



高等学校人工智能通识课程融合系列规划教材

医学信息与人工智能 素养教程

主编 张鸿来 顾金媛 胡月



YIXUE XINXI YU AI
SUYANG JIAOCHENG



南京大学出版社

图书在版编目(CIP)数据

医学信息与人工智能素养教程 / 张鸿来, 顾金媛,
胡月主编. — 南京: 南京大学出版社, 2026. 6.
(高等学校人工智能通识课程融合系列规划教材).

ISBN 978-7-305-30072-1

I. R319

中国国家版本馆 CIP 数据核字第 2026YQ7987 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

丛 书 名 高等学校人工智能通识课程融合系列规划教材

书 名 医学信息与人工智能素养教程

YIXUE XINXI YU RENGONG ZHINENG SUYANG JIAOCHENG

主 编 张鸿来 顾金媛 胡 月

责任编辑 苗庆松 编辑热线 025-83592655

照 排 南京开卷文化传媒有限公司

印 刷 南京新洲印刷有限公司

开 本 787 mm×1092 mm 1/16 印张 16.25 字数 400 千

版 次 2026 年 6 月第 1 版 2026 年 6 月第 1 次印刷

ISBN 978-7-305-30072-1

定 价 49.80 元

网 址: <http://www.njupco.com>

官方微博: <http://weibo.com/njupco>

官方微信号: NJUyuexue

销售咨询热线: (025)83594756

* 版权所有, 侵权必究

* 凡购买南大版图书, 如有印装质量问题, 请与所购
图书销售部门联系调换

南京大学出版社
版权所有

前言

信息素养(Information Literacy)又称信息素质、信息能力,本质是全球信息化背景下需要人们必备的核心能力,最早由美国信息产业协会主席保罗·泽考斯基(Paul Zurkowski)于 1974 年提出,被概括为“利用大量信息工具及主要信息源使问题得到解答的技术和技能”。2015 年初,美国大学与研究型图书馆协会(ACRL)发布的《高等教育信息素养教育框架》,进一步强化了大学生在学术研究、批判性思维、反思性学习等方面的综合信息素养培养要求。

随着人工智能技术与医学领域的深度融合,信息素养的内涵与外延已全面拓展,如何将 AI 能力融入医学信息素养教育,助力医学生和医疗工作者适应智能医疗时代的发展需求,成为当前亟待探讨和解决的重要命题。当前,人工智能、大数据、云计算等技术正重塑医疗行业生态,医学信息呈现爆发式增长且形态愈发复杂。当代医学生及广大医疗工作者不仅需要掌握医学研究最新动态,还需要具备运用 AI 工具处理、分析、评价医学信息的能力,才能将前沿技术有效转化为临床实践与科研创新的动力。医学信息的智能检索、AI 辅助的文献分析、智能医学应用的实践落地等,已成为医学生专业学习与医务工作者终身学习的核心技能。因此,培养和提升融合 AI 能力的医学信息素养,是高等医学院校人才培养的关键目标之一。

本教程由南京医科大学康达学院张鸿来、顾金媛、胡月担任主编,负责全书整体策划及统稿编写;南京医科大学康达学院闫珮珮、曹悦德、郑洪燕、马静,南

京中医药大学王瑞娟担任副主编,负责相关章节编写。全书共 11 章,分别为:第 1 章人工智能基础,系统阐述了人工智能的定义、发展历程、应用现状及关键技术;第 2 章智能医学,介绍智能医学的定义、研究方法、发展现状、核心应用及面临的挑战与前景;第 3 章信息素养基础,讲解医学信息素养的基础理论、信息类型、信息意识、信息伦理与法律等核心内容;第 4 章信息检索基础,详细说明计算机检索原理、检索语言、检索技术、途径方法及效果评价;第 5 章常用 AI 工具,聚焦学术研究场景下主流 AI 开发框架、应用型 AI 工具(DeepSeek、Kimi AI 助手、秘塔 AI 等)的功能与合规使用;第 6 章书目信息检索,涵盖图书馆馆藏书目检索、分类法、检索系统及服务;第 7 章特种文献,介绍学位论文、会议文献、专利文献等特种文献的检索工具与方法;第 8 章中文数据库,以 SinoMed、中国知网等为例,讲解中文医学信息检索路径;第 9 章外文数据库,围绕 PubMed、Web of Science 等核心数据库,阐述外文医学信息检索技巧;第 10 章网络信息检索,涵盖网络医学资源、搜索引擎及专业信息资源的获取;第 11 章医学信息评价与利用,包括文献管理、论文写作、信息伦理规范等实用内容,全方位助力使用者高效运用医学信息资源。

本教程深入浅出、兼顾理论与实践,既保留传统医学信息素养的核心内容,又强化 AI 技术的融合应用,对医学生及医疗工作者提升智能时代的信息素养具有极强的指导性和实用性。教材宗旨在于帮助读者系统掌握各类医学信息资源的获取方法,熟练运用 AI 工具优化信息处理流程,增强信息意识与批判性思维,全面提升医学信息与人工智能素养。在本书编写过程中,我们广泛吸纳国内外相关研究成果,参考并引用了部分专家学者的著述,在此谨致以诚挚的谢意。

编者

2026 年 1 月

南京出版
社
版权所有

目 录

第 1 章 人工智能基础 001

- 1.1 人工智能定义 001
- 1.2 人工智能的发展历程 003
 - 1.2.1 人工智能的形成期(1956—1976 年) 004
 - 1.2.2 符号智能时期(1976—2006 年) 004
 - 1.2.3 数据智能时期(2006 年至今) 005
- 1.3 人工智能应用现状 006
- 1.4 人工智能关键技术 008
 - 1.4.1 机器学习 008
 - 1.4.2 深度学习 010
 - 1.4.3 自然语言处理 011
 - 1.4.4 计算机视觉 011
 - 1.4.5 知识图谱 012
 - 1.4.6 机器人技术 013
- 1.5 思考题 014

第 2 章 智能医学 015

- 2.1 智能医学定义 015
- 2.2 智能医学主要研究方法 016
 - 2.2.1 医学影像处理与分析 016
 - 2.2.2 生物信息学研究 017
 - 2.2.3 医疗机器人与智能设备应用 018
 - 2.2.4 智能医疗监测与数据挖掘 019
 - 2.2.5 跨学科融合研究 020
- 2.3 智能医学国内外发展现状 022

- 2.3.1 国内发展现状 022
- 2.3.2 国外发展现状 023
- 2.3.3 国内外发展对比、问题挑战及未来发展趋势 024

- 2.4 智能医学的应用 024
 - 2.4.1 医学影像分析 025
 - 2.4.2 3D 打印技术 025
 - 2.4.3 医用机器人 025
 - 2.4.4 可穿戴医疗设备 026
 - 2.4.5 疾病预测与风险评估 027
 - 2.4.6 远程医疗 027
 - 2.4.7 个性化治疗 028
 - 2.4.8 大数据与云平台 028
- 2.5 智能医学面临的挑战及发展前景 029
 - 2.5.1 智能医学面临的挑战 029
 - 2.5.2 智能医学的发展前景 029
- 2.6 思考题 030

第 3 章 信息素养基础 031

- 3.1 信息素养的相关概念 031
 - 3.1.1 信息素养的概念 031
 - 3.1.2 信息素养的内容和内涵 031
 - 3.1.3 信息素养的特征 032
 - 3.1.4 信息素养的表现 033
 - 3.1.5 现有信息素养教育相关框架 033
 - 3.1.6 对大学生信息素养的要求 036
 - 3.1.7 信息素养教育与文献检索课 036

- 3.1.8 医学信息素养教育 037
- 3.1.9 人工智能背景下的信息素养 038
- 3.2 信息意识 038
 - 3.2.1 信息意识基本概念 038
 - 3.2.2 信息意识基本组成 040
 - 3.2.3 信息意识在高校信息素养中的重要作用 041
- 3.3 信息伦理 043
 - 3.3.1 信息伦理基本概念 043
 - 3.3.2 伦理与道德的关系 044
 - 3.3.3 信息伦理基本组成 044
 - 3.3.4 信息伦理主要特征 045
 - 3.3.5 信息伦理发展史 045
 - 3.3.6 信息伦理主要问题 046
 - 3.3.7 信息伦理主要成因 049
- 3.4 信息法律 052
 - 3.4.1 信息法律基本概念 052
 - 3.4.2 信息法律与信息伦理的关系 052
 - 3.4.3 信息法律主要内容 053
 - 3.4.4 我国信息法律现状 054
 - 3.4.5 国外信息法律现状 055
- 3.5 思考题 058

第4章 信息检索基础 059

- 4.1 信息与信息源 059
 - 4.1.1 信息的概念 059
 - 4.1.2 信息的特征 060
 - 4.1.3 信息、知识、情报与文献 060
 - 4.1.4 信息源 062
 - 4.1.5 科学活动中的信息源分类 064
- 4.2 信息检索基本原理 065
 - 4.2.1 信息检索的概念 066
 - 4.2.2 信息检索的发展历程 066
 - 4.2.3 信息检索的原理 069
- 4.3 检索语言 070
 - 4.3.1 检索语言及其作用 070
 - 4.3.2 检索语言类型 071
 - 4.3.3 医学信息检索常用的检索语言 076
- 4.4 信息检索技术 079
 - 4.4.1 信息检索技术的概念 079

- 4.4.2 信息检索技术的热点 084
- 4.5 信息检索的方法、途径和步骤 085
 - 4.5.1 常用的信息检索方法 085
 - 4.5.2 信息检索途径 087
 - 4.5.3 信息检索的步骤 089
- 4.6 检索效果评价 093
 - 4.6.1 检索效果评价指标 093
 - 4.6.2 提高检索效果的方法 096
 - 4.6.3 信息检索的常见误区及改进策略 097
- 4.7 思考题 099

第5章 常用AI工具 100

- 5.1 常用AI工具综述 100
 - 5.1.1 人工智能工具生态系统的演变与发展 100
 - 5.1.2 主流AI开发框架与平台 100
 - 5.1.3 应用型AI工具概览 101
 - 5.1.4 AI辅助学术研究引发的伦理与社会议题 101
 - 5.1.5 构建“人机协同”的健康学术生态 103
- 5.2 DeepSeek 104
 - 5.2.1 DeepSeek的核心功能 104
 - 5.2.2 DeepSeek在文献分析与知识管理方面的应用 105
 - 5.2.3 DeepSeek在学术写作与知识管理方面的应用 107
 - 5.2.4 DeepSeek的学术伦理与AI使用边界 109
- 5.3 Kimi AI助手 111
 - 5.3.1 Kimi AI助手概述 111
 - 5.3.2 Kimi在文献分析与知识管理方面的应用 112
 - 5.3.3 Kimi在学术写作辅助方面的应用 113
 - 5.3.4 Kimi的学术伦理与AI使用边界 115
- 5.4 秘塔AI 117
 - 5.4.1 秘塔AI简介 117
 - 5.4.2 秘塔AI在文献分析与知识管

理方面的应用	118	6.5.2 联合目录	140
5.4.3 秘塔 AI 学术写作辅助方面的应用	120	6.5.3 国内联合目录	142
5.4.4 学术伦理与 AI 使用边界	121	6.5.4 CALIS 联合目录公共检索系统	142
5.5 思考题	121	6.5.5 世界书目(OCLC WorldCat)	142
第 6 章 书目信息检索	122	6.6 思考题	143
6.1 中国图书馆分类法	122	第 7 章 特种文献	144
6.1.1 简介	122	7.1 认识特种文献	144
6.1.2 《中国图书馆分类法》的结构	123	7.2 会议文献	144
6.1.3 《中国图书馆分类法》的体系	123	7.2.1 会议文献概述	144
6.1.4 类目的划分与排列	125	7.2.2 会议文献的类型	144
6.1.5 《中国图书馆分类法》的标记符号和标记制度	129	7.2.3 会议文献的出版形式	145
6.1.6 《中国图书馆分类法》的意义	130	7.2.4 会议文献检索工具	145
6.2 图书馆书目检索系统	132	7.3 学位论文数据库	146
6.2.1 馆藏目录的概念	132	7.3.1 学位论文相关概念	146
6.2.2 馆藏目录的作用	132	7.3.2 学位论文数据库检索	146
6.2.3 馆藏目录的分类	133	7.3.3 学位论文数字化项目资源	147
6.2.4 图书馆网站	133	7.4 专利文献	148
6.2.5 馆藏书目检索系统	133	7.4.1 专利文献的相关概念	148
6.2.6 OPAC 功能介绍	134	7.4.2 专利的网上检索	149
6.2.7 图书馆数字资源的利用	134	7.5 标准文献	150
6.3 图书馆服务	135	7.5.1 标准文献的概念	150
6.3.1 图书馆服务简介	135	7.5.2 标准文献的检索	151
6.3.2 借阅服务	135	7.6 政府出版物	152
6.3.3 文献检索	136	7.6.1 政府出版物的概念	152
6.3.4 参考咨询	136	7.6.2 政府出版物的检索	152
6.3.5 科技查新	136	7.7 科技报告	155
6.3.6 定题服务与学科服务	137	7.7.1 科技报告概述	155
6.3.7 宣传辅导与用户培训	137	7.7.2 科技报告检索工具	156
6.3.8 馆际互借与文献传递	138	7.8 思考题	157
6.4 馆藏图书利用	138	第 8 章 中文数据库	158
6.4.1 图书分类和排架	138	8.1 中国生物医学文献服务系统(SinoMed)	158
6.4.2 分类号的排列	139	8.1.1 SinoMed 概述	158
6.4.3 期刊排架	139	8.1.2 SinoMed 检索方式与技巧	159
6.5 馆藏目录	139	8.2 中国知网	166
6.5.1 馆藏目录查询	139	8.2.1 中国知网概述	166
		8.2.2 中国知网检索方式与技巧	167

