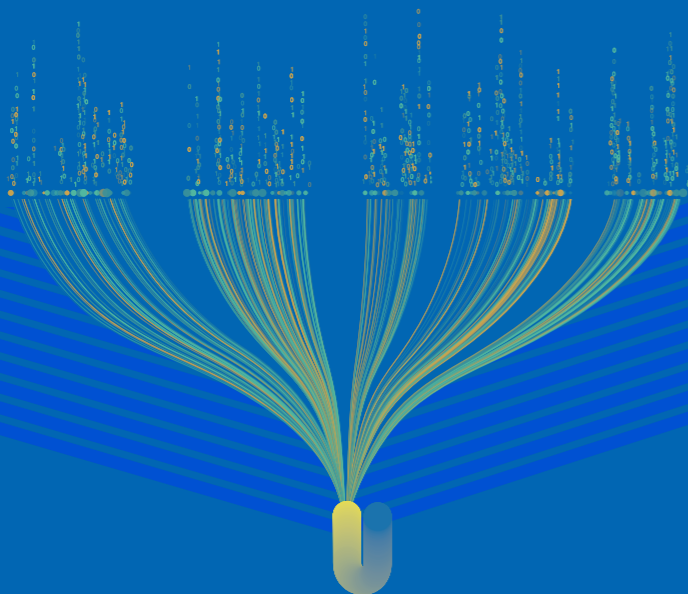


★ 服务热线: 400-615-1233
★ 配套精品教学资料包
★ www.huatengedu.com.cn

DAXUESHENG XINXI SUYANG
大学生信息素养



策划编辑: 王少松
责任编辑: 高宇
封面设计: 黄燕美



定价: 39.90元

高等职业教育公共基础课教材

大学生信息素养

主编 赵丽 李喆时 李杰

北京邮电大学出版社

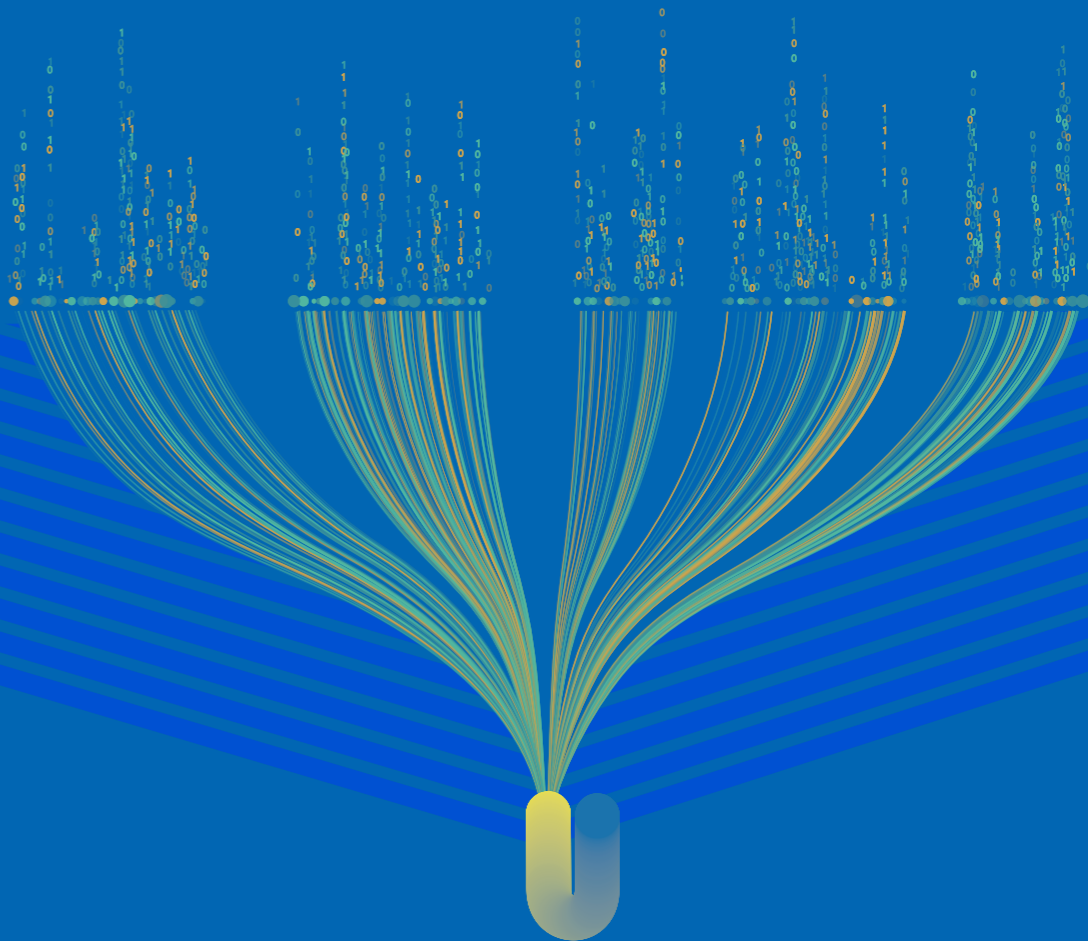


高等职业教育公共基础课教材

DAXUESHENG XINXI SUYANG

大学生信息素养

主编 赵丽 李喆时 李杰



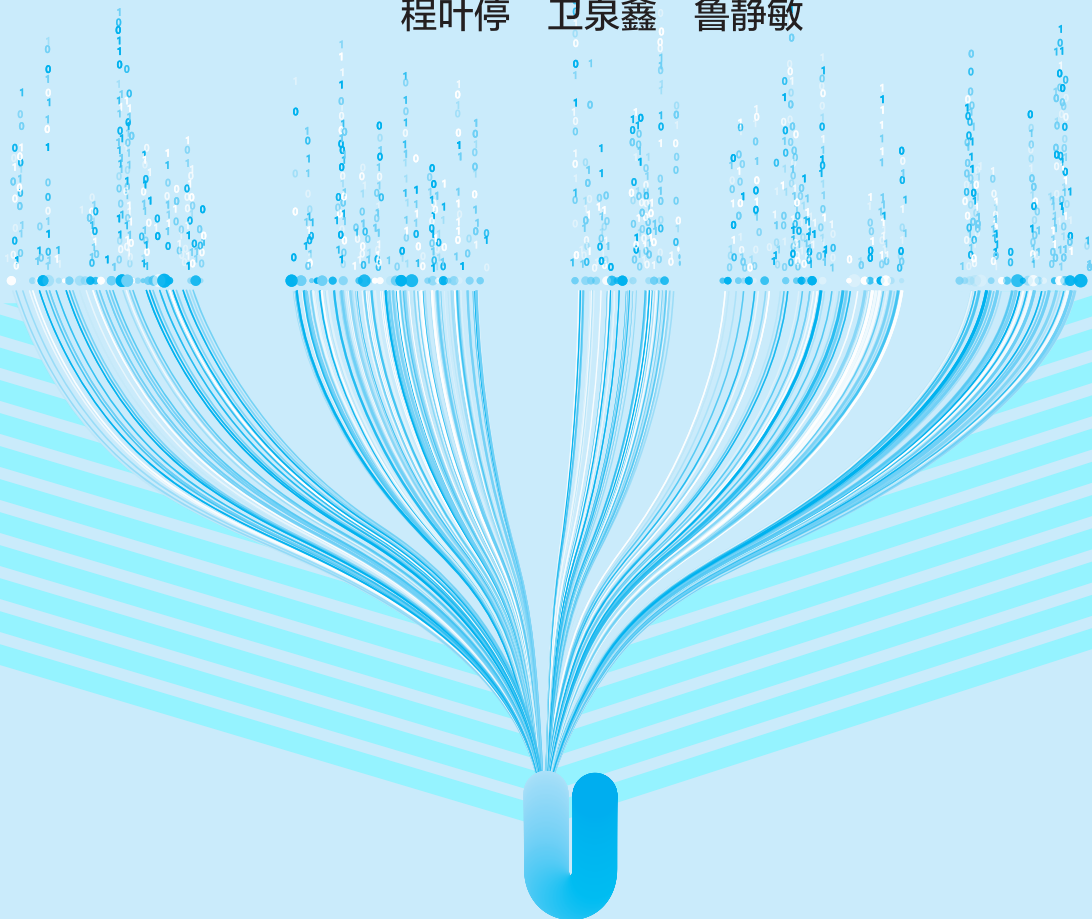
 北京邮电大学出版社
www.buptpress.com

高等职业教育公共基础课教材

DAXUESHENG XINXI SUYANG

大学生信息素养

主 编 赵 丽 李喆时 李 杰
副主编 段 杰 侯晓敏 程 川
程叶停 卫泉鑫 鲁静敏



北京邮电大学出版社
www.buptpress.com

内 容 简 介

内容简介:本书共分8章,包括信息素养基础知识、信息检索的基本原理与方法、信息检索工具、中英文数据库检索系统、电子图书与数字图书馆、计算机与互联网信息检索、特种文献检索、网络信息检索与利用中的有关问题。

本书可作为大学生信息素养教育教材,也适合大学生治学借鉴参考。

图书在版编目(CIP)数据

大学生信息素养 / 赵丽, 李喆时, 李杰主编. -- 北京: 北京邮电大学出版社, 2023. 11

ISBN 978-7-5635-6337-1

I. ①大… II. ①赵… ②李… ③李… III. ①大学生—信息素养—信息教育—研究 IV. ①G254.97

中国国家版本馆 CIP 数据核字(2023)第 225897 号

策划编辑: 王少松 责任编辑: 高 宇 封面设计: 刘文东

出版发行: 北京邮电大学出版社

社 址: 北京市海淀区西土城路 10 号

邮政编码: 100876

发 行 部: 电话: 010-62282185 传真: 010-62283578

E-mail: publish@bupt.edu.cn

经 销: 各地新华书店

印 刷: 三河市骏杰印刷有限公司

开 本: 787 mm×1 092 mm 1/16

印 张: 12.25

字 数: 253 千字

版 次: 2023 年 11 月第 1 版

印 次: 2023 年 11 月第 1 次印刷

ISBN 978-7-5635-

定 价: 39.90 元

· 如有印装质量问题, 请与北京邮电大学出版社发行部联系 ·

服务电话: 400-615-1233

前言

PREFACE

信息社会的到来引发了信息环境的巨大变化,面对信息社会万千变化的特征、机遇与挑战,信息素养渐渐走进了我们的视野,它区别于先天禀赋,是一个与时俱进的动态概念,包含了技术和人文两个层面的意义,既反映了人们利用信息的意识和能力,也反映了人们面对信息的修养。

信息素养既是知识经济时代大学生必备的基本素养,也是大学生成为创新型人才的关键素养和能力基础,是其终身学习的必备素养。如何让信息素养成为 21 世纪公民的基本素养,成为当代大学生成才与创业就业的能力基础,信息素养教育创新与改革尤为紧迫。

高校开设大学生信息素养课程是提升大学生信息素质、培养大学生适应信息社会发展的最直接、有效的途径。大学生信息素养是集理论与实践教学于一体的课程,学生学习的自主性对教学效果影响很大。我们根据课程改革和计算机检索的要求并结合不同学生的实际需要,组织具有丰富教学经验的教师编写了本书。本书在编写过程中,全面落实党的二十大精神,深入实施科教兴国战略、人才强国战略、创新驱动发展战略,秉承立德树人的教学理念,积极贯彻素质教育和创新教育的精神,突出当代大学生对文献信息利用能力的培养,还大量参阅了综合性、多科性和单科性文献检索的著作及有关论文,吸收了众多同类教材的优点。

本书的主要内容有信息素养基础知识、信息检索的基本原理与方法、信息检索工具、中外文数据库检索系统,电子图书与数字图书馆、计算机与互联网信息检索、特种文献检索、网络信息检索与利用中的有关问题等内容。本书融文理、古今、中外文检索工具书于一体,使广大在校大学生能较快地掌握多学科文献信息检索的技能和方法。另外,针对学生的学习特点在检索技巧上进行了设计,使所列实例具有较强的示范性和实用性。

