收录文献不全;