8.2.3 中国知网 AI 功能与应用指南	172	10.2.3 综合型搜索引擎	218
8.3 万方数据知识服务平台	176	10.2.4 学术搜索引擎	220
8.3.1 万方数据库概述	176	10.2.5 医学专业搜索引擎	221
8.3.2 万方数据库检索方式与技巧	177	10.3 网络医学参考工具书	222
8.3.3 万方 AI 功能与应用指南	180	10.3.1 网络参考工具书概述	222
8.4 读秀	182	10.3.2 网络医学参考工具书举例	223
8.4.1 读秀学术搜索概述	182	10.4 基础医学信息资源	224
8.4.2 读秀检索方式与技巧	183	10.4.1 基础医学文献信息概述	224
8.5 思考题	185	10.4.2 基础医学信息资源	224
第 9 章 外文数据库	186	10.5 临床医学信息资源	226
9.1 PubMed	186	10.5.1 临床医学信息概述	226
9.1.1 PubMed 概述	186	10.5.2 常用网络上的临床医学信息资源	226
9.1.2 PubMed 检索方式与技巧	188	10.6 思考题	228
9.2 Wiley Online Library	195	第 11 章 医学信息评价与利用	229
9.2.1 Wiley Online Library 概述	195	11.1 医学文献管理	229
9.2.2 Wiley Online Library 检索方式与技巧	195	11.1.1 文献的整理	229
9.3 Web of Science	199	11.1.2 文献的鉴别	229
9.3.1 引文检索概述	199	11.1.3 文献的分析与评价	230
9.3.2 Web of Science 核心合集检索方式与技巧	204	11.2 常用文献管理软件	232
9.4 Science Direct	208	11.2.1 Endnote	233
9.4.1 Science Direct 数据库概述	208	11.2.2 NoteExpress	234
9.4.2 Science Direct 数据库检索方式与技巧	209	11.3 医学论文写作	236
9.5 思考题	211	11.3.1 医学论文写作概述	236
第 10 章 网络信息检索	212	11.3.2 医学论文写作特点	236
10.1 网络医学信息资源概述	212	11.3.3 医学论文写作基本格式及要求	240
10.1.1 网络医学信息资源定义	212	11.3.4 医学论文写作方法与步骤	245
10.1.2 网络医学信息资源特点	212	11.4 信息伦理道德及规范	249
10.1.3 网络医学信息资源类型	213	11.4.1 信息伦理失范现象	249
10.2 搜索引擎	216	11.4.2 信息伦理规范和信息道德的培养	250
10.2.1 搜索引擎概念	216	11.5 思考题	251
10.2.2 搜索引擎分类	217	参考文献	252

第 1 章

人工智能基础

1.1 人工智能定义

人工智能是极具挑战性的领域。伴随着大数据、云计算、深度学习和类脑智能等技术的发展,人工智能的浪潮又一次掀起。目前信息技术、互联网等领域几乎所有主题和热点,如搜索引擎、智能硬件、机器人、无人机和工业 4.0,其发展突破的关键环节都与人工智能有关。

1956 年四位年轻学者:麦卡锡(McCarthy J.)、明斯基(Minsky M.)、罗彻斯特(Rochester N.)和香农(Shannon C.)共同发起和组织召开了用机器模拟人类智能的夏季专题讨论会。会议邀请了包括数学、神经生理学、精神病学、心理学、信息论和计算机科学领域的 10 名学者参加,为期两个月。此次会议是在美国的新罕布什尔州的达特茅斯(Dartmouth)学院召开的,也称为达特茅斯夏季讨论会。

会议上,科学家运用数理逻辑和计算机的成果,提供关于形式化计算和处理的理论,模拟人类某些智能行为的基本方法和技术,构造具有一定智能的人工系统,让计算机去完成需要人的智力才能胜任的工作。其中明斯基的神经网络模拟器、麦卡锡的搜索法、西蒙(Simon H.)和纽厄尔(Newell A.)的“逻辑理论家”成为讨论会的三个亮点。

在达特茅斯夏季讨论会上,麦卡锡提议用“Artificial Intelligence”(人工智能)作为这一交叉学科的名称,定义为制造智能机器的科学与工程,标志着人工智能学科的诞生。半个多世纪来,人们从不同的角度、不同的层面给出对人工智能的定义。下面介绍 4 种对人工智能的定义。

(1) 类人行为方法

库兹韦尔(Kurzweil R.)提出人工智能是一种创建机器的技艺,这种机器能够执行需要人的智能才能完成的功能(Kurzweil,1990)。这与图灵测试的观点很吻合,是一种类人行为定义的方法。1950 年,图灵(Turing A.)提出图灵测试,并将“计算”定义为:应用形式规则,对未加解释的符号进行操作(Turing,1950)。图 1.1 给出了图灵测试的示意图,将一个人与一台机器置于一间房间中,而与另外一个人分隔开来,并把后一个人称为询问者。询问者不能直接见到屋中任一方,也不能与他们说话,因此他不知道到底哪一个实体是机器,只可以

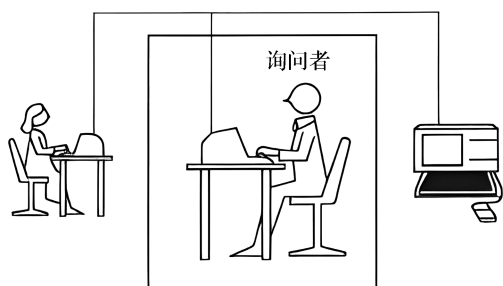


图 1.1 图灵测试

通过一个类似终端的文本设备与他们联系。然后让询问者仅根据通过这个仪器提问收到的答案辨别出哪个是计算机,哪个是人。如果询问者不能区别出机器和人,那么根据图灵的理论,就可以认为这个机器是智能的。

图灵测试具有直观上的吸引力,成为许多现代人工智能系统评价的基础。如果一个系统已经有可能在某个专业领域实现智能,那么就可以通过把它对一系列给定问题的反馈与人类

专家的反馈相比较来对其进行评估。

图灵测试也引发了很多争议,其中最著名的是塞尔(Searle J.)的“中文屋论证”(Searle, 1980)。塞尔设想自己被锁在一间屋子里,给了他大批的中文文本,塞尔本人对中文一窍不通,既不会写也不会说,甚至也不能将中文文本与日文中的汉字和平假名/片假名一样的图形相区别。这时他又得到了与这个中文文本相联系的英文规则书,由于塞尔的母语是英文,所以他认为自己可以轻易地理解并把握这本规则书。接下来塞尔将接收到屋外传来的英文指令和中文问题,指令教他怎样将规则书与中文文本联系起来,得到答案。当塞尔对规则书和脚本足够熟悉的时候,他就可以熟练地输出处理编写后的中文答案。一般人也难以区分塞尔与母语讲中文的人,但是事实上,塞尔认为整个过程中他根本不懂、不理解中文,只是执行规则书上的“程序”。这种行为在中国人看来与计算机用中文作答没有什么区别,但却成功地通过了图灵测试,并不具有理解中文的智能。基于这一点,塞尔认为即使机器通过了图灵测试,也不一定说明机器就真的能像人一样有思维和意识。

(2) 类人思维方法

1978年,贝尔曼(Bellman R. E.)提出人工智能是那些与人的思维、决策、问题求解和学习等有关活动的自动化(Bellman, 1978),主要采用的是认知模型的方法——关于人类思维工作原理的可检测的理论。为确定人类思维的内部是怎样工作的,可以有两种方法:通过内省或者通过心理学实验。一旦有了关于人类思维足够精确的理论,就可能把这种理论用计算机程序实现。如果该程序的输入/输出和实时行为与人的行为相一致,就证明该程序可能是按照人类模式运行的。例如,纽厄尔和西蒙开发了“通用问题求解器”(General Problem Solver, GPS),他们并不满足于仅让程序能够正确地求解问题,而是更关心对程序的推理步骤轨迹与人对同一个问题的求解步骤的比较。作为交叉学科的认知科学,把来自人工智能的计算机模型与来自心理学的实验技术相结合,试图创立一种精确而且可检验的人类思维工作方式理论。

20世纪50年代末,人们对神经元的模拟中提出了用一种符号来标记另一些符号的存储结构模型,这是早期的记忆块(Chunks)概念。20世纪80年代初,纽厄尔认为,通过获取任务环境中关于模型问题的知识,可以改进系统的性能,记忆块可以作为对人类行为进行模拟的模型基础。通过观察问题求解过程,获取经验记忆块,代替各个子目标中的复杂过程,可以明显提高系统求解的速度。由此奠定了经验学习的基础。1987年,纽厄尔、莱尔德(Laird J.)和罗森布鲁姆(Rosenbloom P. S.)提出了一个通用解题结构 SOAR(Newell et al., 1987),并希望该解题结构能实现各种弱方法。SOAR 是“State, Operator and Result”

的缩写,即状态、算子和结果之意,意味着实现弱方法的基本原理是不断地用算子作用于状态,以得到新的结果。SOAR 是一种理论认知模型,它既从心理学角度对人类认知建模,又从知识工程角度提出一个通用解题结构。SOAR 的学习机制是由外部专家的指导来学习一般的搜索控制知识。外部指导可以是直接劝告,也可以是给出一个直观的简单问题。系统把外部指导给定的高水平信息转化为内部表示,并学习搜索记忆块。

(3) 理性思维方法

1985 年,查尼亚克(Charniak E.)和麦克德莫特(McDermott D.)提出人工智能是用计算模型研究智力能力(Charniak et al., 1985)。这是一种理性思维方法,一个系统如果能够按它所知范围内正确行事,它就是理性的。古希腊哲学家亚里士多德(Aristotle)是首先试图严格定义“正确思维”的人之一,他将之定义为不能辩驳的推理过程。他的三段论方法给出了一种推理模式,当已知前提正确时总能产生正确的结论。例如,专家系统是推理系统,所有的推理系统都是智能系统,所以专家系统是智能系统。这些思维法则被认为支配着心智活动,对它们的研究创立了“逻辑学”研究领域。

19 世纪后期至 20 世纪早期发展起来的形式逻辑给出了描述事物的语句以及事物之间关系的精确的符号。到了 1965 年,原则上已经有程序可以求解任何用逻辑符号描述的可解问题。在人工智能领域中,传统上所谓的逻辑主义希望通过编制逻辑程序来创建智能系统。

这种逻辑方法有两个主要问题。首先,把非形式的知识用形式的逻辑符号表示是不容易做到的,特别是当这些知识不是 100%确定的时候。其次,原则上可以解决一个问题与实际解决问题之间有很大的不同。甚至对于仅有几十条事实的问题进行求解,如果没有一定的指导来选择合适的推理步骤,都可能耗尽任何计算机的资源。

(4) 理性行为方法

尼尔森(Nilsson N. J.)认为人工智能关心的是人工制品中的智能行为(Nilsson, 1998),这种人工制品主要指能够动作的智能体(Agent)。行为上的理性指的是已知某些信念,执行某些动作以达到某个目标。智能体看作是可以进行感知和执行动作的某个系统。在这种方法中,人工智能可以认为就是研究和建造理性智能体。

在“理性思维”方法中,它所强调的是正确的推理。做出正确的推理有时被作为理性智能体的一部分,因为理性行动的一种方法是逻辑地推出结论。另外,正确的推理并不是理性的全部,因为在有些情景下,往往没有某个行为一定是正确的,而其他的是错误的,也就是说没有可以证明是正确的应该做的事情,但是还必须要做某件事情。

当知识是完全的,并且资源是无限的时候,就是所谓的逻辑推理。当知识是不完全的,或者资源有限时,就是理性的行为。理性思维和行为常常能够根据已知的信息(知识、时间和资源等)做出最合适的决策。

简言之,人工智能主要研究用人工的方法和技术,模仿、延伸和扩展人的智能,实现机器智能。人工智能的长期目标是实现达到人类智力水平的人工智能。

1.2 人工智能的发展历程

人类对智能机器的梦想和追求可以追溯到三千多年前。早在我国西周时代(公元前

1046—公元前 771 年),就有巧匠偃师献给周穆王艺伎的传说。东汉(公元 25—公元 220 年)张衡发明的指南车是世界上最早的机器人雏形。

古希腊的亚里士多德(公元前 384 年—公元前 322 年)所著的《工具论》,为形式逻辑奠定了基础。布尔(Boole)创立的逻辑代数系统,用符号语言描述了思维活动中推理的基本法则,被后世称为“布尔代数”。这些理论基础对人工智能的创立发挥了重要作用。

1936 年,图灵创立了理想计算机模型的自动机理论,提出了以离散量的递归函数作为智能描述的数学基础,给出了基于行为主义的测试机器是否具有智能的标准,即图灵测试。

1943 年,神经生理学家麦卡洛克(McCulloch W. S.)和数理逻辑学家皮茨(Pitts W.)在《数学生物物理公报》(*Bulletin of Mathematical Biophysics*)上发表了关于神经网络的数学模型(McCulloch et al., 1943)。这个模型,现在一般称为 M-P 神经网络模型。他们总结了神经元的一些基本生理特性,提出神经元形式化的数学描述和网络的结构方法,从此开创了神经计算的时代。

1956 年,在美国的达特茅斯学院召开了为期两个月的学术研讨会,提出了“人工智能”这一术语,标志着这门学科的正式诞生。

60 多年中,人工智能的发展经历了形成期、符号智能和数据智能时期。

1.2.1 人工智能的形成期(1956—1976 年)

这一时期的主要贡献包括:

(1) 1956 年,西蒙和纽厄尔的“逻辑理论家”程序,该程序模拟了人们用数理逻辑证明定理时的思维规律。

(2) 1956 年,乔姆斯基(Chomsky N.)提出形式语言的理论。该理论对计算机科学有着深刻的影响,特别是在程序设计语言的设计、编译方法和计算复杂性等方面发挥着重大的作用。

(3) 1958 年,麦卡锡提出表处理语言 LISP,它不仅可以处理数据,而且可以方便地处理符号,成为人工智能程序设计语言的重要里程碑。时至多年,LISP 语言仍然是人工智能系统重要的程序设计语言和开发工具。

(4) 1965 年,鲁宾逊(Robinson J. A.)提出归结法,被认为是一个重大的突破,也为定理证明的研究带来了又一次高潮。

(5) 1965 年,斯坦福大学的费根鲍姆(Feigenbaum E. A.)和化学家勒德贝格(Lederberg J.)开始合作研制化学分析专家系统 DENDRAL,并于 1968 年研制成功。1972—1976 年,费根鲍姆成功开发医疗专家系统 MYCIN。

1.2.2 符号智能时期(1976—2006 年)

这一时期的主要贡献包括:

(1) 1975 年,西蒙和纽厄尔荣获计算机科学最高奖——图灵奖。1976 年,因在获奖演讲中提出了“物理符号系统假说”,他们成为人工智能中影响最大的符号主义学派的创始人和代表人物。