本书由石家庄理工职业学院赵丽、李喆时和李杰任主编,石家庄理工职业学院段杰、侯晓敏、程川、程叶停、卫泉鑫、鲁静敏任副主编。

本书在编写过程中参考了大量国内外相关的著作、论文、教材和网站,并在书后列出了参考文献,在这里向这些文献的作者致以衷心的感谢。

限于编者水平有限,本书难免存在不足之处,恳请广大读者提出批评意见。

编者

目录

CONTENTS

知识篇

第 1 章

信息素养基础知识

1.1 信息概述	2
1.2 信息素养的概念	5
1.3 信息素养的标准	6
1.4 信息素养的内涵	9
1.5 信息素养教育	11

第 2 章

信息检索的基本原理与方法

2.1 信息检索语言	18
2.2 信息检索的原理、类型和手段	27
2.3 信息检索的研究范围与研究对象	30
2.4 信息检索方法与程序	31
2.5 信息检索策略	32
2.6 信息检索结果的评价方法	33

技能篇

第 3 章

信息检索工具

3.1 检索工具的含义	36
3.2 检索工具的类型	36
3.3 检索工具的排检方法	38
3.4 常用的中文检索工具	41
3.5 外文检索工具	49

第 4 章

中外文数据库检索系统

4.1 CALIS 数据库	55
4.2 CASHL 数据库	60
4.3 NSTL 数据库	65
4.4 万方数据资源系统	69
4.5 中国知网数据库	73
4.6 维普资讯数据库	78
4.7 外文数据库	84

第 5 章

电子图书与数字图书馆

5.1	电子图书	87
5.2	超星(汇雅)数字图书馆	89
5.3	书生之家数字图书馆	93
5.4	方正 Apabi 数字图书馆	95
5.5	中国国家数字图书馆	96

第 6 章

计算机与互联网信息检索

6.1	计算机信息检索的发展过程和网络基础知识	98
6.2	搜索引擎	101
6.3	计算机信息检索技术	105
6.4	电子信息资源	107
6.5	百度和谷歌	110
6.6	计算机检索过程与基本功能	115
6.7	网络信息检索策略与技巧	119

第 7 章

特种文献检索

7.1	学位论文及其检索	124
7.2	会议文献及其检索	131
7.3	科技报告及其检索	140
7.4	标准文献及其检索	144
7.5	专利文献及其检索	152
7.6	年鉴	160
7.7	政府出版物	162

素养篇

第 8 章

网络信息检索与利用中的有关问题

8.1	网络信息利用中的知识产权问题	166
8.2	信息法规与信息道德	173
8.3	信息利用规范	178

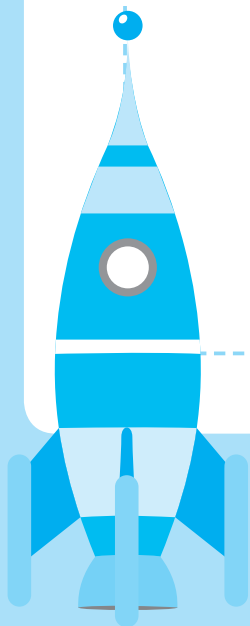
参考文献	189
------	-----



知 识 篇

第 1 章 信息素养基础知识

第 2 章 信息检索的基本原理与方法



第 1 章 信息素养基础知识



学习目标

1. 了解信息的定义、特征和类型。
2. 了解信息素养的概念和标准。
3. 掌握信息素养的内涵。
4. 了解信息素养教育的内涵与现状。

1.1 信息概述



当今,社会信息化已形成巨潮席卷整个人类社会。人类各种实践活动、各个领域都与信息和信息技术息息相关、密不可分,使得整个社会的生产方式、生活方式、活动方式、行为方式、思想观念、道德规范等方面都在发生根本变化。信息促使人类大脑在时空领域得到极大扩展,从而加速了世界经济的区域集团化、一体化的进程。

信息作为一种社会资源很早就被人们所认识,而且自古以来人类就在利用这种资源,只是利用的能力和水平很低而已。例如,我国古代发明的原始感测和显示技术的指南针、原始通信技术的烽火台,以及记录存储信息的造纸术和印刷术等。

1.1.1 信息的定义和特征

1. 信息的定义

当今世界,“信息”已成为社会生活中使用频率最高的词汇,与之相连的“信息经济”“信息产业”也成为人们身边的现实。信息作为重要的经济,战略资源和推动社会生产力的新动力,日益受到人们的青睐,“信息时代”“信息社会”正在向我们走来,信息化成为当代社会发展的大趋势。

“信息”一词被高频率地使用,但究竟什么是信息呢?怎样认识信息的科学意义呢?我们知道世界万物都有自己的运动规律和表现特征。社会生活中人们了解到的事物间的差异都是通过一定的形式表现出来的。例如,声音、图像、文字等,这些形式所表达的内容就是信



息。信息不是物质本身,而是物质的一种基本属性。

拉丁语中的“信息”一词是陈述、描述,概要的意思;英语中的“信息”是 information,有知识,资料,数据、情报,消息,报道、报告,通知等多种含义;汉语中的“信息”一词,“信”和“息”二字的意思相近,均有音信、消息的意思,但前者侧重于消息、征兆,后者强调情况、音信。

那么,信息到底是什么呢?目前,学术界还没有形成一个一致的、完整系统的定义。《辞海》中是这样定义的:信息是指客观的消息、情况、情报等,是指对消息接受者来说预先不知道的报道等。《新华字典》将信息表述为:信息是指音信、消息。信息论中是指用符号传输的报道,报道的内容是接受者预先不知道的,是事物的运动状态和关于事物运动状态的陈述。《韦伯斯特词典》将信息表述为:信息是知识和情报的通信和接受,是通过调查,研究或要求而得到的知识,是情报、消息、新闻、事实和数字等。我国科学家钱学森也对信息有过这样的阐述:信息是激活了、激化了的,是为了解决一个特定的问题所需要的知识。

综上所述,信息就是消息,人们在学习、工作、日常生活中随时随地都在接受和利用信息;信息是资源,它具有使用价值和价值;信息是财富,并且是无价之财富;信息是生产力要素,更是一种不可估量的促进生产力发展的新动力。虽然学术界难以形成统一的信息定义,但从文献信息探索的角度,我们应对信息做一个基本的约定:信息是自然界和人类社会中一切事物自身的运动状态与方式,是事物的内在联系和含义的表征,信息通过一定的载体形式反映事物的本质特征及其运动变化。也就是说,信息是一种可以用来消除不定性的东西,它泛指一切事物运动的状态和运动的方式,是关于事物运动状态和运动方式的反映。正因为信息是事物运动状态和运动方式的反映,所以它才能用来消除人们认识上相应的不定性,不断提高人类认识世界和改造世界的能力。因此,人类社会的发展、科学技术的进步,都离不开信息资源的开发和利用,而且信息资源的真实状况及开发利用程度,已成为衡量一个国家经济、文化、科技,乃至综合国力的重要指标。

2. 信息的特征

信息与物质、能量并称为人类社会的三大基本资源,但是与物质和能量不同,信息具有以下特征。

(1) 信息具有客观真实性。信息既不是物质,也不是能量,信息就是信息。不论你是否认识到,信息总是存在的。信息的这一最重要的本质特征是其生命所在。

(2) 信息具有共享性。信息的共享性是它与其他资源最大的不同之处。物质和能量被专人占有和使用,进而被消耗而不复存在。信息显然不同,当某一信息被多人共享时,信息的占有者不会丧失该信息,而且,一般来说也不会因为多次使用而改变信息的自身价值。作为一种共享性的资源,信息可以通过传输和扩散的方式实现共享,在交换、传递的过程中,信息不但不会丧失,而且还可能同时获取新的信息。因此,信息还可称为一种无限再生的资源。另外,我们在实际生活中应当注意区分信息的共享性与现实共享范围,了解独占信息、有价信息和公共信息的特性,合理合法地充分享有信息。

(3) 信息具有可存储性。信息的可存储性指的是信息必须依附于各种形式的载体,信息和物质(事物)、能量具有密切的关系,但它仍需依附于另一种形式的载体。所以说,没有物质,没有能量,就没有事物的状态和方式,也就没有信息,而且信息的识别、转换、扩充、压缩、生成、分享等所有的过程都必须有物质和能量的支持。信息通过物质载体来存储,从远



古的结绳记事、甲骨、泥板、纸张到磁、激光、大规模集成电路的广泛应用,大大丰富了信息的存储形式,也极大地增加了物质载体存储的信息量。

(4) 信息具有可转换性。信息可以被加工、整理、概括、归纳使之浓缩精练和不断扩充并变换各种形态,由原始一次信息变为二次信息、三次信息等,实现信息价值的增值。运用现代信息处理系统,将原始文件进行压缩,可达到便于传输和节省存储空间的目的。信息技术的快速发展可使信息进行声音、图形、文本、数据等多种形式的相互转换,有效地促进信息的传播和交流。例如,可将声音转换为文字或将文字转换为声音、将统计表转换为拼图等。

(5) 信息具有传输性。信息不同于物质,作为一种既无大小又无重量的非实体,它的传递速度仅取决于其承载体的性能。因此,计算机技术和光纤技术的高速发展使信息的传递超出了以往载体的限制。当信息的载体是电磁波和光波时,其传输的速度将达到每秒 30 万千米,而运用广播、电视、Internet 等大众传媒传递信息,可在瞬间将信息迅速、广泛、全方位地扩散。

(6) 信息具有极强的时效性。信息与物质、能量的价值观不同,信息有很强的时效性,过时的信息就是马后炮。例如,天气预报信息、军情与警情信息、战略预测信息等,如果不在最需要的时间内被准确地获取、传递、使用,它将分文不值。也就是说,信息的利用价值与时间成反比,及时有效地把握信息就能实现信息的最佳价值。

(7) 信息具有增值性。信息通过人脑思维或人工技术的综合、加工和处理,不断积累丰富,提高其质量和利用价值。信息交换的结果是信息的增值。

其实,信息的特征还有许多,如知识性、识别性、生成性、动态性等,这里限于篇幅,不再一一介绍。

1.1.2 信息的类型

信息类型繁多,从不同角度划分会有不同的信息类型。

(1) 依据信息存在的领域及信息的内容可划分为自然信息和社会信息。自然信息是指存在于人们主观意识之外,来源于自然界,在自然界中传递的信息。社会信息则来源于人类社会,是人类对于社会现象和社会活动的认识、反映,是在人际间传播的信息及思维信息。

(2) 依据信息的符号种类及信息的性质可划分为语言信息和非语言信息。语言信息是由作为信息沟通最基本,最重要的工具——语言来表现的信息,它由语法信息,语文信息和语用信息构成。人们总是以信息为媒介与事物打交道,通过语法信息获得语文,语用信息,经分析比较产生决策信息,根据决策信息使事物按人们预期的目标发展。非语言信息则是指表情、体态等非语言显现的信息。

(3) 依据信息的状态可划分为动态信息和静态信息。动态信息是指时效性极强的瞬息多变的的信息,如军情、警情、新闻等类信息。静态信息则是指呈现稳定、相对静止状态的信息,如历史文献、档案资料等类信息。

(4) 依据信息的加工处理程度可划分为一次信息、二次信息、三次信息。一次信息是在生产实践,社会实践和科学实验中产生的新成果、新知识和经验性总结等方面的信息;二次信息是依据客观需要,按照一定的科学方法,将特定分散的一次信息进行加工整理而形成的信息;三次信息是指依据二次信息提供的线索并选用一次信息的内容进行综合分析形成的信息。



(5) 依据信息的媒体和信息的载体可划分为单媒体、多媒体信息和文献及电子信息。用来表示信息的声音、图像、文字、数据,以及广播、电视等物质载体是信息的表示媒体。只有一种媒体表示的信息是单媒体信息,运用多种媒体综合、协调地表示的信息是多媒体信息。多媒体信息不同于多媒体技术和多媒体计算机系统的概念,而现代化的多媒体手段可以将多媒体信息带入一个崭新的应用领域。

(6) 依据信息的社会属性可划分为政治信息、经济信息、文化信息、科技信息、学术信息、法律信息、军事信息、市场信息等。

1.2 信息素养的概念



信息素养(information literacy, IL)最早使用在1974年,美国信息产业协会(Information Industry Association, IIA)主席保罗·泽考斯基(Paul Zurkowski)在提交给美国全国图书馆与情报科学委员会(National Commission on Libraries and Information Science, NCLIS)的报告中,他认为信息素养是人们利用大量的信息工具及主要信息源解决具体问题的技能。由此,信息素养概念得到广泛传播和使用,而且各个时期对信息素养概念的解释也不尽相同。