(2) 1977 年,美国斯坦福大学的计算机科学家费根鲍姆在第五届国际人工智能联合会议上提出知识工程的新概念。20 世纪 80 年代,专家系统的开发趋于商业化,创造了巨大的经济效益。知识工程是一门以知识为研究对象的学科,使人工智能的研究从理论转向应用,

从基于推理的模型转向知识的模型。

(3) 1981年,日本宣布了第五代电子计算机的研制计划。其研制的计算机的主要特征是具有智能接口、知识库管理、自动解决问题的能力,并在其他方面具有人的智能行为(史忠植,1988)。

(4) 1984年,莱斯利·瓦利安特(Leslie Valiant)在计算科学和数学领域的远见及认知理论与其他技术结合后,提出可学习理论,开创了机器学习和通信的新时代(Valiant,1984)。

(5) 1993年,肖哈姆(Shoham Y.)提出面向智能体的程序设计(史忠植,2000)。

(6) 1995年,罗素(Russell S.)和诺维格(Norvig P.)出版了《人工智能》一书,提出“将人工智能定义为对从环境中接收感知信息并执行行动的智能体的研究”(Russell et al.,1995)。

(7) 2011年,朱迪亚·珀尔(Judea Pearl)获得图灵奖,奖励他在人工智能领域的基础性贡献:提出概率和因果性推理演算法,彻底改变了人工智能基于规则和逻辑的方向(Pearl,2000)。

1.2.3 数据智能时期(2006年至今)

这一时期的主要贡献包括:

(1) 2006年,杰弗里·辛顿(Geoffrey Hinton)等发表深度信念网络,开创了深度学习的新阶段。

(2) 2016年,AlphaGo采用深度强化学习,击败最强的人类围棋选手之一李世石,推动了人工智能的发展和普及。

(3) 2016年,寒武纪推出了国际上首个稀疏深度学习处理器 Cambricon-X,在速度(响应时间)和功耗这两个关键指标上实现了对当时传统 CPU 和 GPU 的显著超越。

(4) 2018年,图灵奖被共同授予深度学习领域的三大奠基人:杰弗里·辛顿(Geoffrey Hinton)、杨立昆(Yann LeCun)、约书亚·本吉奥(Yoshua Bengio),他们开创了深度神经网络,为深度学习算法的发展和应用奠定了基础。

(5) 2022年11月30日,美国 OpenAI 发布聊天机器人程序 ChatGPT。2023年3月16日,百度文心一言正式发布。2024年1月5日,DeepSeek 发布了首个大模型。大语言模型吸引企业和机构客户使用 API 和基础设施,共同搭建 AI 模型、开发应用,实现产业 AI 普惠。

我国的人工智能研究起步较晚。智能模拟纳入国家计划的研究始于1978年。1984年召开了智能计算机及其系统的全国学术讨论会。1986年起把智能计算机系统、智能机器人和智能信息处理(含模式识别)等重大项目列入国家高技术研究“863计划”。1997年起,又把智能信息处理、智能控制等项目列入国家重大基础研究“973计划”。进入21世纪后,在《国家中长期科学和技术发展规划纲要(2006—2020年)》中,将“脑科学与认知科学”列入八大科学前沿问题之一。2017年7月8日,国务院正式发布《新一代人工智能发展规划》,部署有关研究计划。信息技术将继续向高性能、低成本、普适计算和智能化等主要方向发展,寻求新的计算与处理方式和物理实现是未来信息技术领域面临的重大挑战。

1981年起,我国相继成立了中国人工智能学会(CAAI)、全国高校人工智能研究会、中国计算机学会人工智能与模式识别专业委员会、中国自动化学会模式识别与机器智能专业委员会等学术团体,1989年首次召开了中国人工智能联合会议(CJCAI),1989年创刊了《模

式识别与人工智能》期刊,2006 年创刊了《智能系统学报》和《智能技术》期刊,2011 年创刊了国际刊物 *International Journal of Intelligence Science*。

我国的科技工作者已在人工智能领域取得了具有国际领先水平的创造性成果,其中尤以吴文俊院士关于几何定理机器证明的“吴方法”最为突出,荣获 2000 年国家科学技术最高奖励。现在,我国已有数以万计的科技人员和大学师生从事不同层次的人工智能研究与学习。人工智能研究已在我国深入开展,它必将为促进其他学科的发展和我国的现代化建设做出新的重大贡献。

1.3 人工智能应用现状

人工智能(AI)作为当今最前沿的技术领域之一,目前在多个领域都有广泛且深入的应用。截至 2025 年,人工智能(AI)技术已在多个领域实现深度渗透,重塑产业生态,展现出了强劲的发展活力,推动着社会经济的智能化转型,以下是一些主要领域的应用现状。

(1) 医疗健康领域

人工智能技术被广泛用于疾病诊断与影像分析,如 X 光、CT 扫描和 MRI 等。它能够快速准确地识别影像中的模式和异常情况,辅助医生进行诊断,提高诊断的速度和准确性,减少人为错误。例如,Google Health 的 AI 系统在乳腺癌检测中,诊断准确率超越了放射科医生;IDx-DR 是第一个获得美国 FDA 批准的人工智能眼科诊断系统,能够在没有医生参与的情况下检测糖尿病视网膜病变。

(2) 个性化医疗

通过分析患者的基因信息、病历等多源数据,人工智能可以为患者制定个性化的治疗方案。比如 IBM Watson for Oncology 能分析患者病历和研究文献,提供个性化的癌症治疗建议;Deep Genomics 利用 AI 预测基因突变对人体健康的影响,帮助设计针对性的治疗方案。

(3) 药物研发

在药物研发过程中,AI 可以加速药物发现过程,识别潜在的药物分子,并优化临床试验。例如 Atomwise 使用 AI 进行药物分子结构分析,帮助发现新药物组合;Benevolent AI 利用机器学习算法分析生物医学数据,寻找潜在的治疗药物。

(4) 金融服务领域

算法交易:人工智能用于自动化交易策略的开发,能够快速分析海量的市场数据,实时做出交易决策,大大提高了市场反应速度和交易效率。例如 Kensho Technologies 利用 AI 进行市场分析,提供投资策略建议;Renaissance Technologies 的量化基金使用复杂的 AI 算法进行交易,获得了显著的投资回报。

(5) 风险管理与欺诈检测

AI 可以识别和预测金融风险,检测欺诈行为,保护用户资产安全。例如 PayPal 使用机

机器学习模型监控交易活动,实时识别潜在的欺诈行为;JP Morgan's Coin 通过 AI 分析合约,自动化地处理银行的合规事务,减少人为错误。

(6) 客户服务与个性化推荐

金融机构采用 AI 驱动的聊天机器人和虚拟助手提供个性化的客户服务,同时分析用户数据以推荐金融产品。例如美国银行的虚拟助手 Erica 帮助用户管理财务事务,提供个性化理财建议;Acorns 利用 AI 分析用户消费习惯,提供智能投资建议。

(7) 制造业领域

智能化与机器人:人工智能应用于工业机器人和自动化系统,优化生产流程,提高生产效率和产品质量。例如 FANUC 机器人在制造业中使用机器学习技术进行质量控制和自动化操作;Siemens 的 MindSphere 通过 AI 分析工业设备数据,进行预测性维护和优化。

(8) 预测性维护

通过分析设备数据,人工智能可以预测潜在故障,提前进行维护,减少停机时间和维修成本,提高设备利用率和生产连续性。例如 GE 的 Predix 平台利用 AI 进行工业设备的实时监控和预测性维护;Uptake Technologies 提供工业数据分析平台,通过 AI 预测机器故障,优化维护策略。

(9) 供应链优化

AI 用于分析供应链数据,优化库存管理,提升供应链的效率和灵活性。例如 Amazon 的仓库机器人利用 AI 技术优化仓储管理,提高物流效率;DHL Supply Chain 通过 AI 分析物流数据,优化配送路线和库存策略。

(10) 零售业领域

智能购物体验:人工智能用于提供个性化推荐、虚拟试衣间和自动化客服等服务,提升购物体验,增加顾客的购物满意度和忠诚度。例如 Stitch Fix 利用 AI 分析用户偏好,提供个性化时尚建议;Sephora 的虚拟试妆工具通过 AI 技术让顾客在虚拟环境中试妆。

(11) 库存管理

AI 可以优化库存管理,预测需求变化,减少库存浪费和运营成本。例如 Zara 使用 AI 进行市场分析和库存管理,提高产品供应的准确性;Ocado 利用 AI 技术优化仓储和配送流程,提高物流效率。

(12) 消费者行为分析

通过分析顾客数据,人工智能可以识别消费趋势,帮助零售商更好地理解市场需求,优化产品和服务,支持市场策略调整。例如 Walgreens 利用 AI 分析顾客购买习惯,优化店内商品摆放和促销策略;阿里巴巴通过数据分析和 AI 技术实现精准营销,提升销售业绩。

(13) 交通与物流领域

自动驾驶技术:AI 在汽车自动驾驶中用于环境感知、决策制定和路径规划等。目前,越来越多的自动驾驶汽车投入商业运营或进行测试,像谷歌旗下的 Waymo、特斯拉的 Autopilot 等。自动驾驶技术有望减少交通事故,提高道路安全性,并提升交通运输的效率。

(14) 智能交通管理

人工智能用于交通流量分析和优化,提高城市交通管理的效率。例如一些智能交通系统利用 AI 分析实时交通数据,优化信号灯控制,减少交通拥堵;智能导航应用基于 AI 提供实时路况和路线建议,提高驾驶体验。

(15) 物流优化

AI 用于物流路径规划、车辆调度和仓储管理,优化配送流程,降低运输成本,提高客户满意度和企业竞争力。例如 UPS 的 ORION 系统利用 AI 优化快递路线,节省燃油成本和配送时间;京东的智能物流通过 AI 分析和预测物流需求,提高配送效率。

(16) 教育领域

个性化学习:全球已有超过 50% 的学校引入了人工智能辅助教学工具,能够根据学生的学习进度和风格,推荐合适的学习内容和路径,提高学生的学习效率和成绩。例如美国的 Knewton 公司,中国的猿辅导、作业帮等智能教育平台。

(17) 智能辅导

语音识别和自然语言处理技术在教育中的应用越来越广泛,智能语音助教系统能够通过语音识别技术实现与学生的实时互动,提供个性化辅导。

(18) 教育评估和测试

人工智能在教育评估和测试中的应用日益成熟,能够通过分析学生的答题过程,评估其真实水平,预测学生的未来表现,并且可以在在线教育平台中实现自动评分,提高评分的客观性和效率。例如英国剑桥大学与 IBM 合作开发的智能测试系统,以及 Coursera 和 edX 等教育平台。

AI 已从实验室走向规模化落地,在提升效率、降低成本、解决复杂问题等方面展现出了巨大潜力,但其发展仍须应对技术、伦理和社会的多重挑战。未来,AI 将更加注重与人类价值观对齐,向“负责任 AI”方向演进,同时会进一步渗透至垂直领域,成为推动全球数字化转型的核心引擎。

1.4 人工智能关键技术

人工智能的关键技术涵盖机器学习、深度学习、自然语言处理、计算机视觉等多个领域。

1.4.1 机器学习

机器学习是人工智能的核心领域之一,下面从定义、主要类型、常用算法及其应用场景等方面详细介绍。

(1) 定义

机器学习是一门多领域交叉学科,涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。它专门研究计算机怎样模拟或实现人类的学习行为,以获取新的知识或技能,重新组织已有的知识结构使之不断改善自身的性能。

(2) 主要类型

① 监督学习

分类:目标是将数据分类到不同的类别中。例如,根据电子邮件的内容分为垃圾邮件和非垃圾邮件,根据肿瘤的特征判断是良性还是恶性等。

回归:预测一个连续的数值。比如,根据房屋的面积、房间数量、地理位置等特征来预测房屋的价格,根据历史销售数据预测未来的销售额等。

② 无监督学习

聚类:将数据点划分为不同的簇,使得同一簇内的数据点相似度较高,不同簇之间的数据点相似度较低。例如,对客户的消费行为数据进行聚类,将客户分为不同的群体,以便进行精准营销。

降维:在不损失太多信息的前提下,将高维数据转换为低维数据,以便于数据的可视化、存储和分析。比如,主成分分析(PCA)常用于图像数据的降维,减少数据量的同时保留图像的主要特征。

③ 半监督学习:利用少量的标注数据和大量的未标注数据进行学习,结合了监督学习和无监督学习的优点,在标注数据获取成本较高的情况下有广泛应用。如图像识别中,只有少量图像有标注信息,通过半监督学习可以利用大量未标注图像来提高模型的性能。

④ 强化学习:智能体在环境中采取一系列行动,通过接收环境反馈的奖励信号来学习最优行为策略。例如,机器人通过不断尝试不同的动作来完成任务,根据任务完成的情况获得奖励,从而学习到最佳的行动策略;游戏中的智能角色通过强化学习,来学习如何在游戏中取得更高的分数。

(3) 常用算法

① 决策树:基于树结构进行决策,每个内部节点是一个属性上的测试,分支是测试输出,叶节点是类别或值,可解释性强,常用于分类和回归任务。

② 支持向量机:在特征空间中寻找一个最优的超平面来将不同类别的数据分开,在数据线性可分和非线性可分的情况下都有很好的表现,常用于二分类问题。

③ 朴素贝叶斯:基于贝叶斯定理,假设特征之间相互独立,计算每个类别在给定特征下的概率,选择概率最大的类别作为预测结果,常用于文本分类等领域。

④ K-Means:是一种经典的聚类算法,随机选择 K 个初始聚类中心,将数据点分配到距离最近的聚类中心所在的簇,然后更新聚类中心,重复此过程直到聚类结果稳定。

(4) 应用场景

① 金融领域:用于风险评估、信用评级、股票价格预测、欺诈检测等。通过分析大量的金融交易数据和用户信息,识别潜在的风险和欺诈行为,为金融机构提供决策支持。

② 医疗领域:辅助疾病诊断、治疗方案推荐、药物研发等。例如,通过分析患者的病历、影像数据等,帮助医生更准确地诊断疾病,预测疾病的发展趋势,为患者制定个性化的治疗方案。

③ 交通领域:交通流量预测、智能交通调度、自动驾驶等。根据历史交通数据和实时路况信息,预测交通流量,优化交通信号灯时长,提高交通效率,同时也是自动驾驶技术的重要支撑。

1.4.2 深度学习

深度学习是机器学习的一个分支领域,它基于人工神经网络,通过构建具有多个层次的神经网络模型,让计算机自动从大量数据中学习特征和模式,以实现数据的分类、预测、生成等任务。

(1) 神经网络架构

① 卷积神经网络(CNN)

结构特点:包含卷积层、池化层和全连接层。卷积层通过卷积核在数据上滑动进行卷积操作,自动提取数据中的局部特征,大大减少了模型的参数数量,降低计算量,提高了训练效率和泛化能力。池化层则对卷积层的输出进行下采样,减少数据维度,同时保留主要特征。

应用场景:在图像识别、目标检测、图像分割等计算机视觉任务中表现出色。例如,在人脸识别系统中,CNN 可以准确地提取人脸的特征,用于身份识别;在医学图像分析中,CNN 可以帮助医生检测肿瘤、识别病变等。

② 循环神经网络(RNN)及其变体

结构特点:RNN 能够处理序列数据,其内部的隐藏状态可以记住之前的信息,并将之传递到当前时刻,从而对序列中的长期依赖关系进行建模。LSTM 和 GRU 是 RNN 的变体,它们通过门控机制有效地解决了 RNN 中的长期依赖问题,能够更好地捕捉序列中的长期信息。

应用场景:广泛应用于自然语言处理领域,如机器翻译、文本生成、语音识别等。在机器翻译中,RNN 及其变体可以将源语言句子中的信息依次编码,并根据这些信息生成目标语言句子。

③ 生成对抗网络(GAN)

结构特点:由生成器和判别器组成,生成器负责生成与真实数据相似的样本,判别器则负责区分生成的样本和真实样本。两者通过对抗训练的方式不断优化,生成器努力生成更逼真的样本以欺骗判别器,判别器则努力提高辨别能力,最终达到一个动态平衡。

应用场景:在图像生成、数据增强、风格迁移等方面有广泛应用。可以生成逼真的人脸图像、风景图像等,还可以用于将一种风格的图像转换为另一种风格,如将油画风格转换为摄影风格。

(2) 训练算法

① 反向传播算法

原理:基于链式法则,将误差从输出层反向传播到输入层,依次计算每一层的误差梯度,然后根据梯度来更新神经网络的权重,使得损失函数朝着减小的方向变化,从而使模型逐渐收敛到最优解。

作用:是深度学习中计算梯度的核心方法,为神经网络的训练提供了有效的优化手段,使得模型能够学习到数据中的复杂模式和特征。

② 优化器

常见类型:随机梯度下降(SGD)及其变体 Adagrad、Adadelata、Adam 等。SGD 每次使用一个小批量的数据来计算梯度并更新权重,Adagrad 和 Adadelata 可以根据每个参数的历史梯度信息自适应地调整学习率,Adam 则结合了动量和自适应学习率的优点,能够更快地

收敛到最优解。

作用:用于更新神经网络的权重,调整模型的参数,以最小化损失函数。不同的优化器在不同的任务和数据集上可能有不同的表现,选择合适的优化器可以提高训练效率和模型性能。

(3) 应用领域

① 计算机视觉:除了前面提到的图像识别、目标检测和图像分割,还包括图像超分辨率、图像去噪、视频理解等。例如,图像超分辨率技术可以将低分辨率图像恢复为高分辨率图像,提高图像的清晰度和细节;视频理解任务可以对视频中的动作、场景,进行识别和分析,用于视频监控、自动驾驶等领域。

② 自然语言处理:除了机器翻译和文本生成,还包括情感分析、知识问答、信息检索等。例如,情感分析可以判断文本表达的情感倾向是正面还是负面,用于舆情监测、产品评价分析等;知识问答系统可以根据用户的问题从大量文本数据中提取准确的答案,为用户提供便捷的知识获取服务。

③ 语音处理:在语音识别、语音合成、语音增强等方面有重要应用。语音识别技术可以将语音信号转换为文字,用于语音助手、语音输入等场景;语音合成技术能够将文字转换为自然流畅的语音,为盲人阅读、智能客服等提供便利。

随着互联网技术的飞速发展,数据量呈爆炸式增长。如何有效地组织和利用这些海量数据,成了当前信息技术领域面临的重要挑战。知识图谱的出现,为解决这一问题提供了新的思路和方法,将在本章后续介绍。它能够将海量的非结构化和半结构化数据转化为结构化的知识,使得计算机能够更好地理解和处理人类的知识,从而推动人工智能技术的发展。

1.4.3 自然语言处理

(1) 语言模型

如 BERT、GPT 等,能够学习语言的语法、语义和语用知识,用于文本生成、知识问答、机器翻译等任务,GPT-4 可以根据给定的提示生成连贯、有逻辑的文本内容。

(2) 语义理解

包括词法分析、句法分析、语义角色标注等技术,旨在理解文本的语义结构和意图,使计算机能够正确理解和处理人类语言,如智能客服系统能够理解用户的问题并给出准确回答。

(3) 文本生成

根据给定的主题、语境或条件生成自然语言文本,如新闻写作、故事生成、对话生成等,一些智能写作助手可以帮助用户快速生成文章初稿。

1.4.4 计算机视觉

(1) 图像识别

通过提取图像的特征,与已知的图像类别进行匹配,实现对图像内容的分类和识别,如在安防监控中识别行人、车辆等目标。

(2) 目标检测

不仅要识别图像中的物体类别,还要确定物体的位置和边界框,广泛应用于自动驾驶、

智能安防等领域,如自动驾驶汽车需要检测道路上的行人、车辆、交通标志等目标。

(3) 图像分割

将图像划分为不同的区域,每个区域对应一个特定的物体或语义类别,在医学图像处理、遥感图像分析等领域有重要应用,如医学影像中的器官分割,有助于医生进行疾病诊断。

1.4.5 知识图谱

(1) 定义

知识图谱本质上是一种语义网络,它由节点和边组成。节点代表实体,如人物、事件、地点等;边代表实体之间的关系,如“出生于”“包含”“属于”等。通过这种方式,知识图谱能够将分散的知识有机地整合在一起,形成一个庞大的知识网络。

(2) 组成要素

- ① 实体:是知识图谱中的基本元素,具有明确的语义含义和唯一的标识。
- ② 属性:用于描述实体的特征和性质,如人物的年龄、性别等。
- ③ 关系:表示实体之间的关联,是知识图谱的核心组成部分。

(3) 知识图谱的关键技术

① 知识抽取

实体抽取:从文本中识别出命名实体,如人名、地名、组织机构名等。常用的方法有基于规则的方法、基于统计的方法和深度学习方法。

关系抽取:确定实体之间的语义关系,如因果关系、并列关系等。主要方法包括基于模板的方法、监督学习方法和无监督学习方法。

属性抽取:提取实体的属性信息,如人物的身高、体重等。通常采用基于规则和机器学习相结合的方法。

② 知识表示

符号表示:以逻辑符号和规则来表示知识,如谓词逻辑、语义网络等。

向量表示:将知识图谱中的实体和关系映射到低维向量空间中,通过向量的运算来表示知识之间的关系。常用的模型有 TransE、TransR 等。

③ 知识融合

实体对齐:将不同来源的知识图谱中的相同实体进行匹配和合并,消除实体的歧义性。

知识合并:将多个知识图谱中的知识进行整合,形成一个更大、更完整的知识图谱。

④ 知识推理

基于规则的推理:根据预先定义的规则,从已知的知识中推导出新的知识。

基于机器学习的推理:利用机器学习算法,如神经网络、决策树等,从数据中学习推理模型,实现知识的推理。

(4) 知识图谱的应用场景

① 智能问答系统

知识图谱能够为智能问答系统提供丰富的知识支持,使得系统能够准确理解用户的问

题,并从知识图谱中快速检索到相关的答案。例如,在智能客服中,知识图谱可以帮助客服人员快速回答用户的问题,提高服务效率和质量。

② 信息检索

通过将知识图谱与搜索引擎相结合,能够实现语义检索,提高检索结果的准确性和相关性。用户输入的查询语句不再仅仅是基于关键词的匹配,而是基于知识图谱中的语义理解,从而返回更符合用户需求的结果。

③ 推荐系统

知识图谱可以为推荐系统提供丰富的用户和物品信息,通过分析用户的兴趣偏好和物品之间的关系,实现个性化的推荐。例如,在电商推荐中,知识图谱可以帮助推荐系统更好地理解用户的购买行为,推荐更符合用户需求的商品。

④ 医疗领域

知识图谱可以用于疾病诊断、药物研发、医疗决策支持等。例如,通过构建疾病知识图谱,医生可以更准确地诊断疾病,制定治疗方案;药物研发人员可以利用知识图谱挖掘药物的潜在作用机制,加速药物研发进程。

(5) 知识图谱的未来发展趋势

① 多模态知识图谱

将文本、图像、音频、视频等多种模态的数据融合到知识图谱中,实现更全面、更丰富的知识表示和推理。

② 知识图谱与深度学习的深度融合

进一步探索知识图谱与深度学习的结合方式,利用知识图谱的先验知识来指导深度学习模型的训练,提高模型的性能和可解释性。

③ 知识图谱的可解释性研究

随着知识图谱在各个领域的广泛应用,其可解释性问题越来越受到关注。未来需要研究如何使知识图谱的推理过程和结果更加透明、可解释,以满足用户对解释性的需求。

④ 知识图谱在新兴领域的应用拓展

不断探索知识图谱在物联网、区块链、量子计算等新兴领域的应用,为这些领域的发展提供新的技术支持和解决方案。

1.4.6 机器人技术

(1) 运动控制

使机器人能够实现各种动作和任务,涉及电机控制、动力学建模、轨迹规划等技术。如工业机器人通过精确的运动控制完成焊接、装配等任务。

(2) 传感器融合

将机器人的多种传感器(如摄像头、激光雷达、超声波传感器等)的数据进行融合处理,使机器人能够更全面、准确地感知周围环境。如移动机器人通过传感器融合,实现自主导航和避障。

1.5 思考题

1. 比较 AI 的四种定义方法,结合图灵测试论证现代 AI 的智能特征。
2. 简述 AI 发展三时期的关键突破,并分析其演进动力。
3. 以 Google Health 或 IBM Watson 为例,评价 AI 在医疗诊断中的效率提升与职业替代风险。
4. 解释卷积神经网络(CNN)在图像识别中的原理,并与循环神经网络(RNN)对比在处理序列数据方面的优劣。
5. 设计知识图谱在智能客服或医疗中的应用场景,分析其决策价值及多模态融合挑战。



【微信扫码】

本章相关资源

南京大学出版社
版权所有

第 2 章

智能医学

2.1 智能医学定义

智能医学起源于计算机技术在医学领域的辅助应用,如早期的医学影像和诊疗专家决策系统。随着人类社会从机器时代、电气时代迈入信息时代,互联网技术与人工智能及其核心技术(机器学习与深度学习算法等)迅速发展,传统医学也逐渐迈入了智能医学时代。智能医学可视为电子、计算机、互联网及物联网、区块链、云计算、人工智能、3D 打印、虚拟现实、人机交互等技术与医学影像学、生物医学、医学检验学、医学信息学、医学诊断学等领域紧密交融的一门新型医、理、工学交叉学科,其研究方向包含了医学感知、医学大数据智能分析、医学智能决策、精准医疗、医学智能人机交互等,被广泛部署于神经工程、康复工程、组织工程、基因工程、智能医学仪器、智能远程医疗、智能医学教育、新药研发、智能医学图像分析、智能诊疗、智能手术、精准放疗等领域。当前机器还远未达到自主思考解决问题的能力阶段,且在各应用领域须考虑医学伦理、个性化诊断、精准化治疗等多方面因素,因此,当前智能医学依然是人类智能为主导,人工智能作为辅助模式运作。未来,随着技术的不断进步和应用的深入拓展,智能医学有望为医学领域带来更多的创新和突破,成为推动医学发展的重要动力引擎。

智能医学的特征在于“信息技术+医学”的深度融合与协同发展。它并非简单地以信息技术取代医学,而是两者相互促进、共同发展。智能医学广泛融合了人工智能、虚拟现实、增强现实、医疗大数据、机器人、可穿戴医疗设备、云平台、远程医疗、区块链、计算机辅助手术导航和 3D 打印等众多前沿科技领域。在智能医学体系中,“智能”不仅包含了人工智能,还涵盖了更广泛的智能技术。智能作为手段,医学作为目的,两者在智能医学的发展中互相补充,缺一不可。

智能医学的核心理念是“交叉融合”。通过众多学科的前沿技术与医学的密切融合,智能医学能够解决现代医学发展中的一系列难题。它为医生和技术人员在疾病的诊治方面提供了全新的思路,极大地促进了医学的进步。

2.2 智能医学主要研究方法

智能医学作为一门跨学科领域,结合了医学、机器学习、数据处理等多个领域的知识和技术。其主要研究方法可以归纳为医学影像处理与分析、生物信息学研究、医疗机器人与智能设备应用、智能医疗监测与数据挖掘、跨学科融合研究等方面。

2.2.1 医学影像处理与分析

医学影像中的智能方法主要是指应用人工智能技术,特别是机器学习和深度学习技术,对医学影像进行分析、处理、诊断和预测的一系列方法。

(1) 识别与分类

人工智能能够快速准确地识别医学影像中的特定结构或异常,如肿瘤、骨折、肺结节等,并进行分类。这种方法可以辅助医生进行早期诊断,提高诊断的准确性和效率。

(2) 辅助诊断与预测

基于大量历史病例的学习,人工智能系统可以辅助医生做出更准确的诊断。例如,在肺癌筛查中,AI 可以帮助识别肺部 CT 扫描中的小结节,预测是否为恶性。此外,AI 还能对疾病的发展趋势进行预测,为医生制定预防和治疗策略提供依据。

(3) 量化分析

人工智能可以用于对医学影像数据进行量化分析,比如测量肿瘤大小的变化、评估治疗效果等。这种方法可以为医生提供更为精准的诊断依据,有助于制定个性化的治疗方案。

(4) 影像分割与定位

医学影像分割是将图像中的特定区域或结构提取出来的过程,对于疾病诊断和治疗方案的制定具有重要意义。人工智能,特别是深度学习算法,能够实现高精度的影像分割,帮助医生准确分割出病变区域,为手术提供精确指导。同时,AI 还能对病变位置进行精确定位,为放射治疗等提供重要依据。