1989年美国图书馆协会(American Library Association, ALA)关于信息素养的总结报告中提出“具备信息素养的人,必须能够识别何时需要信息,知道如何查找、评估和有效利用所需要的信息,来解决实际问题或者做出决策”。

2003年9月,联合国教育、科学及文化组织(以下简称教科文组织)和美国全国图书馆与情报科学委员会(NCLIS)联合召开的信息素养专家会议,将信息素养定义为一种能够确定、查找、评估、组织和有效地生产、使用和交流信息,并解决面临的问题的能力。会议发表了《布拉格宣言:走向信息素养社会》,宣称信息素养是人们有效参与信息社会的一个先决条件,是终身学习的一种基本人权。

2005年11月,教科文组织召开信息素养和终身学习高层研讨会,发表了《亚历山大宣言》,提出信息素养和终身学习是信息社会的灯塔,照亮了信息社会发展、繁荣和走向自由的进程。信息素养是终身学习的核心,使人们在生活中能够有效地搜索、评估使用和创建信息,以实现个人教育和职业目标。

2015年2月,美国大学与研究图书馆协会(Association of College and Research Library, ACRL)通过的《高等教育信息素养框架》中指出,信息素养是反映发现信息、理解信息生产和价值、使用信息创造新知识和参与社群学习的综合能力的集合。

综上所述,信息素养是指个体在信息社会中,能够确定何时需要信息,依法合理有效地运用现代信息技术获取、评估、管理、利用、创建和传播交流信息,解决实际问题的能力与道德修养。信息素养是个人利用信息解决问题的综合能力与基本素质,是个人适应信息化社会和实现自我发展的关键素养。



1.3 信息素养的标准



1.3.1 国内的信息素养标准

我国在 1999 年的国家文件《中共中央国务院关于深化教育改革全面推进素质教育的决定》中规定培养学生信息素养的 6 个标准。

- (1) 对信息的关注。
- (2) 能够研究和判别可供选择的信息及观点的优劣、可行性。
- (3) 能够选择各种信息源,熟悉使用各种信息工具以获取信息的能力。
- (4) 能对某一课题找到相应的信息,进行综述及纵观其现状和发展趋势。
- (5) 获得继续自我教育的基础。
- (6) 发挥主动性和独立性。

2005 年,北京高校图书馆学会在参照美国大学与研究图书馆协会(ACRL)、澳大利亚大学图书馆委员会(CAUL)、英国国立图书馆和大学图书馆协会(SCONUL)的信息素养标准,研制了《北京地区高校信息素质能力指标体系》,共分 7 个维度,19 项标准,61 条具体指标项目。7 个维度分别如下。

(1) 具备信息素质的学生能够了解信息以及信息素质能力在现代社会中的作用、价值与力量。

(2) 具备信息素质的学生能够确定所需信息的性质与范围。

(3) 具备信息素质的学生能够有效地获取所需要的信息。

(4) 具备信息素质的学生能够正确地评价信息及其信息源,并且把选择的信息融入自身的知识体系中,重构新的知识体系。

(5) 具备信息素质的学生能够有效地管理、组织与交流信息。

(6) 具备信息素质的学生作为个人或群体的一员能够有效地利用信息来完成一项具体的任务。

(7) 具备信息素质的学生了解与信息检索、利用相关的法律、伦理和社会经济问题,能够合理、合法地检索和利用信息。

2008 年,图书馆工作委员会信息素养教育工作组组织专家在《北京地区高校信息素质能力指标体系》基础上修改,提出《高校大学生信息素养指标体系(讨论稿)》,由 6 项一级指标 17 项二级指标和 35 个知识点构成,其中一级指标减少了信息道德指标。

1.3.2 国外的信息素养标准

1. 美国的信息素养标准

1998 年,美国图书馆协会和教育传播协会制定了学生学习的九大信息素养标准,从信



息素养、独立学习和社会责任等三个方面进行了表述,进一步明确和丰富了信息素养在技能、态度、品德等方面的要求。

(1) 信息素养。包括以下3个标准。

- ① 能够有效地和高效地获取信息。
- ② 能够熟练地、批判性地评价信息。
- ③ 能够精确地、创造性地使用信息。

(2) 独立学习。包括以下3个标准。

- ① 能探求与个人兴趣有关的信息。
- ② 能欣赏作品和其他对信息进行创造性表达的内容。
- ③ 能力争在信息查询和知识创新中做得最好。

(3) 社会责任。包括以下3个标准。

- ① 能认识信息对民主化社会的重要性。
- ② 能履行与信息和技术相关的符合伦理道德的行为规范。
- ③ 能积极参与小组的活动来探求和创建信息。

2000年1月,美国大学与研究图书馆协会(ACRL)发布《高等教育信息素养能力标准》(Information Literacy Competency Standards for Higher Education),分为标准、执行指标和效果三个板块,有5个一级指标,22个二级指标和具有直接可测性的86个三级指标。5个一级指标如下。

(1) 具有信息素养能力的学生能决定所需的信息种类和程度。

(2) 具有信息素养能力的学生能有效地获取所需要的信息。

(3) 具有信息素养能力的学生能评判性地评价信息及其来源,并能把遴选的信息与原有的知识背景和评价系统结合起来。

(4) 具有信息素养能力的学生,无论是个体还是团体的一员,都能有效地利用信息达到某一特定的目的。

(5) 具有信息素养能力的学生懂得有关信息技术的使用所产生的经济、法律和社会问题,并能在获取和使用信息中遵守公德和法律。

2015年2月,美国大学与研究图书馆协会(ACRL)通过《高等教育信息素养框架》(Framework for Information Literacy for Higher Education),其中提出了基于信息素养的元素养(metaliteracy)概念,强调元认知(即批判性反省),以元素养的培养为核心。

2. 澳大利亚的信息素养标准

2000年10月,澳大利亚大学图书馆委员会(CAUL)批准了澳大利亚高等教育所使用的信息素质标准,它是在美国标准的基础上修改而成的,但又增加了2个一级指标:具有信息素质的人能够对收集与产生的信息进行分类、保存、管理和改写;能够认识到信息素质是终身学习和具有参与感的公民所必需的。

2001年,澳大利亚与新西兰高校信息素养联合工作组(Australian and New Zealand Institute for Information Literacy, ANZIIL)正式发布了《澳大利亚与新西兰信息素养框架:原则、标准及实践》。2004年颁布了澳大利亚与新西兰高校信息素质能力指标体系,其中6个一级指标、19个二级指标和67个三级指标。6个一级指标如下。



- (1) 能确认信息需要并决定所需信息种类和程度。
- (2) 能够迅速、准确地获取所需要的信息。
- (3) 能对搜索到的信息做出正确的判断。
- (4) 能管理所搜集到的或采集到的信息。
- (5) 能将初始的信息和新信息应用到构建新概念或创新知识中。
- (6) 使用信息时,必须了解和遵守使用信息的相关道德、法律、文化和社会问题。

3. 英国的信息素养标准

2011年,英国国立图书馆和大学图书馆协会发布了《信息素养的七大支柱》,指出信息素养涵盖数字素养、视觉及媒体素养、学术能力、信息处理、信息技能、数据监护及数据管理七个方面。

4. 日本的信息素养标准

2014年7月,日本国立大学图书馆协会和教育学习支援检讨特别委员会联合公布了日本的《高等教育信息素养标准》。该标准具体包括高等教育中信息素养应该掌握的知识、技能以及实践过程。接受信息素养培训的人,在以下阶段中通过以下过程进行信息的探索。

(1) 认识信息需求和所面临的课题。能够明确认识信息需求,对所需要的信息范围进行具体地确定。

- ① 正确认识自己所面临的课题。
- ② 认识对课题所必要的信息。
- ③ 认识当时所具有的信息。

(2) 计划对所需要的信息进行合适的、有效的探索。对所需要的信息从经济、法律、道德规范的角度出发,并且制定有效地进行探索的计划。

- ① 理解信息如何生成,如何传播。
- ② 通常所理解的能够得到的信息种类和特征。
- ③ 理解获取信息的方法以及能够得到的服务。
- ④ 理解查找信息时符合法律、道德规范以及经济的合理性。

(3) 准确、有效地获取信息。能够利用所掌握的获取信息的手段,准确、有效地获取所需要的信息。

- ① 有效地利用图书馆以及所提供的服务。
- ② 根据信息的种类,采用恰当的访问方法和检索工具获取信息。
- ③ 掌握获取信息必要的检索技能。
- ④ 正确理解信息,选择所需要的信息。

(4) 对所收集的信息进行评价、分析、整理和组织。能够评价和分析所需要的信息,使用恰当的工具处理和组织信息。

- ① 从可靠性、关联性、正确性等角度出发评价和分析所收集的信息和数据。
- ② 能够组织对所获取的信息进行有效的整理、表述和利用。

(5) 知识体系的重新构建。对所需要的信息能够批判性地整合到自身的知识体系当中,重新构建知识体系。

- ① 对所获取的新信息批判性地整合到自身的知识体系当中。



② 整合新的信息,重新构建自身的知识体系。

(6) 对所获取的信息进行有创造性的运用和发布。能够对所获取的信息在符合法律、道德规范的前提下进行有效且有创造性的利用、发布和交流。

① 掌握在利用信息时有关法律和道德规范的知识。

② 理解发布信息的对象以及交流对象。

③ 掌握发布信息的技能。

1.4 信息素养的内涵



信息素养是一种综合能力,包括能够认识到何时需要信息,能够有效地检索、评估和利用信息;能够将获取的信息与自己已有知识相融合,构建新的知识体系,解决所遇到的问题;能够了解利用信息所涉及的经济、法律和社会问题,合理、合法地获取和利用信息。信息素养是在信息化社会中个体成员所具有的各种信息品质,主要包括信息意识、信息知识、信息能力和信息道德四个要素。

1.4.1 信息意识

1. 信息意识的概念

信息意识即人的信息敏感程度,是人们对自然界和社会的各种现象、行为、理论观点等,从信息角度的理解、感受和评价。信息意识的强弱决定了获取、判断和利用信息能力的自觉程度。在日常生活中,处处都可能隐藏着有效信息,只要我们具有较强的信息意识,留心观察媒体信息和生活,就会有所收获。

2. 信息意识的主要内容

(1) 信息的价值。信息的价值主要是指信息主体对信息的作用、功能及其在社会中的价值应有的充分认识,即信息价值观的树立。

(2) 信息获取与传播意识。信息获取意识是指信息主体应具有主动寻求和发现信息的意识。信息传播意识是指将自己创造或获取的信息传播给他人的意识。

(3) 信息保密意识。特定信息需要某种程度的保密,如国家政治、经济、军事秘密、尖端科技成果、个人隐私等。

(4) 信息守法意识。信息内容的任意散播,有时不仅是道德问题,而且还可能涉及法律问题。

(5) 信息安全意识。防止有用的信息泄露、丢失或被更改的意识。如保管好个人的存折密码、上机密码等;对重要的信息应当备份;保护与自己相关的信息系统的安全,防止他人破坏。

(6) 信息动态变化意识。信息的价值不是一成不变的,旧的信息会不断地被新的信息所取代和超越。



(7) 信息的经济意识。信息是商品,可以用货币去购买信息,通过销售信息可以获取利润。它包含三方面的内容。

- ① 信息是可以卖的。
- ② 信息是可以买的。
- ③ 存在专门的信息服务企业从事信息买卖服务。

(8) 信息污染意识。信息在传播的过程中存在着错误的、虚假的、过时的、没有价值的信息。

1.4.2 信息知识

1. 信息知识的内涵

信息知识既是信息科学技术的理论基础,又是学习信息技术的基本要求。只有掌握信息技术的知识,才能更好地理解与应用它。信息知识是指对信息基本常识的了解,包括信息文化知识,各种信息源、信息检索工具、检索方法等方面的知识,计算机和网络技术常识以及信息法规知识。

2. 信息知识的主要内容

(1) 传统文化素养。传统文化素养包括读、写、算的能力。进入信息时代之后,读、写、算方式产生了巨大的变革,被赋予了新的含义,但传统的读、写、算能力仍然是人们文化素养的基础,而信息素养是传统文化素养的延伸和拓展。在信息时代,快速阅读的能力是有效地在各种各样的海量信息中获取有价值的信息的根本保障。