(5) 个性化治疗建议

结合患者的临床信息和其他检查结果,人工智能能够提供个性化的治疗方案建议,帮助医生选择最适合患者情况的治疗方法。这种方法可以提高治疗效果,减少不必要的医疗干预,降低医疗成本。

(6) 远程医疗服务

人工智能还促进了远程医疗服务的发展,使得偏远地区的患者也能获得高质量的诊断服务。通过云端实时分析影像数据,AI 可以为偏远地区的医生提供诊断支持,提高医疗服务的可及性和质量水平。

(7) 跨学科融合与创新

医学影像中的智能方法还涉及跨学科融合与创新。例如,量子成像、光声成像等新型成像技术将与 AI 深度融合,为医生提供更清晰、更详尽的人体内部结构图像。同时,整合生

物医学工程、计算机科学等多个领域的资源和技术优势,共同推动医学影像技术的发展和

创新。
综上所述,医学影像中的智能方法包括图像识别与分类、辅助诊断与预测、量化分析、影像分割与定位、个性化治疗建议、远程医疗服务以及跨学科融合与创新等多个方面。这些方法的应用极大地提升了医学影像诊断的准确性和效率,为现代医学诊断带来了革命性的变化,主要体现如下:

① 技术基础。利用计算机视觉和机器学习技术。

② 应用实例。如人工智能算法辅助医生识别肺部 CT 扫描中的微小结节,提高肺癌的早期诊断率。据统计,利用人工智能辅助诊断,医生的诊断准确率可以提高 10%~20%。此外,该技术还广泛应用于病灶定位、病理分析等领域。

③ 前沿方向。深度学习与影像组学,即利用深度学习技术进行影像分析,探索疾病的影像特征。

2.2.2 生物信息学研究

生物信息学中的智能方法主要是指应用人工智能技术,包括机器学习、深度学习等,对生物数据进行处理、分析和解释的一系列方法。

(1) 基因组分析

通过比较不同种类的基因组序列,人工智能可以帮助研究人员找到共同的基因和基因变异,从而更好地理解生物进程和疾病发生机制。具体来说,机器学习算法可以用于预测基因功能、识别基因变异等任务;而深度学习算法,如卷积神经网络(CNN),则可以用于更复杂的基因序列分析和变异检测。

(2) 蛋白质结构预测

人工智能通过分析蛋白质序列来预测蛋白质的三维结构,从而帮助研究人员更好地理解蛋白质的功能。传统的蛋白质结构预测方法通常需要大量的计算资源和时间,且准确度不高。而利用机器学习算法,特别是结合深度学习技术,可以显著提高蛋白质结构预测的准确性和效率。例如,AlphaFold 等深度学习模型已经在蛋白质结构预测方面取得了突破性进展。

(3) 药物研发

人工智能在药物研发领域的应用也非常广泛。通过机器学习算法,研究人员可以识别潜在的药物靶点,预测药物的活性和毒性,从而加快药物研发过程。此外,深度学习算法还可以用于药物分子的生成和优化,以及药物与靶标相互作用的模拟和预测。这些智能方法的应用可以显著降低药物研发的成本和时间,提高药物研发的成功率。

(4) 个性化医疗

人工智能通过分析个体的基因、环境因素和生活方式等信息,可以为个体提供个性化的医疗建议和治疗方案。这种方法结合了生物信息学和人工智能的优势,可以根据个体的具体情况制定最适合的治疗计划,从而提高治疗效果和患者的生活质量。

(5) 多组学数据整合分析

多组学数据整合分析是指结合两种或者两种以上的组学数据集来研究生物系统中多种

物质之间的相互作用。人工智能方法,特别是机器学习和深度学习算法,在多组学数据整合分析中发挥着重要作用,可以从大量且繁杂的多组学数据中找到它们之间的内在联系,帮助人们全面地认识生命系统,对研究生命科学问题具有重要意义。

(6) 可解释人工智能方法

为了提高人工智能在生物信息学中的可信度和透明度,研究人员还开发了各种可解释人工智能方法。这些方法包括基于规则的方法、模型可视化方法、特征重要性方法等。它们可以帮助研究人员理解机器学习算法在生物数据中的应用过程,从而提高算法的精确度和可靠性。例如,LIME(Local Interpretable Model-agnostic Explanations)是一种用于解释机器学习算法预测结果的工具,它已经被应用于基因组学、蛋白质预测和药物发现等领域。

综上所述,生物信息学中的智能方法涵盖了基因组分析、蛋白质结构预测、药物研发、个性化医疗、多组学数据整合分析以及可解释人工智能方法等多个方面。这些方法的应用极大地推动了生物信息学的发展和创新,为生命科学研究和医学实践提供了强有力的支持,主要体现如下:

- ① 研究内容。主要研究生物数据的获取、存储、分析和解释等技术。
- ② 应用实例。通过大规模基因组测序和数据分析,生物信息学可以帮助科学家揭示疾病的遗传基础和发病机制。例如,在病毒的应对研究中,生物信息学技术便可发挥重要作用,可帮助科学家快速解析病毒基因组序列,为疫苗研发和药物治疗提供关键信息。
- ③ 前沿方向。机器学习与生物信息学分析,即利用机器学习进行基因和蛋白质组学等生物信息学分析。

2.2.3 医疗机器人与智能设备应用

医疗机器人与智能设备应用中的智能方法涉及多个方面,这些方法的应用极大地提升了医疗服务的效率和质量。

(1) 图像识别与处理

① 深度学习算法。医疗机器人和智能设备利用深度学习算法,如卷积神经网络(CNN)等,可对医学影像进行自动识别和分类,有助于快速准确地诊断疾病,减少医生的诊断负担,提高诊断效率。

② 图像增强与恢复。针对医学影像中存在的噪声、模糊等问题,智能方法通过图像增强与恢复技术,可提高影像的清晰度和对比度,从而让医生更易于观察和诊断。

(2) 精准操作与辅助手术

① 高精度定位与导航。医疗机器人利用先进的传感器和定位技术,实现高精度的手术操作。例如,集成有 AI 模块的达·芬奇手术机器人可以实现亚毫米级的精确切割和缝合。

② 智能辅助决策。通过整合病患历史数据、基因组学信息以及最新的科研成果,医疗机器人能够为临床决策提供智能化建议。在制定治疗方案时,AI 能够根据患者的具体病情匹配最优的治疗策略。

(3) 远程监控与诊疗

① 物联网技术。医疗机器人和智能设备通过物联网技术,实时采集患者的生命体征数据,并运用 AI 算法分析病情发展趋势,为远程会诊和紧急干预提供实时信息支持。

② 虚拟现实与增强现实。结合虚拟现实(VR)和增强现实(AR)技术,医疗机器人可以为医生提供沉浸式的手术模拟和远程操作体验,提高手术的准确性和安全性。

(4) 个性化治疗与康复

① 个性化治疗方案。通过分析患者的基因数据、生理参数等信息,医疗机器人能够制定个性化的治疗方案,提高治疗效果和患者的生活质量。

② 智能康复训练。康复机器人能够根据患者的康复进展和个体差异动态调整康复计划,通过传感器收集患者的动作数据,机器人可实时提供个性化的康复训练,量化评估恢复情况。

(5) 智能药物管理与配送

① 药物筛选与模拟。AI技术可以加速新药的研发过程,降低研发成本;通过AI驱动的药物筛选和模拟,可以预测药物的活性和毒性,优化药物结构。

② 智能药物配送。AI驱动的护理机器人可以协助护士执行常规药物配送等工作,减少医护人员的工作负担;同时,机器人还能确保药物在正确的时间、以正确的剂量送达患者手中。

(6) 语音交互与智能助手

① 语音识别与合成。通过语音识别技术,医疗机器人和智能设备可以实现与患者的语音交互,提供健康咨询、用药提醒等服务,提高患者的自我管理能力和健康意识。

② 智能虚拟助手。基于AI的虚拟助手可以为医生提供手术指导、病历管理等支持,同时还可为患者提供预约挂号、检查结果查询等便捷服务。

综上所述,医疗机器人与智能设备应用中的智能方法涵盖了图像识别与处理、精准操作与辅助手术、远程监控与诊疗、个性化治疗与康复、智能药物管理与配送以及语音交互与智能助手等多个方面。这些方法的应用不仅提高了医疗服务的效率和质量,还为患者带来了更加便捷、个性化的医疗体验,主要体现如下:

① 研究内容。主要研究医疗机器人的设计、控制和应用等技术。

② 应用实例。医疗机器人可以辅助医生进行手术操作、康复训练等任务,提高医疗服务的效率和质量。例如,达·芬奇手术机器人已经广泛应用于微创手术领域,可以减少手术创伤和并发症风险。

2.2.4 智能医疗监测与数据挖掘

智能医疗监测与数据挖掘中的智能方法涉及多个关键技术和算法,这些方法的应用极大地提升了医疗服务的智能化水平。

(1) 智能医疗监测中的智能方法

① 可穿戴设备与远程监测

患者佩戴的可穿戴设备(如智能手表、健康监测手环等)能够持续监测并收集患者的生理数据,如心率、血压、血氧饱和度等。这些设备通常具有无线连接功能,能够将数据实时传输到云端或医生的终端设备上,实现远程监控。AI算法对收集到的数据进行实时分析,能够自动识别异常数据并触发预警机制,及时通知患者和医生采取相应措施。

② 智能家居系统与健康监测

智能家居系统可以集成各种健康监测设备,如智能体重秤、智能血压计等。这些设备通

过智能家居平台与 AI 系统相连,实现数据的自动上传和分析。AI 通过分析这些数据,可以评估患者的整体健康状况,并提供个性化的健康建议。

③ 医疗传感器与实时数据分析

医疗传感器可以放置在患者体内(如植入式传感器)或体外(如皮肤贴片传感器),持续监测患者的生理指标。这些传感器通过物联网技术与 AI 系统相连,实现数据的实时传输和处理。AI 系统能够对传感器收集到的数据进行深度分析和挖掘,发现数据中的异常模式和潜在风险,为患者提供及时的健康预警。

(2) 数据挖掘中的智能方法

① 机器学习是数据挖掘的核心技术之一,在医疗数据挖掘中起着至关重要的作用。通过机器学习算法,可以从大量的医疗数据中发现潜在的模式和规律,为医疗决策提供支持。机器学习方法包括监督学习、无监督学习和强化学习等,广泛应用于疾病预测、患者分类、药物反应预测和个性化治疗等领域。

② 自然语言处理(NLP)技术在医疗数据挖掘中尤为重要。因为大量医疗数据以文本形式存在,如医生笔记、病历记录等。通过 NLP 技术,可以将这些非结构化数据转化为结构化数据,便于进一步分析。NLP 技术包括文本分类、情感分析、实体识别等,在医疗数据挖掘中发挥着重要作用。

③ 关联规则挖掘是一种用于发现数据集中变量之间有趣关系的方法。在医疗数据挖掘中,关联规则挖掘广泛应用于药物与疾病的关联、症状与疾病的关联以及治疗方案与疗效的关联等方面的发现。

④ 聚类分析是一种将数据分成不同组的方法。在医疗数据挖掘中,聚类分析可以用于患者群体分类、疾病亚型识别等。通过聚类分析,可以发现数据中的隐藏模式和结构,为医疗决策提供支持。

⑤ 时间序列分析是一种统计方法,主要用于分析和预测时间序列数据。在医疗数据挖掘中,时间序列分析可以用于疾病流行趋势预测、患者病情监测等,可以提高医疗决策的及时性和准确性。

⑥ 数据可视化是将复杂的医疗数据转化为直观的图形和图表的过程。通过数据可视化技术,可以帮助医疗从业者快速发现数据中的异常和趋势,提高决策效率。常用的数据可视化方法包括热图、散点图、折线图和饼图等。

综上所述,智能医疗监测与数据挖掘中的智能方法涵盖了多个关键技术和算法。这些方法的应用不仅提升了医疗服务的智能化水平,还为患者带来了更加精准、个性化的医疗体验。随着人工智能和大数据技术的不断发展,智能医疗监测与数据挖掘的应用前景将更加广阔,主要体现在如下:

① 技术基础。结合物联网、大数据等技术。

② 应用方向。通过实时监测患者的生理指标,及时发现异常情况并采取措施;同时,从海量医疗数据中挖掘出有用信息,为医疗决策提供支持。

2.2.5 跨学科融合研究

跨学科中的智能研究方法是一种综合性的研究范式,它结合了多个学科的知识和技术,利用智能技术来推动跨学科研究的深入和发展。

(1) 核心要素

① 跨学科整合。智能技术作为连接不同学科的桥梁,能够促进学科间的交叉融合,其研究的核心在于整合不同学科的知识和方法,形成新的研究视角和解决方案。

② 智能技术应用。人工智能、机器学习、数据挖掘等智能技术在跨学科研究中发挥着重要作用,能够处理和分析复杂的数据,揭示数据背后的规律和模式,为跨学科研究提供有力支持。

(2) 主要方法

① 文献综述与比较分析。广泛收集多学科文献,确定与研究主题相关的多个学科领域,通过图书馆、学术数据库等渠道全面收集文献资料。仔细阅读并分析收集到的文献,梳理不同学科对研究问题的观点、方法和结论。比较不同学科观点的异同,找出它们之间的联系和冲突,综合提炼出跨学科研究的思路和问题解决方案。

② 案例研究与多学科视角分析。寻找具有跨学科特征的实际案例,确保案例具有代表性和典型性,能够涵盖多个学科的知识和问题。组织来自不同学科的研究人员组成团队,对案例进行深入分析。每个学科的研究人员从自己的专业角度出发,分析案例中的问题、解决方案和影响。通过团队讨论和交流,整合不同学科的观点,形成对案例的全面认识,并总结跨学科合作的经验和教训。实验设计与多学科团队合作确定研究问题和假设,结合多个学科的理论和方法,设计实验方案。实验设计要考虑不同学科的需求和限制,确保实验的可行性和有效性。组建由不同学科的研究人员组成的实验团队,共同实施实验。每个学科的研究人员负责实验的特定部分,团队成员之间密切合作,及时沟通和协调实验过程中的问题。收集和分析实验数据,从不同学科的角度解读实验结果,综合不同学科的分析结果,得出跨学科的结论和建议。模型构建与验证明确研究问题和目标,确定模型中需要包含的跨学科要素。收集相关学科的理论 and 数据,为模型构建提供基础。选择合适的模型构建方法,如系统动力学模型、生态经济模型等,将不同学科的要素整合到模型中,建立跨学科模型框架。对模型进行验证和调整,确保模型的准确性和可靠性。将验证后的模型应用于实际问题的解决,为跨学科研究提供决策支持。