(2) 信息的基本知识。信息的基本知识包括信息的理论知识;对信息、信息化的性质、信息化社会及其对人类影响的认识和理解;信息的方法与原则,如信息分析综合法、系统整体优化法等。

(3) 现代信息技术知识。现代信息技术知识包括信息技术的原理、信息技术的作用、信息技术的发展及其未来等。

(4) 外语。信息社会是全球性的,在互联网上有 80% 的信息是英语,此外还有其他国家语种。要相互沟通,就要了解国外的信息,表达我们的思想观念,这就要求我们掌握 1~2 门外语,适应国际文化交流的需要。

1.4.3 信息能力

1. 信息能力的内涵

信息能力包括信息系统的基本操作能力,信息的采集、传输、加工处理和应用的能力,以及对信息系统与信息进行评价的能力等。这也是信息时代重要的生存能力。身处信息时代,如果只是具有强烈的信息意识和丰富的信息常识,而不具备较高的信息能力,还是无法有效地利用各种信息工具去搜集、获取、传递、加工、处理有价值的信息,不能提高学习效率和质量,无法适应信息时代对未来人才的要求。

2. 信息能力的主要内容

(1) 信息获取能力是指信息主体根据特定的目的和需求,从外界信息载体中提取所需



信息的能力,包括信息的查找能力和现场信息的搜集能力。

(2) 信息理解能力包括信息的识别与理解能力、评价判断能力和选择能力。信息主体应能够正确地识别与理解所遇到的信息的含义,知道它们反映了什么客观规律与现象。同时,能够正确地判断与估计所查找到的信息的价值与意义,并能在浩瀚的资源中选择自己需要的信息。

(3) 信息表达能力包括信息的生成能力和信息的表示能力。它要求我们能够在社会实践和社会调查数据中获取并生成有用的信息,并且能够运用正确的形式表达出来,以便于信息传播。

(4) 使用信息技术的能力是指能够使用计算机以及常用的软件系统处理工作、学习、生活等方面的问题。

1.4.4 信息道德

1. 信息道德的内涵

信息道德是指在信息的采集、加工、存储、传播和利用等信息活动的各个环节中,用来规范其间产生的各种社会关系的道德意识、道德规范和道德行为的总和。它通过社会舆论、传统习俗等,使人们形成一定的信念、价值观和习惯,从而使人们自觉地通过自己的判断规范自己的信息行为。

2. 信息道德的主要内容

(1) 在获取和使用信息过程中应遵循的伦理规范的总和。例如,不得危害社会或侵犯他人的合法权益,自觉保护他人的知识产权、隐私权等,不传递不良信息等。高尚的信息道德是正确信息行为的保证,信息道德关系到整个社会信息素养的发展方向。

(2) 信息道德作为信息管理的一种手段,与信息政策、信息法律有密切的关系,它们各自从不同的角度实现对信息及信息行为的规范和管理。

信息素养的四个要素共同构成一个不可分割的统一整体。信息意识是先导,信息知识是基础,信息能力是核心,信息道德是保证。

1.5 信息素养教育



信息素质教育是终身教育体系的有机组成部分,改变了传统的信息获取方式。通过信息素质教育,大学生可利用各种信息手段获取所需信息,对高校的人才培养起到了积极的作用,也成为衡量高校教育水平高低的标志之一。

从传统的信息素养教育“文献检索课”到信息社会“互联网+”时代基于新技术、新理念的信息素养教育,其内容和形式都在持续地推进、拓展和创新。“互联网+”时代,高校教学模式正在经历深刻变革。因此,随着时代发展,信息素养教育不断丰富延伸,今后还有更广阔的发展空间。



1.5.1 信息素养教育的内涵

信息素养教育是指为启发人的信息意识、提高人的信息能力和信息道德水平所进行的一系列社会教育和培训活动。简而言之,信息素养教育是以提高人的信息素养为中心的教育活动。

信息素养教育的目标可以分为狭义和广义两个层面。信息素养教育的狭义目标是培养学生的信息素养,也就是培养一种基于信息解决问题的能力,具体包括:信息需求的识别能力、信息获取能力、信息评价能力、信息管理能力、信息应用能力。信息素养教育的广义目标是培养学生的元素养和终身学习能力。就是能够想到而且能够做到通过获取和利用信息解决所遇到的问题,并且能够把所获取的信息融入自身的知识结构,把获取信息的过程和经历内化为自身的学习能力。

个体适应社会的发展以及个体自身的发展都需要不断地学习,即终身学习。终身学习主要强调不断获取新知识、新技能,提高使用外部环境的能力,这与信息素养的核心理念一致。信息素养是终身学习的重要基础,信息素养所强调的知识体系与能力体系重构实质上就是通过解决问题实现终身学习。

1.5.2 信息素养教育的现状

1. 美国的信息素养教育

20世纪60年代美国首先开展信息素养教育。1985年,美国政府各类相关组织开始关注和重视信息素养教育。由美国科学促进会(AAAS)1989年制订了《美国2061计划》,其中,将信息科学、计算机技术纳入科学教育体系中,把提高全民的“科学素养”作为科学教育的首要目标和解决教育问题的主要手段。1987年,美国大学与研究图书馆协会成立了信息素养总统委员会,该委员会由教育界和图书馆学界的领军人物组成,担负起3个使命。

- (1) 明确了信息素养对于学生的能力发展、终身学习和成为积极公民的重要性。
- (2) 针对正式或非正式的学习环境,设计一个或多个信息素养发展模型。
- (3) 明确继续教育和教师发展的内涵。

1989年,在美国大学与研究图书馆协会信息素养总统委员会倡导下,成立了美国国家信息素养论坛(National Forum on Information Literacy, NFIL)。在信息素养总统委员会《总结报告》中,号召国家和全体民众共同重视信息素养,强调国家的经济独立性和公民的生活质量越来越依赖于终身学习。2015年美国大学与研究图书馆协会正式颁布《高等教育信息素养框架》,开启审视信息素养的全新视野,即信息素养是发现信息、理解信息的价值、使用信息创造新的知识和参与社群学习的综合能力的集合。

总的来说,20世纪90年代以后,美国大学信息素养教育在教学内容和方法上都有了深入研究,而且在全美大学得到实施,逐渐成为美国大学素质教育的有机组成部分。同时美国大学图书馆在信息素养教育中的重要作用与地位也越发凸显出来,其中基于Web的在线信息素养教育已经逐渐成为美国大学图书馆信息素养教育的主要形式。在线信息素养教育主要具有开放性、形象性、交互性和个性化等特点。其中TILT(texas information literacy tutorial)是在得克萨斯大学系统数字图书馆的资助下,得克萨斯大学奥斯汀分校开发的信息



素养教育在线指南,其设计形式、内容的创新性和互动性已经得到了普遍认可。到目前为止,TILT可以说是广为使用、评价较好的美国在线信息素养教育指南之一。

从美国信息素养的发展来看,美国在信息素养教育方面更注重人文和社会因素,以信息获取和信息组织为基础,以社会道德、法律意识和创造性能力的培养为核心。

2. 英国的信息素养教育

信息素养教育在英国具有悠久的历史。1981年,在牛津召开了第二次国际会议,研讨各级各类图书馆的用户教育,将图书馆用户教育的发展推向新的高度。就信息素养教育这个体系内部来说,英国的信息素养教育在初等教育和中等教育中开展得较好。初等教育阶段就开设了信息教育课,并于1998年列为必修课,到高中阶段信息通信技术仍为必修课,对学习内容和要达到的目标制定了国家课程标准。为了更好地在高等教育中开展信息素养教育,1999年,英国高校与国家图书馆协会成立了一个特别工作组,专门研究了高等教育中的信息素养教育问题,最后形成了名为《高等教育信息技能意见书》的研究报告,并提出了信息素养教育中应培养的七个基本能力及信息素养的基本模式。2002年,英国联合信息系统委员会又在曼彻斯特城市大学图书馆和利兹大学图书馆的协助下开展了The Big Blue的研究项目,为英国的高等教育和16岁以上社会成员的信息素养教育提出了14条建议,并建立了自己的信息素养教育模式。

3. 日本的信息素养教育

日本的信息素养教育工作具有长期稳定性和连续性的特点。1985年,日本“回应信息化社会的初等、中等教育和各方调研协作会议”就提出了信息素养教育的必要性。日本文部省自1986年开始着手促进计算机在中小学的应用,地方教育当局负责教师训练的任务。1989年,日本教育部规定在小学和中学都要开展信息素养教育,并且利用计算机和多媒体改进教学,加强信息道德教育。

自1993年以来,日本中学的课程开始出现信息素养教育的内容,但分别出现在不同的科目中。在日本的高中阶段,职业学校才有信息科技科目,普通高中只有在数学课上教授些有限的相关知识,而高中生由于准备竞争激烈的高考,无暇顾及非考试科目的信息科目,使信息素养教育的连贯性在高中出现了断层。

1996年7月,日本中央教育审议会首次咨询报告《展望21世纪日本的教育发展趋势》,详细论述了信息化教育,将培养学生“信息综合能力”的必要性放在首位进行论述,并提议把国会教育会馆作为全国信息教育中心。随着网络远程教学越来越普遍,日本政府近年来实施了一项在基础教育领域有重大影响的“百所中小学联网”的试验研究项目。该试验项目力争让学生在全日本乃至全世界范围内进行广泛的信息交流,增强他们获取信息、分析信息和处理信息的能力,从而培养出有高度创造力的、能适应21世纪激烈的国际竞争的全新人才。

1998年日本图书馆协会出版了《图书馆利用教育准则》,同年京都大学开始在全校开设“信息探索入门”的基础课程,图书馆对该课程给予了大力支持。以此为契机,信息素养教育发展到了日本全国的各个大学。2000年以后,包括“信息素养”等课程在内的“信息相关课程”成为所有大学的必修课程。同年,日本各个高校在参照上述美国标准的基础上,结合实际制定了各自的信息素养能力标准,开展信息素养教育,为日本高等教育信息素养标准的制



定积累了丰富的经验。

2014年,日本制定出了《高等教育信息素养标准》。在该标准中对信息素养的概念又赋予了新的内涵,给出了如下较为完整的定义:所谓信息素养,就是当需要信息的时候,能够认识信息,有计划地收集、评价、整理、管理信息,以及利用信息并且有效地发布信息的能力。

4. 国际化的信息素质教育

随着信息技术和互联网技术的迅猛发展,信息素养教育受到世界各国重视和普遍开展,相关国际组织积极支持和推动信息素养教育,促进全球信息素养教育研究和推广。

1972年,教科文组织撰写的研究报告《学会生存》自始至终贯穿一条主线,教育应该扩展到整个人生。书中还指出:“未来的文盲,不再是不识字的人,而是没有学会怎样学习的人。”

2003年,教科文组织和美国全国图书馆与情报科学委员会、美国国家信息素养论坛联合发表了以“走向具有信息素养的社会”为主题的《布拉格宣言》,以推进信息素养的国际化。

2004~2005年,联合国有关信息素养的项目着重于培养具有信息素养的公民,鼓励在世界范围内进行有效的信息素养实践,支持信息素养创新课程,促进政府官员、研究学者、图书馆员与媒体专业人员之间的合作。

5. 我国的信息素养教育

我国的信息素养教育主要以高校为主,教育过程又以文献检索课为核心。从20世纪80年代到90年代初,国家出台了一系列关于文献检索课开设及管理的相关条例和通知,对文献检索课的课程建设、教学大纲、教学评估、内容标准等进行规范管理,并取得很好的效果。其中,1984年2月教育部印发《关于在高等学校开设文献检索与利用课的意见》的通知中明确指出,“各高等学校(包括社会科学和理、工、农、医各专业院校)应当积极创造条件,开设文献检索与利用课,有条件的学校可作为必修课,不具备条件的学校可作为选修课或先开设专题讲座,然后逐步发展、完善。研究生更应该补上这一课”。1992年,教育部在颁发的《文献检索课教学基本要求》中将信息意识和信息能力的培养列为文献检索课的教学目标。1998年,教育部颁布的《普通高等学校专业目录和专业介绍》中对每个专业的培养目标都提出了“掌握文献检索、资料查询的基本方法,具有独立获知识信息的能力”的具体要求。

进入21世纪,随着网络数据库、各类电子资源在我国各大高校大幅增长,文献由纸本向网络化、数字化发展。2002年1月,“全国高校信息素质教育学术研讨会”上将“文献检索课学术研讨会”改名为“信息素质教育学术研讨会”。之后,信息素养教育得到国内学术界的广泛关注和研究。

2010年以来,我国信息素养教育理论与实践探索紧密结合,积极应用现代教育技术,创新教学形式和内容,不断完善教学体系。将嵌入式教育、课程整合式教育、翻转课堂、信息共享空间和学科服务等引入教学实践。增加文献信息管理、信息分析、信息系统等教学内容,针对不同年级、专业学生开设相应的信息素养课程,逐步形成层级式信息素养教学体系。

2016年6月,世界教育创新峰会(WISE)与北京师范大学中国教育创新研究院在北京共同发布《面向未来:21世纪核心素养教育的全球经验》研究报告。报告指出,最受重视的公民七大素养分别是:沟通与合作、创造性与问题解决、信息素养、自我认识与自我调控、批



判性思维、学会学习与终身学习、公民责任与社会参与。

2018年4月13日教育部印发了《教育信息化2.0行动计划》，指出到2022年要基本实现信息化应用水平和师生信息素养普遍提高，推动从提升信息技术应用能力向全面提升信息素养转变。由此可见信息素养教育的重要性。

2023年2月，世界数字教育大会在北京开幕，会上发布了《中国智慧教育蓝皮书(2022)》与2022年中国智慧教育发展指数报告。报告显示，我国智慧教育基础设施设备环境基本建成，中国接入互联网的学校比例已接近100%；近年来，通过大力加强学生信息素养培育，近八成中小学生数字素养达到合格及以上水平；中小学教师数字素养全面提升，超过86%的教师数字素养达到合格及以上水平；学校管理信息化与网络安全制度建设完成度较高，已有近85%的学校具备网络安全管理制度。

1.5.3 信息素养教育的内容

信息素养教育的内容是丰富的、发展的。一方面，人们对信息素养的认识程度在不断深入和提高；另一方面，随着信息技术的广泛应用和飞速发展，信息量爆炸式增长，世界进入大数据时代，工业、农业、金融、交通、旅游、物流等行业进入信息化发展轨道。新兴的信息咨询和信息服务业异军突起，要求从业者具备更高的信息素养。因此，信息素养教育呈现职业化和多元化发展趋势。

信息素养教育可以分为多层次、多系列的专门教育。从幼儿园到大学，有不同的对象、不同的模式和不同的层次，也分为不同的等级和要求，如基础教育（包括基础的图书馆知识和基础的信息技术知识）、普通教育（以信息检索和利用为主的公共课程教育）、专业教育（信息管理与信息系统及其相关专业教育）。

信息素养教育可分为以下3个层次。

(1) 拓展学生的视野和知识面，了解各种各样的信息资源。在大数据时代，信息的内容与形式发生了变化，不同的信息内容与形式相互促进导致信息资源更加丰富和多变。以图书、期刊、学位论文、报纸、标准、专利、科技报告、百科全书、字词典、年鉴、地图为主的传统信息资源依然存在，但其存在方式、生产流程和载体形式已经发生了明显变化。大量新媒体不断涌现，如数字杂志、数字报纸、数字广播、手机短信、移动电视、网络、桌面视窗、数字电视、数字电影、触摸媒体等，创造出大量新的信息资源，借助微信公众号、微博、网络论坛、知识社区、网络文库、慕课(MOOC)等平台进行信息推送和传播。这些网络平台为终身学习提供了丰富的学习资源，有必要被所有的学习者知悉，因此在这个层面上信息素养教育就要让学生了解这些优质的网络信息资源。

(2) 提升学生的信息获取能力，让学生知道如何找到，并获取所需的信息资源。由于资源的形式、内容、来源不同，获取的方法和手段也有区别。检索方法与技巧是这个层面信息素养教育的重点。

(3) 培养学生的信息应用能力，使学生具有敏锐的信息意识和基于信息解决问题的综合能力，信息应用能力的培养更为复杂，是信息素养教育的重点和难点。

另外，信息素养教育还应包括信息价值意识教育、信息获取与传播意识教育、信息道德与法治意识教育、信息安全与保密意识教育。



1.5.4 信息检索与信息素养教育

掌握信息检索的基础知识和基本技能,对于提高信息素养具有非常重要的意义,具体表现在以下方面。

1. 有利于创新人才的培养

当今社会正处于知识经济时代,科学技术飞速发展,新知识层出不穷。社会对人才的需求也已发生巨变。智力劳动者的社会需求日益紧缺。培养高素质,高智商的创新人才,已是目前我国高校教育迫在眉睫的重大使命。所谓创新人才,必然是发掘和捕捉最新科技知识信息的行家里手,因为无数创新成果表明:谁能优先吸收、获取和利用最新科技知识信息,谁就能优先占领科技发展前沿阵地,谁就能在竞争中立于不败之地。为此,开设本课程就是培养学生捕捉最新科技知识信息的情报意识,学习和掌握获取最新科技知识信息的技能,为科技创新锻炼过硬的文献信息检索本领。

2. 有利于开发、利用文献信息(智力)资源

在当今的信息化社会里,文献信息迅猛膨胀,数量惊人,类型繁杂,新旧知识更新加快,人们处于浩如烟海的文献信息当中,迅速获取广博精准的文献信息越来越困难。学习和掌握文献信息检索的科学技术和方法,即可全面、迅速、精准地获取文献信息,为进入丰盛的文献信息的智能宝库提供了金钥匙。人们利用文献信息检索技术和方法,可以遨游文献信息大海。在这个知识经济的时代里,谁能优先掌握信息资源,谁就能在竞争中站在科学发展的前沿阵地。因此,我们必须要有被时代淘汰的紧迫感,树立强烈的信息意识,积极主动地获取所需文献信息。

3. 有利于继承、借鉴和创新

继承、借鉴和创新是社会发展的客观规律。科学技术的发展、人类物质和精神文明的进步也不例外。正如牛顿所说:“如果说我看得远一点,是因为我站在巨人的肩膀上。”这就是继承、创新规律的生动描述。因此,我们从事学习、科研和生产实践,必须在前人或他人的理论、技术和科研成果的基础上,深入理解知识、学精技术、吸取研究的经验教训和发现尚待解决的问题等,只有这样才能制定出可行性创新方案,对已学知识提高认识,使技术更加精湛,研究的起点更高。尤其在当今知识经济时代,继承、借鉴和创新的作用更加突出。例如,一项新的科技成果中,前人、他人的成果占90%,而创新的成果只占10%。为此,只有学习和掌握过硬的文献信息检索技能,才能在学习、科研时迅速准确地发现和了解前人和他人的研究成果。显而易见,只有强化文献信息意识和能力修养,才能有效地继承、借鉴和创新。

4. 有利于节时、省力和高效

学习、科研、工作都离不开查阅有关文献信息资料,它占据整个学习、研究的大量时间。美、日等国的调查结果表明,一个科研周期的各项工作环节所占时间为:计划思考占7.7%,查阅文献占50.9%,研究实验占32.1%,撰写论文占9.3%。

可见,为了提高查阅效率,必须选用最科学、最实用的文献信息检索方法,缩短时间,节省大量的精力,以便更好地学习。开设此课程的目的就是培养学生科学的检索技术,尤其是当今信息技术已引入文献信息检索领域,不仅要学习和掌握手工检索技能、技巧,而且要学



习和掌握计算机检索技能、技巧,以实现省时、省力、高效。

5. 有利于终生知识更新

世界进入知识经济时代以来,知识半衰期迅速缩短,新旧知识代谢加快。大学生迈入社会,走向新的岗位,如不及时更新知识,适应社会发展的新形势,将会成为新时代的知识盲人,落后于时代,必将被淘汰。有关专家估测:科技人员已具有的科技知识,12%是在大学获取的,87.5%是在工作岗位上通过学习与实践积累的。大学生在新的工作岗位上如果不更新知识,3~5年内已学的知识50%陈旧失效,8~10年再不更新知识,原有的知识将失效80%。为此,从长远出发,就要提高独立自主的学习能力、文献信息检索与利用的能力和不断更新知识的能力。

6. 有利于改善我们的生活质量

在日常生活中,小到买一本书,大到出国留学或买房买车这样的大额消费,如果掌握了信息检索的基本技能,就能及时准确地获取与生活有关的社会与市场信息,使我们的生活不仅更加丰富多彩,而且生活成本也会降低,从而为改善生活质量提供了极大的可能性。

7. 有利于缓解信息压力,有效解决信息污染

无处不在的信息虽然给我们的生活、学习、工作带来了许多便利,但不可否认,也带来了巨大的信息压力,尤其是无处不在的“垃圾”信息,往往使我们无所适从,干扰我们的思维,影响我们做出正确决策。掌握正确的信息检索技术,可以有效化解信息污染带来的风险,缓解巨大的信息量带来的精神压力。总之,学习信息检索知识是信息素养教育的重要内容,是培养高素质人才的基本要求。

第2章 信息检索的基本原理与方法



学习目标

1. 了解信息检索语言。
2. 掌握信息检索的原理、类型和手段。
3. 了解信息检索的研究范围和研究对象。
4. 了解信息检索程序、策略和效果评价。

2.1 信息检索语言



2.1.1 检索语言在信息检索中的作用

检索语言是根据检索的需要而发明的一种专门的人工语言,是在文献检索领域中用来描述文献特征和表达信息检索提问的一种专用语言。

信息检索是信息传递过程中的一个重要环节。随着科学技术的飞速发展,科技文献量急剧增长,内容专深且互相渗透,而科研、生产、教学及行政部门对情报的需求十分迫切,尤其是在检索课题较专深的情况下,信息检索总是变得相当复杂,为了达到较满意的检索效果,就必须做出很大的努力。

检索语言与检索效果之间有着密切的关系,它在检索过程中起着极其重要的作用。信息检索的全过程包括信息的存储过程和检索过程。当存储信息时,文献标引人员首先要对各种文献进行主题分析,通过分析选出若干能代表文献主题的概念,并用信息检索语言把这些概念标引出来,然后纳入信息检索系统中。当检索信息时,信息检索人员也首先要对检索课题进行分析,并且通过分析明确其检索范围,选出若干能代表信息需要的概念,并把这些概念转换成信息检索语言,然后从信息检索系统中查找用该检索语言标引的文献,从而找到用户所需的信息。

由此可见,信息检索语言是信息检索系统的主要组成部分,是标引人员与检索人员之间沟通思想、取得一致意见的桥梁。



2.1.2 信息检索语言的组成、特点和功能

1. 信息检索语言的组成

实质上信息检索语言是表达、概括文献信息内容的概念及相互关系的概念标识系统。它可以是从自然语言中精选出来并加以规范化的一套分类号,也可以是代表某类事物的某一方面特征的一套代码(如化合物的各种代码),用以对文献内容和信息需要进行主题标引、逻辑分类或特征描述。

信息检索语言是由词汇和语法组成的。在这里词汇是指登录在分类表、词表中的全部标识,一个标识(分类号、检索词、代码)就是它的一个语词,而分类表及词表则是它的词典;语法是指如何创造和运用那些标识(单个标识或几个标识的组合)来正确表达文献内容和信息需要,以有效地实现信息检索的一整套规则。

2. 信息检索语言的特点

检索语言不同于自然语言,它所表述的概念只有一种解释,不允许一词多义、多词一义,而使概念的表述模棱两可。检索语言的这种单一性保证了表述概念的唯一性和标引与检索的一致性,从而使信息检索人员全面、准确、快速地检索到含有所需信息的文献。

3. 信息检索语言的功能

通过检索语言标引文献的主题概念,不仅能简明地提示文献包含的信息内容及其外表特征,而且还能将同一主题概念的文献集中在一起,使文献的存储集中化、系统化、组织化,便于进行有规律的检索。

2.1.3 检索语言的分类

目前世界上使用的检索语言有一两千种,但各种信息检索语言的基本原理是一致的,只是它们在表达各种概念及其相互关系和解决对它们提出的共同要求时所采用的方法不同,因而形成了不同的类型和语种。

(1) 信息检索语言按其构成原理可分为分类检索语言、主题检索语言和代码检索语言三大类。

(2) 信息检索语言按其包括的学科或专业范围可分为综合性语言和专业性语言。

(3) 信息检索语言按其适用范围可分为一种检索工具或检索系统专用的语言,某一图书、信息机构专用的语言,各类图书信息机构通用的语言。此外,信息检索语言可分为手工检索工具适用的语言和机器检索系统适用的语言。