(3) 应用场景

① 智能医疗监测与数据挖掘。结合计算机科学、医学、生物学等多个学科的知识和技术,利用智能技术进行医疗数据的收集、分析和挖掘,为医疗决策提供支持。

② 生成式人工智能在跨学科教学中的应用。利用生成式人工智能创建虚拟实验和模拟环境,支持跨学科实践和探索,提高学生的学习效率 and 创新能力。

综上所述,跨学科中的智能研究方法是一种综合性的研究范式,但跨学科中的智能研究方法也面临诸多挑战,如数据整合与挖掘的复杂性、模型构建的准确性等。不过随着人工智能和大数据技术的不断发展,跨学科中的智能研究方法将不断完善和拓展,为更多领域的跨学科研究提供有力支持,在多个领域都将有广泛的应用前景,为跨学科研究提供了新的思路和方法,主要体现如下:

① 研究趋势。随着技术的不断发展,智能医学的研究方法也在不断创新和拓展。例如,深度学习与全景病理组学结合,进行全方位的病理切片分析;大语言模型与医疗文本挖掘结合,预测疾病发展;深度学习与序列数据建模结合,进行脑机接口、心电图、脑电图等序

列数据的建模等。

② 前沿方向。深度学习与多模态融合,涉及影像多模态、文本影像多模态、影像临床多模态等多种模式的融合研究;机器学习与多组学建模,结合基因组、转录组、蛋白组、代谢组、病理组等多组学数据进行建模分析。

2.3 智能医学国内外发展现状

智能医学作为人工智能与医学深度融合的新兴领域,近年来在国内外都取得了显著进展,正深刻改变着医疗行业的格局。

2.3.1 国内发展现状

我国智能医疗行业近年来呈现爆发式增长,技术快速进步,产业也是蓬勃发展。

(1) 技术快速进步

① 医疗影像与辅助诊断。国内企业和科研机构在医疗影像识别、辅助诊断等领域取得了长足进步。一些本土研发的智能影像诊断产品,基于自主研发的深度学习算法,已获得国家药品监督管理局批准上市,并在各级医疗机构中逐步推广应用。例如,某企业开发的智能肺部影像诊断系统,能够快速准确地识别肺部疾病,包括肺炎、肺结核、肺癌等,为医生提供诊断参考,提高了诊断效率和准确性。

② 远程医疗。随着 5G 技术的快速发展,国内实现了高清、实时的远程会诊和手术指导。偏远地区的患者通过远程医疗平台,能够与大城市的专家进行面对面交流,获得精准的诊断和治疗建议。在一些紧急救援场景中,5G 远程医疗技术可以让专家在第一时间对患者进行评估和指导,为患者的救治争取宝贵时间。例如,在山区发生的突发外伤事件中,当地医护人员可以通过远程医疗平台,在专家的指导下对患者进行紧急处理,随后再将患者转运至条件更好的医院进行进一步治疗。

(2) 产业蓬勃发展

① 产业集聚。以北京、上海、深圳为代表的城市,凭借其丰富的科技资源、完善的产业链和优秀的产业环境,聚集了大量智能医学企业,涵盖医疗机器人、智能诊断、健康管理等多个细分领域。北京依托高校和科研机构的人才优势,在智能医学的基础研究和技术创新方面表现突出;上海凭借其国际化的金融中心地位和完善的医疗资源,吸引了众多智能医学企业的总部和研发中心;深圳则以其强大的电子信息产业基础,在医疗硬件设备和智能可穿戴设备的研发生产上具有优势。

② 创新成果。一些企业在智能康复机器人、智能可穿戴设备等方面取得了创新性成果。例如,某企业研发的智能康复机器人,能够根据患者的康复需求和身体状况,提供个性化的康复训练方案。通过传感器实时监测患者的运动数据,机器人可以自动调整训练强度和方式,帮助患者更好地恢复身体功能。这些产品不仅满足了国内市场的需求,还凭借其高性价比和良好的性能,逐步走向国际市场。

③ 产学研合作紧密。国内医疗机构也积极与科技企业合作,开展智能医学的临床研究和应用实践。例如,一些大型医院与人工智能企业合作,利用医院的临床数据和病例资源,

开发针对特定疾病的智能诊断和治疗系统。通过临床实践不断优化技术,提高产品的可靠性和实用性,推动智能医学技术的不断发展。

④ 政策大力扶持。我国政府高度重视智能医学的发展,发布了一系列政策文件,如《“十四五”国民健康规划》《“互联网+医疗健康服务”行动计划》《关于进一步完善医疗卫生服务体系的意见》等,明确提出加快推进人工智能在医疗卫生领域的应用。政府还通过设立专项科研基金,鼓励高校、科研机构和企业开展智能医学相关的科研项目。同时,建设产业园区,为智能医学企业提供良好的发展环境,促进智能医学产业的集聚发展。例如,在一些产业园区内,政府提供了场地租赁优惠、税收减免等政策支持,吸引了众多智能医学企业入驻,形成了产业集群效应。

2.3.2 国外发展现状

(1) 技术应用广泛且深入

① 医疗影像诊断。欧美发达国家凭借先进的深度学习算法,使智能诊断系统在医疗影像分析上成绩斐然。美国的一些智能诊断工具,基于卷积神经网络(CNN)技术,能够对X光、CT、MRI影像进行细致分析。例如,通过对大量肺癌患者的CT影像数据进行学习,系统可以精准识别出早期肺癌的微小结节,其准确率甚至超越了部分经验丰富的医生,极大地提高了疾病早期诊断的准确性,为患者争取到宝贵的治疗时间。

② 手术治疗。手术机器人技术日臻成熟,以达·芬奇手术机器人为代表,已在全球范围内广泛应用于泌尿外科、妇产科、心胸外科等多种复杂手术。利用其高自由度的机械臂,通过主从式操作系统,医生在控制台就能精确控制机械臂的动作,实现毫米级别的精准操作。在前列腺癌根治术中,达·芬奇手术机器人能够更清晰地分辨组织,减少对周围神经和血管的损伤,有效降低了术后并发症的风险,提升了患者的康复效果。

(2) 产业生态完善

① 科技巨头引领。美国凭借强大的科技实力和成熟的资本市场,在智能医学产业独占鳌头。谷歌、微软等科技巨头纷纷布局,通过收购相关初创企业来获取前沿技术。谷歌旗下的DeepMind公司开发的AI系统,不仅能通过分析患者的电子病历、影像数据等预测病情发展,还能根据患者的个体差异,优化治疗方案。例如,在慢性肾病的治疗中,该系统可以根据患者的肾功能指标、用药历史等数据,预测疾病恶化的可能性,并给出个性化的治疗建议。

② 初创企业活力创新。美国拥有众多专注于智能医学的初创企业,它们在细分领域不断创新,形成了从技术研发、产品生产到临床应用的完整产业链。如专注于智能可穿戴设备的企业,研发出能够实时监测人体生理指标(如心率、血压、睡眠质量等)的设备,并通过数据分析为用户提供健康预警和个性化的健康管理方案。这些初创企业与高校、科研机构紧密合作,加速技术的转化和应用。

③ 欧洲多元发展。欧洲各国也在积极发展智能医学。德国在医疗机器人研发制造方面,凭借其先进的工业制造技术和严谨的科研态度,在手术机器人、康复机器人等领域成果显著。英国则依托其生物医学、数学和计算机科学方面的优势,在医疗大数据分析和人工智能算法研究上表现突出,致力于挖掘医疗数据背后的潜在价值,为疾病预防和治疗提供决策支持。

④ 政策支持力度大。美国政府出台了一系列政策鼓励智能医学发展。在科研经费方面,美国国立卫生研究院(NIH)设立专项基金,资助智能医学相关的科研项目,促进基础研究和临床应用的转化。同时,美国食品药品监督管理局(FDA)简化了医疗器械审批流程,加快了智能医学产品的上市速度。欧盟制定了相关战略规划,推动成员国在智能医学领域的合作与发展,通过建立统一的技术标准和数据共享机制,打破了区域间的技术壁垒,促进了智能医学技术在欧洲的广泛应用。

2.3.3 国内外发展对比、问题挑战及未来发展趋势

智能医学通过 AI、物联网等技术重塑医疗模式,国内外在政策、技术和市场上呈现差异化竞争。我国凭借数据规模和政策红利快速追赶,而欧美在基础研究与高端设备上保持优势。与此同时,智能医学在国内外发展过程中也面临一些挑战,人工智能算法的可解释性不足,在临床应用中可能引发信任危机,未来需突破技术瓶颈;医疗信息跨机构共享困难,各国监管政策和标准尚未完全统一,影响技术的跨境应用和产业合作,需要不断完善法规,加强全球协作,以实现医疗资源的普惠化与智能化。智能医学发展前景广阔,随着技术的不断突破和应用场景的持续拓展,有望为全球医疗健康事业带来更多变革。

(1) 国内外技术差距与我国优势

① 差距。我国在核心算法和高精度医疗设备上仍依赖进口,高端产品国产化率不足。

② 优势:中国在数据规模(医学影像年增长率 30%)和应用场景多样性(如基层医疗覆盖)上领先。

(2) 共性问题与挑战

① 数据质量及隐私安全。数据质量参差不齐,数据隐私保护和安全问题突出,全球面临医疗数据泄露风险,需加强加密技术与法规建设。

② 人才缺口。专业人才缺口明显,尤其是兼具医学与工程背景的专业人员,制约着产业发展。我国放射科医生年增长率仅 1%,远低于影像数据增速,亟须复合型人才。

③ 伦理争议。AI 决策的透明性与责任归属问题引发讨论,需建立跨学科伦理框架。

(3) 未来发展趋势

① 技术深度融合。多模态大模型(整合文本、影像、基因数据)将成为主流,推动精准医疗和药物研发效率提升;脑机接口、数字药物等前沿技术或进入临床阶段。

② 政策与市场驱动。我国“健康中国 2030”战略将进一步开放医保支付范围,支持远程医疗纳入报销体系;欧美则通过公私合作模式加速技术转化。

③ 全球化协作。跨国企业如西门子、GE 医疗等与中国企业合作开发智能设备,共享数据与标准,推动全球医疗资源均衡分配。

2.4 智能医学的应用

作为现代医学发展的重要方向,智能医学不仅拓展了传统医学的边界,也为医疗行业带来了革命性的变革,其应用正在逐步改变传统的医疗模式,显著提高了医疗服务的效率和质量。

2.4.1 医学影像分析

AI 在医学影像分析中的应用已经取得了显著进展,尤其是在图像识别、病灶检测和辅助诊断方面,结合多种医学影像(如 CT、MRI、PET 等)进行综合分析,通过深度学习算法,能够快速、准确地分析大量的医学影像数据,帮助医生发现早期病变,提高诊断的准确性和效率。谷歌旗下的 DeepMind 开发的 AI 系统,在临床试验中表现出色,能够减少假阳性和假阴性的发生率,显著提高了肺癌早期筛查的准确性。美国麻省理工学院(MIT)的研究团队开发了一种 AI 模型,能够通过分析乳腺 X 光片,预测女性未来 5 年内患乳腺癌的风险,该模型在多个数据集上进行了验证,显示出比传统方法更高的预测准确性。IBM Watson Health 开发的多模态 AI 系统能够结合 CT 和 PET 影像,帮助医生更准确地定位肿瘤和评估其恶性程度。

2.4.2 3D 打印技术

3D 打印技术是将数据处理转换后建立三维模型,传递到 3D 打印机,从而快速制造出复杂形状的 3D 物体。目前 3D 打印技术在医学领域的研究和应用主要在医学模型、个性化植入物、手术导板等方面,涉及整形外科、颌面外科、神经外科和心血管外科等多个学科。医学模型的打印可以用于医学教学、医学诊断、术前模拟等方面。可吸收材料在个性化植入物方面的应用,主要集中在人体组织(如骨骼、气管等)的修复重建上,包括 3D 打印的金属假体和生物材料植入物。通过 3D 打印出的个性化骨骼,与患者自身的骨骼形态更加吻合一致,匹配性更好。同时可控地设计假体内部孔隙结构,为种子细胞生长提供有利的空间,促进假体与自身骨组织的融合。比利时 BIOMED 研究所根据下颌骨缺损患者术前的缺损尺寸和形状,使用 3D 打印技术打印了一个与缺损形状完全一致的下颌骨并植入,成功恢复了下颌正常功能。西京医院郭征、孟国林等团队,利用 3D 打印技术成功设计并应用了个性化定制的骨盆、肩胛骨、跟骨、锁骨、肱骨等多种骨骼金属假体,用于肿瘤切除或严重创伤后所致大段骨缺损的修复重建,在实现假体牢固固定的同时,又可保证金属假体置入的准确度,显著提升了手术的整体疗效。

随着 3D 打印技术的不断发展和推广,其应用领域逐渐扩大,目前在肝脏、肾脏等重要器官的生物 3D 打印方面也取得了重要进展。虽然 3D 打印技术尚有一些问题限制了其临床应用,如打印材料性能受限、打印精度不足、打印价格昂贵等,但其医学研究和应用前景仍十分光明,在未来将推动医学的革新和进步。

2.4.3 医用机器人

医用机器人是设计用于从事医疗或辅助医疗工作的智能型装备。按其用途的不同,可分为手术机器人、康复机器人、辅助机器人、教学机器人及服务机器人五类。目前医用机器人的发展十分迅速,相信随着科技的进步,将进一步深入各个医疗领域,发挥越来越重要的作用。

(1) 手术机器人

目前医学应用的热点,它具有人类无法实现的精准度,通过机械化智能化的设备,实现了以往外科手术难以达到的精准性和安全性。第一个被 FDA 认证通过的手术机器人罗伯

特可(ROBODOC)于1989年诞生,它被设计应用于髋关节置换术。1997年美国直觉外科手术公司发明的达·芬奇手术机器人,是目前国内引进应用最为广泛的智能手术系统,能够