(4) 信息检索语言按语种可分为单语种语言和多语种语言。

(5) 信息检索语言按其标识的组合使用方法可分为先组式语言和后组式语言。

上述5种划分方法中,按构成原理划分的方法是最基本、最常用的方法,以下详细说明用这种方法划分的各种信息检索语言。

(1) 分类检索语言是用分类号来表达各种概念,将各种概念按学科性质进行分类和系统排列。分类检索语言包括体系分类语言(体系分类法)和组配分类语言(组配分类法)。体



系分类法主要应用概念划分与概括的方法。组配分类法主要应用概念分析与综合的方法。实际上,无论是前者还是后者,都采用概念划分与概括的方法建立等级体系结构,又采用概念分析与综合的方法进行组配。

(2) 主题检索语言是用语词来表达各种概念,将各种概念按字顺排列。主题检索语言包括标题词型检索语言(标题法)、单元词型检索语言(单元词法)、叙词型检索语言(叙词法)和关键词型检索语言(关键词法)等。它们可统称为主题法系统。

(3) 代码检索语言一般只就事物的某一方面特征,用某种代码系统来加以标引和排列。

下面仅就几种常用语言进行介绍。

1. 体系分类语言

(1) 体系分类语言的定义。体系分类语言是一种直接体现知识分类的等级结构的标识系统。它根据一定的观点,以科学分类为基础,以文献内容的科学性质为对象,运用概念划分与概括的方法,按照知识门类的逻辑次序,从一般到具体,从简单到复杂,进行层层划分,每划分一次就产生许多类目,逐级划分就产生许多不同级别的类目。

所有不同级别的类目,层层隶属,形成一个严格有序直线性的知识门类的等级制体系。每个类目都用分类号作为标记,每个分类号都是表达特定知识概念的词汇,这些词汇即是体系分类语言。

(2) 体系分类语言的特点。

① 体系分类法具有按学科或专业集中地、系统地提示文献信息内容的功能,有利于从学科或专业角度进行全面的检索。

② 体系分类法按结构逐级划分,具有等级结构,便于扩检和缩检。

③ 体系分类法用分类号作为检索标识,不能直接表达概念;检索时,首先要将检索的主题内容转换为分类号,然后进行检索。若不熟悉分类法,转换时就可能发生误差,但用分类号检索,可不受文种的限制。

④ 分类法不宜经常修订,列举在分类表中的类目不能随时改变,因而分类法不能及时反映新的科学技术。遇到检索主题属于新兴学科或主题狭窄的课题时,有可能找不到相应的类目和分类号。

⑤ 分类法是按直线或序列编制类目,遇到跨学科的课题时,与课题相关的文献只能标引在一门学科的类目之下,因此检索时可能漏检。

⑥ 体系分类语言是一种先组式检索语言,因此,不能按检索课题的需要进行组配检索。体系分类语言除了以上特点以外,还有一个独特的优越功能,就是组织图书资料的排架,这是其他检索语言不可比拟的。

2. 标题词语言

(1) 标题词语言的定义。标题词语言是规范化了的自然语言。它以经过标准化处理的名词术语作为标识,来直接表达文献所论及或涉及的事物之间的相互关系(这种关系是借助于参照系统来间接显示的)。

(2) 标题词语言的结构及类型。



① 标题词语言的结构。标题词语言的三大基本构成要素是主标题、副标题和说明语。

- 主标题。这是表达文献核心内容的主题概念,一般由经过控制的自然语言中的词、词组和短语来充当。

- 副标题。它是从一特定方面对主标题进行说明、限定,并经过规范化的词、词组进行限定,说明主标题所表达的文献主题的某个部分、某一应用方面、研究对象和方法、性质、场所以及文献类型。

- 说明语。它是用来详细说明和限定主标题、副标题的内容与方面的词、词组或短语。它由未经控制的自然语言表达,但不纳入词表正文,而是排在主、副标题之下。

② 标题词语言的类型。标题词语言是按事物来集中相关文献的。因而一个标题下的副标题就可能涉及相当于分类表中多个类目的范畴。一般都将某些标题下的内容进行细分。细分方法多种多样,相应地形成了不同类型的标题形式,主要包括以下几种。

- 单级标题。仅由一个名词术语构成的标题,它可以是一个单词,称单词标题或元词标题;也可以是一个词组,称词组标题或复词标题,但都只有一级,故称单级标题。这种标题系统的优点是比较简单,缺点是专指性差,在一个专题下往往集中资料太多,不便于检索。

- 带说明语的单级标题。例如,“起重机,用于建筑施工的”“期刊,建筑工程的”等,这实际相当于一个复词标题。由于用说明语表达复杂概念比较自由,既能表达事物的特征,也能表达事物的方面,因而可以达到较高的专指度。但其缺点也是显而易见的,那就是比较冗长,排列次序不够明确。

- 多级标题。即在标题下再加标题。汉语中常用破折号将两级标题分开,破折号后的标题称为子标题,用以表示该标题所表达事物的某一“方面”。子标题之下还可以有子标题,称为次子标题;次子标题下还可以有子标题,称为再次子标题。不过,实际应用中超过三级的多级标题形式很少见,绝大多数是采用二级或三级标题,各级子标题可统称为副标题,第一级标题则称为主标题。

- 倒置标题。将复词标题中起限定作用的词倒置于被限定词之后,例如。

钢结构桥梁	桥梁,钢结构
木结构桥梁	桥梁,木结构
水泥结构桥梁	桥梁,水泥结构
石结构桥梁	桥梁,石结构

由此可见,倒置排列的结果可以将某一方面的资料集中在一起,这是其主要优点,但它不能单独构成标题系统。

- 带限定词的标题。例如,“建筑学(辞典)”“古建筑(维修)”“苹果(水果)”等。限定词一般都置于括号中,用于做形式复分,或区分两个字面相同但无法用其他方式区分的标题。显而易见,这种标题形式也不能单独构成标题系统。

- 混合标题。同时采用上述几种标题形式中的若干种构成。混合标题系统的优点是能提高检索效率,查找比较方便,但其结构复杂,编制难度较大。

(3) 标题词的书写格式。标题词若是一个词组,则有以下3种书写格式。



① 正序式。按词组的正常书写顺序书写。

② 倒序式。把词组的正常书写顺序颠倒过来,并加逗号表示。利用倒序式可将某些同类事物集中显示其种属关系,例如,“COMPUTERS, ANALOG”(模拟计算机)和“COMPUTERS, DIGITAL”(数字计算机)。

③ 并列式。用“与”的形式把两种概念组合在一起,如“METAL AND ALLOYS”。

一般来说,标题词间的相互关系,即它们所表达的各种概念之间的相互关系,可分为等同关系、等级关系、相关关系 3 种类型。我们可以用“见”和“见自”这一对参照来显示等同关系,用“参见”和“参见自”这一对参照来显示等级关系和相关关系。具体说明如下。

- 等同关系一般有标题词的不同写法、同义词、学名与俗称、新称与旧称、全称与简称、不同译名、其他等同词等。在每组等同关系的几个词(词组)中,都只能选用其中一个作为“标题词”,其余的则作为“非标题词”保留在标题系统中,并用“见”和“见自”这一对参照来标引它们与标题词之间的关系,例如。

电脑 见 电子计算机

这样一来,进行检索时,就不一定要判断清楚哪一个词是被采用的,哪一个词是不被采用的。可以从所想到的任何一个词或任何一种标题形式入手检索。如果入手检索的词或标题形式正好是不被采用的,则“见”参照就会把检索人引到正式被采用的标题词中。

- 等级关系一般包括属种关系、学科与其分支的关系、整体与部分的关系等。

等级关系实际上是广义词(泛指词)与狭义词(专指词)之间的关系。广义词是上位词,即上级标题词;狭义词是下位词,即下级标题词。在标题法中,一般只做上级标题词“参见”下级标题词的参照,而不做下级标题词“参见”上级标题词的参照。例如。

电机	参见	电动机	发电机
海洋学	参见	海洋地质学	
消化系统	参见	肠胃	

- 相关关系一般有交错关系,矛盾关系,因果关系,某种原理、方法、材料、设备与其某个应用方面的关系等。由于具有相关关系的两个标题词下的资料是可以互相参考、互相补充的,所以应用相互“参见”,例如。

图书馆学	参见	情报学
情报学	参见	图书馆学

(4) 标题词表。标题词表是把标题词按一定方式组织与展示的词汇表。它由编制说明、主表、副表 3 部分组成。

下面以美国“工程索引”使用的《工程标题词表》(简称 SHE)为例加以说明。《工程标题词表》是《工程索引》(EI)文摘条目编制和检索的依据(1993 年前),EI 的月刊和年刊都是按此标题表编排的。此表由美国工程情报公司编制。

SHE 由主标题词、副标题词、“见”或“参见”项及说明语 4 部分组成。词表的后面有 5 个附录,它们是:附录 1,副标题索引;附录 2,此部分附录中包括计算机辅助分类代码,主题指南和机构索引;附录 3,缩略语表;附录 4,希腊字母及发音;附录 5,美英不同拼写对照。

SHE 的正文是按主标题词的字母顺序排列的。主标题词全文大写黑体印刷,其下含数量不等的若干副标题词,副标题词的第一个字母用大写。



主、副标题词共计 14 000 多个。其具体结构示意图如图 2-1 所示。

BULLDINGS ^①	402, (404) ^②
(Use for general subject ,applying code 404 if for civil defense or military purposes . Use geographical location as subheading . If significant, use heading for specific type of building . For specific types represented by main headings, see BULLDINGS/REFFRECE LIST ^③)	
Acoustics	See ARCHITECTURAL ACOUSTICS
Air Supported Structures	See ^④ INFLATABLESTRUCTURES
Aluminum Applications ^⑤	415
Arctic	
Brick	
Ceilings	
Climate Control	(Beginning 01/83)
Cion Systems	(See also ^⑥ Signal Systems)
Concrete	718

说明:①标题词;②分类代码;③说明语;④“见”项;⑤副标题词;⑥“参见”项。

图 2-1 主、副标题具体结构示意图

SHE 是用计算机编制的,可提供计算机检索磁带,词表中标题词右侧的 3 位数字,便是该标题相对应的计算机分类代码。

“见”(see)项可指引读者选用更切题的标题词,“参见”(see also)项用以指引读者进一步选用含义相关的其他标题词,以扩大检索范围。

说明语位于主标题词或副标题词下,置于括号内,用斜体字排印,用于说明该标题词的定义、适用范围以及使用方法等。

在 SHE 的副标题索引中,每个副标题词的右边分别标有一个符号,它们是关于该副标题的简略范围注释,具体符号及其含义如下。

- S 表示它可以作为普通副标题使用,而且是直接从主标题转过来的。
- # 表示它只能做副标题使用。
- @ 表示它与主标题词的形式不同。
- + 表示它是属于地理位置、材料或船只的名称。

SHE 是查找 EI 的主题索引和相同的辅助索引的工具,通过 SHE,可以准确了解 EI 采用的标题。

(5) 标题词语言的特点。

- ① 标题词语言标引文献,是以事物为中心,从有关该事物出发便于进行特性检索。
- ② 标题词在词表中按字顺排列,无字级性。
- ③ 标题词语言只能通过参照系统进行扩检和缩检,因此,查全率较低。

3. 关键词语言

(1) 关键词语言的定义。关键词语言是为适应主题目录、主题索引编制自动化的需要而产生的一种主题语言型检索语言,是文献量迅猛增长、信息传播速度加快的必然产物。

关键词是指在文献的标题、摘要或正文中出现的,对表达文献内容具有实质意义,能作



为检索入口的,起关键性描述作用的词汇。

关键词语言,就是针对文献中的关键词选定或抽出,不做或极少做词汇控制,按字顺排列,从而提供检索途径的一种检索语言。

(2) 关键词语言的特点。关键词语言可实现检索工具编制过程的计算机化,因而降低对编目人员知识和业务水平的要求,降低标引难度,节约人力。利用计算机自动标引文献,编制关键词索引非常迅速,从而大大缩短检索刊物的报道时差。由于这些优点,关键词语言在国外被普遍应用。关键词语言直接使用自然语言,因此用它作为信息检索语言,质量较差。

4. 单元词语言

单元词又称为元词,它是从文献中抽取出来并经过控制处理的,能表达文献主题的最小、最基本的词汇单位。它可以是一个单纯词,也可以是一个合成词。这些词具有一个共同的特点:它们在概念上不能再进一步分解,如再分解,就再也不能表达原来所代表的特定概念,从而失去检索本意。

单元词语言就是经过词汇控制的,能表达文献主题的元词集合。

元词的灵活组配性使它处理文献的能力很强,但在检索中也产生了相当的误检,因此,这种语言现在很少用,它被叙词语言所代替。

5. 叙词语言

(1) 叙词语言的定义。叙词是经过词汇控制的,在标引中用来显示文献主题,在检索中用来构成提问式的一种检索词汇。

叙词这一术语是由美国人穆尔斯在 20 世纪 40 年代末提出的。叙词语言是以受控的自然语言为语词基础,以字顺和分类系统为词汇显示的基本手段,以语词的概念组配为重要特征的一种标引和检索文献的理论方法。

叙词语言采用科技术语作为标识文献主题概念的标目。叙词语言可谓是博采各种信息检索语言之长,吸取了多种信息检索语言的原理和方法。叙词语言是一种采用规范化的单词或词组,由标引人员或读者自行组配,来表达文献(或课题)主题概念的一种后组式索引语言,或者叫做后组式的检索语言。

(2) 叙词的组配原则。叙词所用的词类主要是名词和动名词。叙词的词序都采用自然序列,尽量采用单词作为叙词,利用单词的组配来表达复合概念,使叙词语言的优越性能(多途径检索和自行扩检、缩检)得以发挥。直接采用专指性的词组是受到适当控制的,只是在用单词组配会造成错误和混乱时才采用词组。语词的组配方式有概念组配和字面组配两种。叙词的组配是概念的组配,应当严守概念组配的原则,只有使之符合概念逻辑,才具有检索意义。

概念组配有以下 3 种。

① 概念相交。时钟和收音机是同级词,谁也不属于谁,组配后得时钟收音机。两个或多个同级词组配后,由于不同概念相交,就会形成一个新概念,这个新概念是组配前各概念的下位概念。若用 A 和 B 表示两个不同概念,则概念相交可用图 2-2(a)表示,其中阴影部分表示组配后形成的新概念。

② 概念并列。两个或多个同级词组配后,由于不同概念并列,也会形成一个新概念,例



如,长篇小说和历史小说这两个概念并列后,得到长篇历史小说。和概念相交不同,概念并列后的新概念是组配前各概念的上位概念,如图 2-2(b)所示。

③ 概念限定。刀具和热处理可以组配成刀具热处理。刀具和热处理不是同级词,刀具是事物,热处理是该事物的某一方面,称方面词,它对事物起限定作用。因此将表示事物的叙词和它的方面词组配后,由于方面词概念的限定作用,就会形成一个新概念,它能说明事物的某一方面。概念限定也可用图 2-2(a)表示。

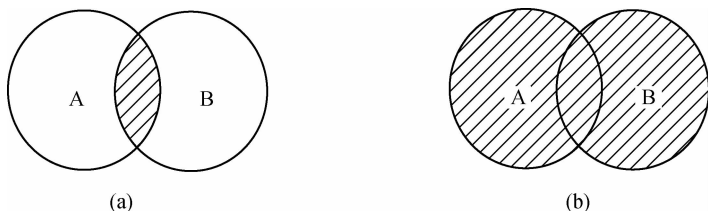


图 2-2 概念组配示意图

在各方面概念组配中,常见的是第一、第三种,特别是概念限定。因此,叙词经过概念组配后,往往能够表达更加复杂的概念,从而提高检索标识的专指性[把标识限制在图 2-2(a)中的阴影部分],以满足特性检索的需要,提高查准率。

(3) 叙词表。叙词语言所对应的叙词表是在吸收了各种词表优点的基础上发展起来的具有现代意义的词表。按其选词的学科范围,可以分为专业性叙词表和综合性叙词表。我国的《汉语主题词表》就是一部综合性叙词表。《冶金专业叙词手册》《机械工程主题表》等属于专业性词表。

叙词表一般都由一个主表和几个辅表组成。主表是叙词表的主体,包括的内容最全,对每个叙词的著录也比较完备,所以可独立存在。辅表则是为方便主表使用而编制的各种辅助索引,不能单独使用。

① 主表。主表即叙词字顺表,包括全部的叙词和非叙词;叙词条目下列有与该词有关的所有上位词、下位词、被代替词和相关词,从而可以详细了解到这类词的组成情况,同时也可以得到相关文献的“参见”项,以便扩大检索,提高查全率。

② 辅表。辅表包括叙词等级索引、叙词分类索引、叙词轮排索引。叙词等级索引按概念的等级关系排列,即将有上下从属关系的词按等级集中在一起,这样的一组词称为词族,最上位的一级词称为族首词,全部款目按族首词的字顺排列。叙词分类索引是按学科类别划分叙词,然后按类汇集,同类下的叙词按字顺排列,因此从学科角度可以查到该类的全部叙词,也可以从已知的类号转换成相应的叙词,以便从主题途径检索文献。叙词轮排索引就是将叙词中的复合词,按其包含的单词做轮番排列。一个叙词词组由几个单词构成,便可轮排几次,在几处都能查到它。

以下是由《工程索引》(EI)的《工程叙词表》中摘录下来的主表的著录格式及几种受控主题词表中的参照符号对照。

《工程叙词表》的主表是由 9 400 个叙词和 9 000 多个非叙词按词的字顺排列而成,在叙词款目下列出一系列参照符号,在非叙词款目下列出 EI 的叙词。其著录格式如下。

叙词款目:Acoustic imaging



SN: Production of read-time images of the internal structure of opaque objects

DT: Ptrdates 1975

UF: Acoustic lenses

BT: Imaging techniques

NT: Ultrasonic imaging

RT: Acoustic holography

Acoustics

Diagnosis

Lenses

非叙词款目: Acoustic Lenses

Use: Acoustic imaging

其中一些符号的含义见表 2-1。

表 2-1 《工程叙词表》中符号的含义

符 号	含 义	说 明
Use	用	非正式叙词转为正式叙词
UF	代	只用于正式叙词,指出正式叙词所代替的非正式叙词
NT	分	指出所列叙词下分得更专指的词,这些词均为正式叙词
BT	属	指出所列叙词的上位概念的词
SN	注释	对所列叙词的词义进行说明
DT	日期	所列叙词起用时间
RT	参见	指出与所列叙词相关的叙词

(4) 叙词语言的特点。叙词语言是在体系分类语言、组面分类语言、标题词语言、单元词语言和关键词语言的基础上,扬长避短而发展起来的一种检索语言,它具有很多优异的性能。

① 叙词语言是后组式检索语言。概念组配是叙词语言的核心。叙词经过自由的、广泛的组配后,能形成概念网罗性高的检索标识,表达主题所含的复杂概念,从而能进行特性检索,提高查准率;能形成若干检索标识,索引深度高,从而可以进行多途径检索。

② 叙词表的等级结构,便于增删、修订叙词。科学技术有了新的进展,增添相应的叙词后,就能标引和检索这些新科技的文献。

③ 叙词语言直接用叙词作为检索标识,不必转换,具有语言的直接性。

④ 叙词语言有一套较完整的参照系统,能显示叙词之间的相互关系,可供扩检或缩检用。

⑤ 利用叙词的词族索引和范畴索引,还可以按学科分类进行族性检索。

总之,与现有其他检索语言比较,叙词语言是一种多功能的较完善的检索语言。科技人员都熟悉自己的专业,从叙词表中找到切题的叙词后,只要组配得当,就能大大提高检索效率。



2.2 信息检索的原理、类型和手段



2.2.1 信息检索的原理与类型

1. 信息检索的原理

人类的信息检索行为总是从特定的信息需求开始,并在特定环境和信息检索系统中完成,这里所说的环境包括产生需求的环境、信息检索系统的运行环境和其他制约因素。特定的检索系统包括完成检索过程所需的一定设施和工具,它可以是图书馆、信息中心或信息经济人,也可以是某种工具书(如文摘索引、目录、资料集、手册、词典等)或机读信息源(如各种机读数据库)。

广义的信息检索包括信息的存储和检索两个过程。信息的存储就是将搜索到的一次信息,经过著录其特征(如题名、著者、主题词、分类号等)而形成款目,并将这些款目组织起来成为二次信息的过程。信息的检索是针对已存储好的二次信息库进行的,是存储的逆过程。存储是为了检索,而为了快速而有效的检索,就必须存储。没有存储,检索就无从谈起。这是存储与检索相辅相成、相互依存的辩证关系。

然而,由于职业、知识水平、个人素质,甚至习惯等因素的差异,信息存储人员(标引者)与信息检索用户(检索者)对同一信息的分析、理解也会存在不同。例如,《计算机在生物化学中的应用》一文,标引者可能将其归为“生物化学”类,而检索者则可能在“计算机”类查找该文。这样,标引者与检索者之间发生了标引错位,存储的信息就无法检索到。

怎样才能保证信息能存储且能取出呢?那就是存储与检索所依据的规则必须一致,也就是说,标引者与检索者必须遵守相同的标引规则。

信息存储与检索共同遵循的规则称为信息检索语言。只要标引者和检索者用同一种检索语言来标引要存入的信息特征和要查找的检索提问,使它们变成一致的标识形式,信息的存储过程与检索过程就具备了相符性。相应地,存入的文献也就可以通过信息检索工具(系统)检索出来。如果检索失败了,就要分析检索提问是否确切地描述了待查课题的主题概念,在利用检索语言标引时是否出了差错,从而导致检索提问标识错误。只有检索提问标识和信息特征标识一致时,相关的文献才能被检索出来。

信息检索正是以信息的存储与检索之间的相符性为基础的,如图 2-3 所示。如果两个过程不能相符,信息检索就失去了基础。检索不到所需的信息,存储也就失去了意义。

2. 信息检索的类型

掌握了信息检索的原理,就有必要了解信息检索的类型。按照不同的标准,信息检索可划分为不同的类型,其特点也各自不同。

(1) 按照检索内容划分。

① 文献信息检索。凡是利用目录、文摘或索引等二次信息查找某一课题、某一著者、某



一地域、某一结构、某一事物的有关信息以及这些信息的出处和收藏单位等,都属于文献信息检索范畴。其检索的结果是文献信息。例如,“设计人行天桥的参考文献有哪些?”便属于该类命题。

② 数据信息检索。凡是利用参考工具书、数据库等信息检索工具检索包含在文献中的某一数据、参数、公式或化学分子式等,都统称为数据信息检索。其检索结果为数据信息。例如,“某一新型载货汽车的载重量是多少?百公里油耗是多少?”就属于数据检索。

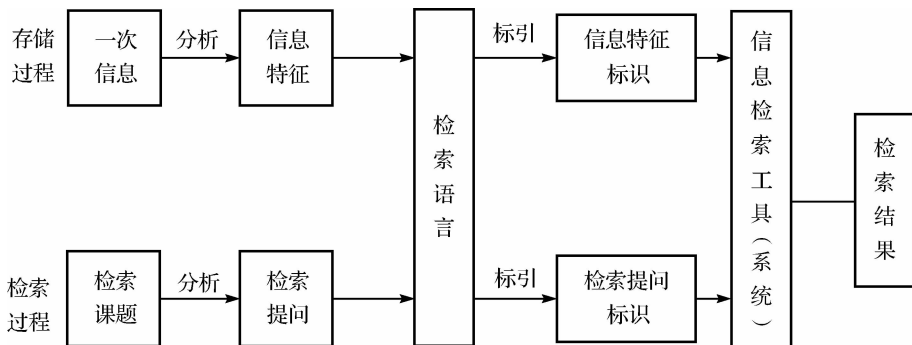


图 2-3 信息检索基本原理示意图

③ 事实信息检索。凡是利用百科全书等信息检索工具从存储事实的信息系统中查找出特定事实的过程,都称为事实信息检索。其检索结果是基本事实。例如,“世界上最长的斜拉桥是哪座?”“该桥位于什么地方?”“何时建成?”等。

文献信息检索是一种相关性的检索,检索的结果是文献线索,还必须进一步查找才能检索到有关的一次信息;数据与事实信息检索是一种确定性检索,检索的结果是可供用户直接利用的信息。一般情况下,文献信息检索通过二次信息来实现,而数据与事实信息检索则通过 3 次信息来完成。