图 2.1 “天玑”骨科导航机器人

通过 AI 技术辅助医生进行微创手术;该系统能够提高手术的精确度和稳定性,减少手术创伤和术后并发症,在胃肠外科、妇产科、泌尿外科等多个科室广泛应用。近年来,国外开始研制更小型的外科手术机器人,如2012年美国华盛顿大学研制出一批带有鸟翼状机械臂的“乌鸦”医用机器人,相比达·芬奇机器人更加小巧、价格更低,并且可以开源代码,根据需求扩展其功能。国内在机器人的研制上也取得一定成果,比如我国第一台自主研发的“天玑”骨科导航机器人已经成功用于临床,取得了良好的临床应用效果,如图 2.1 所示。

(2) 康复机器人

可以辅助或替代机体功能,以达到帮助患者恢复或日常活动的目的。康复机械手是其中的代表,如犹他州/麻省理工学院(Utah/MIT)人工臂,可模拟人手多个轴向的活动;汉迪1(Handy1)床边机器人,是销量最大的康复机器人之一,可实现帮助喂食、洗漱、化妆等功能;移动智能轮椅如惠勒斯利(Wheelesley)机器轮椅系统,能获取外界信息并自动规划移动路线。

(3) 辅助机器人

目前包括救援机器人和教学机器人,救援机器人可以自主对人体进行生命体征的检测、分析判断及救援。比如国际救援系统研究所研制的救援机器人 T-25 恩宇(T-52 Enryu),可以实现伤员的定位寻找及障碍物的清理。

(4) 教学机器人

通过模拟临床实际情况,让学员有感官上的直接感受,实现教学的目的。如美国研发的诺埃尔(Noelle)教学机器人通过模拟待产的孕妇,用于产科的教学;我国现在多家医院也已采用多种机器人模型用于教学或培训,如进行胸外按压、气管插管及各种腔镜训练等,取得了良好的教学效果。

(5) 服务机器人

是指服务于医院、诊所或提供健康服务等作用的智能机器人,比如日本机械工程研究所开发的护理机器人迈空(MELKONG),能够实现平托患者将并之送往目的地;美国研发的帮手机器人(Help-mate Robot),可以实现在医院范围内定点投送食品和药品;英国研发的 PAM 机器人,可以帮助搬运瘫痪患者。

2.4.4 可穿戴医疗设备

可穿戴医疗设备目前在医疗领域发展迅速,出现了多种多样的产品和技术。可穿戴医疗设备可以直接穿戴在人体,方便携带,通过软件和网络支持,测量和记录分析,甚至干预人体的生理状态或疾病进程。目前可穿戴医疗设备根据可实现的功能不同可以分为五种,分别是健康监护、安全监测、康复辅助、疗效评测及疾病监测。

健康监护领域的应用主要是通过可穿戴设备实现对人体日常活动,如心率、活动量等,甚至对心电图、血氧饱和度等更加复杂生理信号的监测,评估其健康状态。安全监测方面的应用可以通过可穿戴设备及时发现异常信息,传递求救信息发送至家人或急救中心。有报道称使用紫蜂(ZigBee)通信技术,可以及时发现老人在室内的不慎摔倒,同时进行定位,迅速通知医疗人员实施救援。在康复方面,可穿戴设备可对患者的康复训练起到督促提醒的作用,并对康复训练活动的全过程进行分析和指正。在疗效评测上,可穿戴设备可以对参加临床随机对照实验的患者进行监督,以获得更加准确、客观的数据,达到更加有效评测新药疗效的目的。在疾病监测上,通过佩戴可穿戴设备,可以对患者的疾病进程和发展进行有效的监测,获得相关数据,从而实现对疾病的防控,避免疾病的恶化。苹果公司的 Apple Watch 通过内置的 AI 算法,能够实时监测用户的心率、血氧水平等健康指标,并在检测到异常时发出警报,帮助用户及时就医,如图 2.2 所示。



图 2.2 实时监测心率的可穿戴式智能手表

目前可穿戴设备在医学的应用方面仍受到多种相关技术水平的限制,比如电池续航技术、信号传输技术、材料技术、传感器技术等。只有各个领域整体技术水平的提高,才能使得可穿戴设备在医学领域的应用更加广泛。

2.4.5 疾病预测与风险评估

AI 通过分析患者的电子健康记录(EHR)、基因组数据、生活方式等多维度信息,能够预测个体患病的风险,并提供个性化的预防建议。这种预测模型在慢性病管理、早期干预和公共卫生政策制定中具有重要应用。英国牛津大学的研究团队开发了一种 AI 模型,能够通过分析患者的电子健康记录,预测未来 10 年内患心血管疾病的风险。该模型在英国国家健康服务(NHS)中得到了广泛应用,可帮助医生识别高风险患者并进行早期干预。美国斯坦福大学的研究人员开发了一种基于机器学习的糖尿病预测模型,能够通过分析患者的血糖水平、体重、家族史等数据,预测个体患 2 型糖尿病的风险。该模型在多个临床研究中表现出色,能够提前数年预测糖尿病的发生。

近年来,AI 技术在基因组学领域的应用日益广泛。例如,谷歌的 DeepVariant 项目利用深度学习技术,能够从基因组测序数据中准确识别出与疾病相关的基因变异,帮助医生进行更精准的疾病风险评估。

2.4.6 远程医疗

远程医疗的研究和应用目前主要包括远程医疗会诊、远程医疗教育咨询、远程医疗监护。国内外对远程医疗的研究投入都很大,出现了多种远程医疗的技术方案和项目应用。

我国首次远程医疗的活动是在 1988 年,由中国人民解放军总医院通过卫星通信与德国

的一家医院进行了一例神经外科的远程病例讨论。1994年,原国家卫生部主导并实施了“金卫工程”2号工程,该工程主要的目的是建设国内远程医疗会诊系统。1997年,中国金卫医疗网络即原卫生部卫生卫星专网正式开通,我国远程医疗进入实际应用阶段。

AI技术在远程诊断、患者监护和健康管理中的应用日益广泛。例如,美国的Teladoc Health公司利用AI技术,能够通过视频会诊和远程监测,为患者提供及时的医疗服务。目前实现远程医疗的技术手段已经比较完善,但受到专家资源、医疗体制等因素的影响,有些相对落后地区仍缺乏建设远程医疗系统的能力和机会。因此,在相对落后地区可更多地投入人力和物力,加大远程医疗的推广应用,真正起到缓解优质医疗资源分配不均、提高落后地区医疗水平的效果。

2.4.7 个性化治疗

AI在精准医疗中的应用主要体现在药物基因组学、治疗方案优化和个性化药物设计等方面。通过分析患者的基因组数据、临床数据和药物反应数据,AI能够为患者提供量身定制的治疗方案,提高治疗效果并减少副作用。IBM Watson for Oncology是一个基于AI的癌症治疗推荐系统,能够通过分析患者的基因组数据和临床数据,为医生提供个性化的治疗方案建议;该系统已经在全球多个癌症治疗中心得到应用,帮助医生制定更精准的治疗计划。美国FDA批准的药物Abacavir是一种用于治疗HIV的药物,其疗效和副作用与患者的基因型密切相关,AI系统能够通过分析患者的基因组数据,预测该患者对Abacavir的反应,从而帮助医生决定是否使用该药物。

近年来,AI技术在药物研发中的应用越来越广泛。例如,英国的Exscientia公司利用AI技术加速了新药的研发过程,成功开发出多种针对癌症和罕见病的个性化药物。AI能够通过分析大量的化学和生物数据,快速筛选出潜在的药物候选分子,大大缩短了药物研发周期。

2.4.8 大数据与云平台

随着经济社会对云计算、云存储、云迁移等服务需求的持续增长,我国的云服务市场正步入成熟发展阶段。其中,医疗行业尤为突出,数据量极大,在处理数据上就遇到了海量数据和非结构化数据的挑战,而近年来很多国家都在积极推进医疗信息化发展,这使得很多医疗机构参与了大数据分析和云平台的构建,有力地推动了我国医疗事业的发展。

随着云技术在医疗领域的应用不断拓展,医疗机构可以使用云平台这种协作方式有效处理和交付数据,并将数据分析成有意义的信息,这使得医疗资源紧张、就医贵、就医难等问题有所缓解。通过使用云计算服务,医疗机构只需为使用的资料和服务支付费用,例如存储、应用程序和基础设施服务,降低了患者就医的成本。

智能医学的应用领域正在不断扩展,AI技术在以上多领域的应用已经取得了显著成果。随着技术的不断进步,AI将在未来的医疗健康领域发挥更加重要的作用,为患者提供更精准、高效的医疗服务。

2.5 智能医学面临的挑战及发展前景

2.5.1 智能医学面临的挑战

智能医学已经迈入新一轮发展阶段,呈现出大数据、群体性、自主化、人机融合的发展新特征。它将给未来医学带来深刻的变化,是未来医学创新和改革的强大动力,将改变医疗手段甚至医疗模式,推动医学事业发展,重塑医疗产业结构,同时也必将对医师的未来规划产生重大影响。

随着科技前沿技术不断转化为医学应用,智能医学领域的应用场景会越来越丰富,智能医学将逐渐成为影响医疗行业发展、提升医疗服务水平的重要方向。不过智能医学领域的研究与发展也面临着严峻挑战,主要集中在两个层面:技术发展层面和管理规范层面。

(1) 技术发展层面

对于深度学习而言,模型越复杂,越具有强表达能力,就越需要海量数据积累和技术创新。人才专业水平是智能医学发展的关键因素之一,医学与理工科结合的高端复合人才队伍建设是重中之重,需要投入大量的人力物力进行学科建设。在智能医学与前沿科学技术同步发展之时,如何保护独立的人类思想可能会成为一个难题。

智能医学从学术牵引式发展迅速转变为需求牵引式发展,“人的智能”须面对“超级人工智能”的挑战,应提前考虑并采取预防措施以避免不可控问题的发生,积极寻求应对方法,确保医疗领域应用人工智能以“为人类利益服务,绝不伤害人类”为原则。

(2) 管理规范层面

主要是智能医学相关伦理与法律约束。智能医疗是通过打造健康档案区域医疗信息平台,利用最先进的物联网技术,实现患者与医务人员、医疗机构、医疗设备之间的互动,逐步达到信息化。智能医学应用于医疗领域的伦理挑战包括公平受益、失业、患者隐私、医疗安全、责任划分和监管等问题。其原因可能有未遵守基本伦理原则、技术缺陷、立法和监管缺失、隐含算法偏见、数据质量欠佳、公众素养不足等。

人工智能、3D 打印、大数据等技术应用于医疗领域的发展需要有效的监管机制,包括政府、组织以及来自公众的监管,不仅需要相关组织树立行业道德规范,也需要国家层面出台必要的法律法规,从法律层面处理科学技术带来的伦理难题。

总体而言,智能医学的发展对我们既是机遇,同时也是挑战。在智能医学时代,利用好前沿科技,提高自身的竞争优势,才能更好地迎接这个时代的来临。

2.5.2 智能医学的发展前景

(1) AI 技术的未来突破

未来 AI 在医疗领域有望实现更多技术突破。在诊断系统方面,AI 将具备更强大的图像识别和数据分析能力,能够对疾病进行更早期、更精准的诊断。例如,通过对人体生物标志物的分析,AI 可以在疾病尚未出现明显症状时就发现潜在的健康风险。有研究表明,基于 AI 的癌症早期筛查技术能够提前数月甚至数年发现癌症迹象,大大提高患者的治愈率。

手术机器人也将更加智能,具备更高的精度和灵活性。AI可以根据患者的个体解剖结构和手术需求,实时调整手术机器人的操作参数,提高手术的成功率和安全性。同时,远程手术技术也将得到进一步发展,让专家能够通过 AI 技术远程操控手术机器人,为偏远地区的患者进行手术。

(2) 医学 AI 的普及与可及性

AI 技术将有助于改善偏远或医疗资源匮乏地区的医疗状况,实现医疗公平。通过远程医疗平台,借助 AI 辅助诊断工具,偏远地区的患者可以将自己的医疗数据传输给专家,获得及时的诊断和治疗建议。例如,一些基于 AI 的移动医疗应用,可以让患者在家中通过手机采集生理数据,如血压、心率等,AI 系统对数据进行分析后,将异常情况及时反馈给医生,医生可以根据情况为患者提供进一步的指导 and 治疗。

此外,随着技术的发展和成本的降低,AI 医疗设备将更加普及,更多的医疗机构将能够使用更先进的 AI 技术,提高医疗服务的可及性。

(3) AI 与生物医学的融合

未来 AI 与生物技术、基因组学等前沿领域的结合,将会推动精准医疗的发展。AI 可以对大量的基因组数据进行分析,挖掘基因与疾病之间的关联,为个性化治疗提供依据。例如,通过分析患者的基因信息,AI 可以预测患者对不同药物的反应,帮助医生选择最有效的治疗药物和剂量。在生物制药领域,AI 与生物技术的融合将加速新药研发的进程。AI 可以根据生物分子的结构和功能,设计出更具针对性的药物分子,同时利用生物技术进行药物合成和验证,提高新药研发的效率和成功率,为攻克更多疑难病症带来希望。

2.6 思考题

1. 智能医学的定义强调“融合而非替代”,请以医学影像或远程医疗为例,分析其与传统医学的区别。
2. 比较医学影像智能方法与生物信息学方法的技术基础与应用异同。
3. 分析国内外智能医学发展差距的成因,并提出中国的赶超策略。
4. 以“天玑”骨科机器人为例,评价医用机器人在精准手术和资源均衡中的价值与局限。
5. 智能医学面临数据隐私、算法可解释性挑战,请从技术和管理层面提出解决方案。



【微信扫码】

本章相关资源

南京大学出版社
版权所有

第 3 章

信息素养基础

3.1 信息素养的相关概念

3.1.1 信息素养的概念

信息素养(information literacy)又称信息素质或信息能力,其本质是全球信息化需要人们具备的一种基本能力。信息素养最早是由美国信息产业协会主席保罗·泽考斯基(Paul Zurkowski)于 1974 年提出的,并概括为“利用大量的信息工具及主要信息源使问题得到解答的技术和技能”。简单定义来自 1989 年美国图书馆学会(American Library Association, ALA),该定义包括:能够判断什么时候需要信息,并且懂得如何去获取信息,如何去评价和有效利用所需的信息。现在一般认为信息素养是个体(人)对信息活动的态度以及对信息的获取、分析、加工、评价、创新、传播等方面的能力,它是一种对目前任务需要什么样的信息、在何处获取信息、如何获取信息、如何加工信息、如何传播信息的意识和能力。