(2) 按照信息存储和检索的方式和技术划分(详见 2.2.2)。

① 手工检索。手工检索是指人们通过手工的方式来存储和检索信息。

② 计算机检索。计算机检索是指人们利用数据库、计算机软件技术、计算机网络以及通信系统进行的信息存储和检索,其检索过程是在人机协同作用下完成的。计算机会从其存储的大量数据中自动分拣出与用户提问相匹配的信息,用户则是整个检索方案的设计者和操纵者。其检索的本质没有发生变化,发生变化的只是信息的载体形式、检索手段、存储方式和匹配方法。

2.2.2 信息检索的手段

根据信息检索的基本原理,实现信息检索的基本方式可分为传统信息检索和现代信息检索。传统信息检索简称手检;现代信息检索简称机检。

1. 传统信息检索

传统信息检索是检索人员利用手工检索工具手翻、眼看、大脑思维判别,索取原始文献的一种方式。

(1) 传统信息检索的优点如下。



- ① 检索条件简单,成本低。
- ② 在检索过程中可以随时获取反馈信息,及时调整检索策略。
- ③ 可对不同的检索工具同时进行对比,从而提高检索质量。
- ④ 可以参阅检索工具中的附图。

(2) 传统信息检索的缺点如下。

- ① 速度慢、效率低,检出的文献款目必须抄录。
- ② 手工检索工具提供的检索点有限,很难进行多元检索。
- ③ 难以找到涉及几个概念组合的多主题的文献。

(3) 手检的方法有直接检索和间接检索两种。

① 直接检索。直接检索是常用的检索方法,即利用一次文献进行检索。人们常常在需要信息资料时,直接到图书馆查阅各种图书、期刊及其他资料,从中找到所需的信息。

② 间接检索。间接检索是利用各种检索工具(目录、题录、文摘)查找信息的方法。这是文献检索最常用的方法。

2. 现代信息检索

现代信息检索是检索人员利用计算机检索系统查找文献的一种检索方式。计算机检索系统包括数据库技术、计算机技术和通信技术等。机检可以克服手检的缺点,但机检对设备条件的要求比较高,所需的投资比较大。计算机检索已从单机检索、联机检索发展到今天的网络检索,并正向智能化的方向发展。

(1) 单机检索。单机检索是计算机检索的最初形式,随着计算机存储介质的发展变化,也在不断发生变化。目前主要有计算机磁盘检索、计算机磁带检索和计算机光盘检索3种形式。

(2) 网络检索。网络检索是近十几年发展起来的,它目前主要有图书馆局域网的检索、联机检索及 Internet 信息检索3种形式。

3. 传统信息检索与现代信息检索的比较

现代信息检索是在传统信息检索的基础上发展起来的,因此两者的检索原理相同,但在网上的信息检索技术与以往的计算机检索技术又有很大的差别,主要有以下几方面的异同点。

(1) 检索语言方面。检索语言就是在检索信息时使用的专门的语言词汇,手工检索与计算机检索使用的检索语言基本相同,不同的是手工检索使用规范语言较多,而计算机检索则可以使用较多的自然语言,Internet 检索一般使用自然语言。

(2) 概念的组配方面。手工检索无检索运算符,检索时各种概念之间用人脑进行组配;用计算机检索时,则是利用各种逻辑运算符进行概念的组配;网络检索概念组配方式还不成熟,目前一般只用布尔逻辑运算符进行组配。其中计算机检索的概念组配最严谨,手工检索的概念组配最灵活。

(3) 检索入口方面。检索入口就是手工检索工具中的检索途径,是根据信息的某种特征编制的索引进行的检索,如题名、著者等。手工检索的检索入口较少,而计算机检索和网络检索的检索入口较多而且非常灵活,计算机检索过程中一个信息的所有特征都能作为检索入口,如年代、语种、文献类型、各种代号等,检索时还可以多途径同时进行检索。

(4) 检索结果方面。手工检索的检索结果查准率最高,误检率最低,而查全率较低。计



计算机检索的结果查准率要低于手工检索,误检率要高于手工检索,查全率则较高。网络检索结果误检率和查全率也较高。

2.3 信息检索的研究范围与研究对象



信息检索作为一门学科,有它自己的研究范围和对象,也有自己的理论、方法和技术。从总体上看,信息检索的研究对象是比较明确具体的,研究范围广泛而边界有些模糊,理论和方法已经逐渐形成体系。

2.3.1 信息检索的研究范围

信息检索的研究范围包括一切与信息存储与检索有关的系统、过程、理论和方法。一切可供存储和检索利用的信息类型,如文献、数据、事实、知识、声音、图形等;各种信息检索系统及其运行过程,如信息采集、标引、组织、存储、处理、匹配、传送等各种过程中使用的方法;以及在信息检索实践和研究的基础上形成的各种理论和假设,均包括在这个范围内。信息类型侧重于文献,其次是数据和事实。

2.3.2 信息检索的研究对象

信息检索的研究对象主要有以下几方面。

(1) 信息检索理论。它主要包括检索语言与标引理论、信息检索的数学模型、知识表示理论、相关性理论,以及有关的哲学问题。

(2) 信息检索系统。它是实现信息检索的物质基础,是现实的研究对象,主要研究其结构、功能、演变,其设计开发技术、管理维护技术和评价技术,还研究它与其他信息系统乃至整个外部世界的关系。其中,数据库是信息检索系统的核心部分之一。数据库的建造和维护是非常重要的信息技术。数据库的质量直接影响信息检索系统的功能和效率。

(3) 计算机信息检索。它涉及许多计算机设备、软件技术、存储技术革新、检索技术、系统管理和经营知识、市场营销技术等,是一门综合性很强的技术。

(4) 检索策略与方法。它是用户从信息检索系统中获取有关信息所必需的。好的检索策略是检索成功的必要前提,计算机的应用为检索策略和方法的发展提供了有力的支持。近30年来,涌现了许多新颖而有效的检索技术和方法,如布尔检索、位置检索、截词检索、加权检索、聚类检索等,人们利用、研究、评价和完善现有的各种检索策略和方法,研究开发新的更有效的策略和方法。

(5) 用户研究与培训。用户是信息检索系统的生命,无论是系统的研制开发、管理维护、功能和服务的扩展,还是系统评价,都离不开用户研究工作。用户培训是用户研究的继续,是与用户建立紧密联系和发展新用户的一种非常有效的措施。此外,还有自动标引、自动分类和自动摘录,以及相关设备等研究领域。



2.4 信息检索方法与程序



信息检索是根据课题的要求,利用各种有关的检索工具,按照一定的检索程序和方法,采用各种检索策略查找信息的过程。

2.4.1 信息检索方法

常用的检索方法如下。

1. 追溯法

追溯法是指在没有检索工具的情况下,人们利用书、刊之后的引用或参考文献查找资料的方法。

2. 常用法

常用法是指利用各种类型检索工具,通过技巧查找资料的方法。常用法又可分为顺查和倒查两种。

(1) 顺查法是由远而近的查找方法。如已知某研究课题的起始年代,需要了解它的全发展情况时,可从最初研究的年代开始,由远及近地进行查找。

(2) 倒查法也称为逆查法。和顺查法相反,倒查法是由近及远的查找方法,这种方法常用于正在研究课题的过程中。

3. 循环法

循环法又称为分段法,是前两种方法的结合,即首先利用检索工具查找一部分资料以后,为了了解其前因后果,再利用这部分资料的引用文继续追查的方法,利用这种方法能够保证所查的资料较全面、系统。

2.4.2 信息检索程序

查找信息的过程如下。

1. 分析课题

信息检索之际,首先要对课题进行充分的了解,掌握课题有关的基本知识,详细摸清课题所需资料范围及要求,以及弄清课题有关的名词、术语等。

2. 选择检索方法

根据检索人员已掌握的资料情况,确定采用何种检索方法。如果是新立的课题,一般采用顺查法和追溯法结合查找;正在进行中的课题,一般采用倒查法。

3. 确定检索途径

在检索过程中,根据所掌握的有关课题的资料,确定检索途径。一般情况下,所研究的课题较专深、单一时,常用主题途径;课题所涉及的内容较复杂时,常用分类法。



4. 查找信息线索

在确定检索途径之后,根据检索标识,通过有关索引进行信息检索,查找信息线索。

5. 索取原始信息

通过检索获得信息线索后,确认有一定的参考价值,需要详细查阅信息的全文时,即可根据文摘指出的信息出处,向有关部门索取原始信息。

2.5 信息检索策略



信息检索总是根据一定的课题进行,但并不是有关该课题内容的全部信息都需要。要想在数量庞大、类型复杂、内容专深、文种多样的信息海洋中迅速、准确地查找到满足要求的信息,必须制定查找信息的策略。

下面仅介绍手工检索策略的制定方法。

1. 先国内后国外

首先利用国内检索工具或用中文编辑的报道外文的检索工具入手检索。这样,检索者从自己熟悉的文字可以很容易地掌握课题的概况,并可以从中选择出切题的关键词、主题词,为进一步查找外文检索工具提供方便。

2. 检索时多选几个同义词、近义词

信息语言复杂多变,作者使用的语词不尽相同,为了避免漏检,检索时尽量多选几个同义词、近义词作为检索词。

3. 巧妙地利用上、下位词的关系

当要求查全率高而不计查准率时,可利用上、下位词进行检索。

4. 尽量避免从字面出发选择检索词

在检索时,尽量从内容上选词,以避免有关问题因选词不当被漏检。

5. 变换检索词的词序

某些课题,在检索中反复查找而没有结论时,应考虑倒置词序的可能。检索词序的准确与否是检索成败的关键。

6. 利用参见系统,扩大检索范围

检索工具中的参见(see also)项,把内容有关的主题联为一体,以便族性检索。例如,CA中每一大类末尾均列有与该类有关的全部相关大类和具体相关文摘。绝大多数检索工具中的主题词表、主题索引或文摘部分中均没有参见项,检索时分别查找各有关的标题,以获取某一课题的全部信息。

7. 各种索引配合使用

要想把各种索引配合使用,首先熟悉各种索引的结构、特点、款目内容及其功能。例如,



CA 中的化学物质索引与分子式索引、登记号索引、环系索引、杂原子索引关系十分密切。如果从化合物登记号索引入手检索,只能获得化合物的名称与分子式而查不出文摘号,必须转查化学物质索引才能获得信息线索。又例如,对商品名称与俗称,必须首先查索引指南。类似的情况在检索时常常遇到,只有熟练掌握各种检索工具的结构体例、编排特征、各索引款目结构及其相互关系,检索起来才能得心应手,取得满意的效果。

2.6 信息检索结果的评价方法



利用检索工具或检索系统查找信息时,检索结果并不是在任何情况下都能如愿以偿的。其主要原因是:从目前的检索系统看,它只能处理信息标识,而信息的标识只能表示信息的中心主题,不表示信息的次要主题。即系统只能回答信息提问而不一定能满足信息需求。从检索者来看,一般只能大概地表述自己的信息提问,而不能确切地衡量信息与提问的切题程度。再加上,检索课题的性质、范围、深度不断变化,查找信息的方法尚未形成固定模式,补充词表的出版不及时,人们不可避免地受到知识面的局限等,致使信息标引与检索失误在所难免。

在目前情况下,对检索结果的评价方法有以下几种。

1. 查全率

查全率是对所需信息被检出程度的量度。

2. 查准率

查准率是衡量检索系统拒绝非相关信息的能力。

查全率和查准率均可以用公式来表示。

$$\text{查全率}(R) = \frac{a}{a+c} \times 100\%$$

$$\text{查准率}(P) = \frac{a}{a+b} \times 100\%$$

式中, a 表示被检出的相关信息, b 表示被检出的非相关信息, c 表示未检出的相关信息。

除此之外, $c/(a+c) \times 100\%$ 为漏检率,是查全率的补数; $b/(a+b) \times 100\%$ 为误检率,是查准率的补数。

一般来说,查准率与查全率之间存在反变关系。当某一课题的查全率与查准率处于某一比例关系时,继续提高查全率,查准率会降低。

3. 相关性

相关性是用户判断信息与信息需求之间关系的标准。信息系统回答的不是用户提问本身,而是检索式表达的信息提问。因此,检出的是与信息提问相关的信息,但不一定是对用户切题的信息。