随着人们对信息素养进一步的认识,现在认为信息素养是信息意识、信息能力和信息道德的总和,是人的整体素质的一部分。它是未来信息社会生活必备的基本能力,是创新型人才的基本素养。

信息素养已成为各行业必备基本素养之一。21 世纪以来,联合国教科文组织(United Nations Educational, Scientific and Cultural Organization, UNESCO)相继发布了《布拉格宣言》《亚历山大宣言》《菲斯宣言》《巴黎宣言》《里加倡议》等重要文件,并于 2011 年颁布文件将每年的 10 月 24—31 日设立为全球媒体与信息素养周。在 2021 年第 58 届联合国大会上,会议强调媒体与信息素养对可持续发展的重要性,并呼吁各国采取具体措施,以确保人人都能获取和利用信息,推动全民信息素养教育与实践。

3.1.2 信息素养的内容和内涵

(1) 信息素养的内容

信息素养包含了技术和人文两个层面的意义:从技术层面来讲,信息素养反映的是人们利用信息的意识和能力;从人文层面来讲,信息素养也反映了人们面对信息的心理状态。具

体而言,信息素养应包含以下五个方面的内容。

① 热爱生活,有获取新信息的意愿,能够主动地从生活实践中不断地查找、探究新信息。

② 具有基本的科学和文化常识,能够较为自如地对获得的信息进行辨别和分析,正确地加以评估。

③ 可灵活地支配信息,较好地掌握选择信息、拒绝信息的技能。

④ 能够有效地利用信息,表达个人的思想和观念,并乐意与他人分享不同的见解或资讯。

⑤ 无论面对何种情境,能够充满自信地运用各类信息解决问题,有较强的创新意识和进取精神。

美国提出的“信息素养”概念则包括三个层面:文化层面(知识方面)、信息意识(意识方面)、信息技能(技术方面)。经过一段时期之后,正式定义为:“要成为一个有信息素养的人,他必须能够确定何时需要信息,并已具有检索、评价和有效使用所需信息的能力。”

而在《信息素养全美论坛的终结报告》中,再次对信息素养的概念做了详尽表述:“一个有信息素养的人,他能够认识到精确和完整的信息是做出合理决策的基础;能够确定信息需求,形成基于信息需求的问题,确定潜在的信息源,制定成功的检索方案,以包括基于计算机的和其他的信息源获取信息、评价信息、组织信息用于实际的应用,将新信息与原有的知识体系进行融合以及在批判思考和问题解决的过程中使用信息。”

(2) 信息素养的内涵

信息素养的内涵包括四个方面:信息意识、信息能力、信息道德、终身学习的能力。

① 信息意识

信息意识指人们对情报现象的思想观点和人的情报嗅觉程度,是人们对社会产生的各种理论、观点、事物、现象从情报角度的理解、感受和评价能力。具体来说它包含了对于信息敏锐的感受力、持久的注意力和对信息价值的判断力、洞察力。

② 信息能力

信息能力也可以说是信息技能,包括确定信息需求的时机、选择信息源、高效获取信息、处理评估信息、有效利用信息的能力。

③ 信息道德

信息道德指人们在信息活动中应遵循的道德规范,如保护知识产权、尊重个人隐私、抵制不良信息等。

④ 终身学习的能力

获得终身学习的能力是信息素养教育的终极目标。信息素养教育应该把焦点放在学习者身上,即受教育者或者被培训者身上,而不是放在指导者或者教员身上,让学习者学会学习,获得终身学习的能力。

3.1.3 信息素养的特征

信息技术的发展已使经济非物质化,世界经济正加速向信息化迈进,人类已自然地进入信息时代。21 世纪是高科技社会,航天时代、基因生物工程时代、纳米时代、人工智能、经济全球化时代等,但不管怎么称呼,21 世纪的一切事业、工程都离不开信息,从这个意义来说,称 21 世纪是信息时代也是确切的。

在信息社会中,物质世界正在隐退到信息世界的背后,各类信息组成人类的基本生存环境,影响着芸芸众生的日常生活方式,因而构成了人们日常经验的重要组成部分。虽然信息素养在不同层次的人身上体现的侧重面不一样,但概括起来,它主要具有五大特征:① 捕捉信息的敏锐性;② 筛选信息的果断性;③ 评估信息的准确性;④ 交流信息的自如性;⑤ 应用信息的独创性。

3.1.4 信息素养的表现

信息素养主要表现为以下 8 个方面的能力:

(1) 运用信息工具

能熟练使用各种信息工具,特别是网络传播工具。

(2) 获取信息

能根据自己的学习目标有效地收集各种学习资料与信息,能熟练地运用阅读、访问、讨论、参观、实验、检索等获取信息的方法。

(3) 处理信息

能对收集的信息进行归纳、分类、存储记忆、鉴别、遴选、分析综合、抽象概括和表达等。

(4) 生成信息

在信息收集的基础上,能准确地概述、综合、履行和表达所需要的信息,使之简洁明了、通俗流畅并且富有个性特色。

(5) 创造信息

在多种收集信息的交互作用基础上,迸发创造思维的火花,产生新信息的生长点,从而创造新信息,达到收集信息的终极目的。

(6) 发挥信息的效益

善于运用接收的信息解决问题,让信息发挥最大的社会和经济效益。

(7) 信息协作

让信息和信息工具作为跨越时空的、“零距离”的交往和合作中介,使之成为延伸自己的高效手段,以此同外界建立多种和谐的合作关系。

(8) 信息免疫

浩瀚的信息资源往往良莠不齐,需要我们有正确的人生观、价值观、甄别能力,以及自控、自律和自我调节能力,能自觉抵御和消除垃圾信息与有害信息的干扰及侵蚀,并且完善合乎时代的信息伦理素养。

3.1.5 现有信息素养教育相关框架

随着信息技术的发展,信息素养已成为全球教育相关领域的重要议题。各国政府、国际组织、企业等纷纷制定框架,以规范信息素养教育和实践的目标、方法及评估标准,从而优化教育体系,促进终身学习,提高社会整体信息素养水平。以下为全球主要信息素养教育框架梳理及其核心理念与关键要素的分析。

(1) 国际组织层面

① 联合国信息素养系列框架

2011年,联合国举办了首届媒体与信息素养国际论坛,发出了每年庆祝全球媒体与信息素养周的国际倡议。自此开始,联合国相继发布了多个信息素养及其教育相关的重要框架,为推动信息素养教育奠定了坚实基础。2013年,UNESCO首先发布了《全球媒体与信息素养评估框架》,旨在为成员国提供评估指导,监测全民信息素养水平及相关计划执行效果。2018年,UNESCO发布了《数字素养技能全球参考框架》,提出了数字素养全球框架概念,明确了不同教育阶段的数字技能评估标准。2019年,联合国又发布了《全球媒体与信息素养伙伴关系联盟:框架与行动计划》,强调国际合作、政策制定和知识共享,支持各国在教育体系中整合媒体与信息素养,培养具备批判思维和信息技能的公民。

② 国际图联信息素养指南

2006年8月,国际图联(International Federation of Library Associations and Institutions, IFLA)发布了《面向终身学习的信息素养指南》,旨在帮助教育工作者和图书馆员实施信息素养教育,提高用户信息素养水平。指南定义了信息素养的核心技能及其对学术、职业、日常生活、社会参与等的重要性。指南认为,教育界和图书馆界在推动信息素养教育中发挥着关键作用,并提出了整合课程等策略以更好地支持和推广信息素养教育。

③ 国际教育成就评价协会信息素养评价框架

21世纪伊始,国际教育成就评价协会(International Association for the Evaluation of Educational Achievement, IEA)发起了国际计算机与信息素养研究(International Computer and Information Literacy Study, ICILS)项目,旨在借助信息素养评价框架评估学生在计算机与信息素养领域的能力水平。ICILS框架的首次实施在2013年,内容包括对计算机与信息素养的阐释,以及对相关背景因素的分析。随着信息技术的快速进步,ICILS框架在2018年进行了更新和扩展,增加了计算思维框架,旨在捕捉信息与通信技术(Information and Communications Technology, ICT)发展对教育的最新影响。ICILS框架的最近一次更新是在2023年,针对计算思维框架增加了新的模块,旨在更好地反映计算思维中的问题解决过程。

④ 欧盟数字素养框架

欧盟联合研究中心在2011年实施了“数字素养项目”,并于2013年发布了欧盟数字素养框架(European Digital Competence Framework, DigComp)的1.0版本。DigComp框架确定了数字素养的关键组成部分,可用于多种场景,尤其是教育和工作场景。该框架目前已更新至第四代,即2022年3月最新发布的DigComp2.2,在原有基础上提供了更多有关知识、技能和态度的示例,以期助力公民以批判性思维、安全地参与人工智能等数字新兴技术,现已成为欧盟乃至世界范围内数字素养领域的权威框架,具有广泛影响力。

(2) 相关国家层面

① 美国高等教育信息素养框架

2014年6月,美国大学与研究图书馆协会(Association of College and Research Libraries, ACRL)发布了《高等教育信息素养框架》的首个草案,在广泛征求意见并经过多轮反馈与修订后,于2015年1月形成了最终版本。该框架主要围绕着“权威的构建性与情

境性”“信息创建的过程性”等6个核心概念及其知识实践展开,引入了跨学科的信息素养教育新理念,强调了信息的社会背景以及批判性思维,成为许多国家高等教育信息素养相关标准和框架设计的重要参考。

② 英国高等教育信息素养“七大支柱”模型

英国国立与大学图书馆协会(Society of College, National and University Libraries, SCONUL)积极推动信息素养教育在高等教育中的整合,1999年首次发布了信息素养高等教育“七大支柱”模型,并于2011年进行了修订。该模型提供了一个全面的信息素养框架,将信息素养定义为个人在信息环境中有效工作的能力,而这一能力通过7个相互关联的“支柱”来体现,即识别、审视、规划、搜集、评估、管理和展示。该模型旨在帮助学生和研究人员在学术与专业环境中更加有效地使用信息,以在快速变化的信息环境中保持竞争力提供可行路径。

③ 澳大利亚与新西兰信息素养框架

2001年,澳大利亚和新西兰信息素养研究所(Australian and New Zealand Institute for Information Literacy, ANZIIL)制定并发布了《澳大利亚与新西兰信息素养框架》,旨在促进信息素养教育在澳大利亚和新西兰的高等教育系统中的应用和发展,并于2004年进行了修订。该框架制定了“明确信息需求并确定所需信息性质与范围”“高效检索所需信息”等6大核心标准,旨在帮助教育工作者制定课程目标,并为职业教育和高等教育部门提供了与其他教育部门的对齐标准的机会。

(3) 相关企业层面

① 数商全球标准框架

2019年,DQ研究所(DQ Institute)发布了《全球标准报告》,提出了数商(Digital Intelligence Quotient, DQ)框架,作为全球信息素养与数字技能的通用标准。该框架由“数字身份”“数字安全”等8大核心领域,“数字公民”“数字创造力”“数字竞争力”3种发展层级,以及24项具体能力组成,涵盖了个人信息管理、网络安全、数字沟通和创新能力等方面,旨在帮助个人和企业适应数字时代的挑战,提升数字适应能力。DQ研究所还联合电气电子工程师学会(Institute of Electrical and Electronics Engineers, IEEE)等国际组织推动该框架的标准化,使之成为企业、教育和政府部门广泛认可的数字素养发展框架。

② 微软数字技能发展框架

微软公司制定了一套“数字能力发展框架”,旨在帮助企业及其员工提升信息素养和数字技能。该框架涵盖生产力、创造力和创新力3个层次,旨在赋能企业适应数字化转型,提高自身竞争力。该框架采用角色驱动(Role-Based Learning)的学习方法,定义了“门店员工”等4种典型的学习角色,各角色对应不同的数字技能培养方案,并将数字技能分为“青铜”“白银”“黄金”3个等级,以适应不同层次的学习需求。该框架还提供了在线培训、学习资源和实践工具,帮助企业制定信息素养提升计划,确保员工具备应对数字时代挑战的能力。

(4) 不同框架对比分析

综合来看,上述框架有以下相同点:①均强调信息素养作为21世纪必备技能的重要性,以及在应对快速变化的信息环境中的重要作用;②虽结构体系与内容各异,但均强调信息素养不仅仅是对于有关技能的掌握,还包括批判性思维和终身学习等理念;③均强调与

教育系统、图书馆或其他信息机构开展多层次及多维度的广泛合作,帮助公民个人不断提升信息素养;④ 它们都强调在全球化和数字化的背景下,信息素养应在学习、工作和社会参与等方面都发挥积极作用。

上述框架的差异点主要体现在其聚焦点、关键要素和适用群体有所不同。就适用群体而言,部分框架如“全球媒介与信息素养伙伴关系联盟框架”“面向终身学习的信息素养指南”等面向全体公众,具备全民性质;而其他框架则更关注某些特定群体,如“全球数字素养框架”主要面向 15—24 岁的青年及成人,“英国信息素养七大支柱模型”和“美国高等教育信息素养框架”则聚焦高等教育层次的学生和教职工。

3.1.6 对大学生信息素养的要求

(1) 能够准确地确立信息问题

能够准确地确立信息问题是指将学习、生活当中的实际问题、某一项任务或科学研究课题等转变为能够被现有的信息资源系统或其他人所理解和应答的信息问题。这是一切信息活动的起点,又是制订信息获取计划的基本前提。

(2) 能够高效地获取所需要的信息

获取信息是确立信息问题和制订计划后的重要环节。获取信息的技能至少包括传统的图书馆技能、信息检索技能、计算机技能、社会调查能力以及各种科学探究方法等。

(3) 能批判性地评价信息及其来源

批判性思维和评价能力几乎在信息活动的各个环节发生作用。主要包括以下几点:

- ① 对信息问题的评价和调整。
- ② 对信息来源的评价和调整。
- ③ 对信息获取方式和策略的评价与调整。
- ④ 对信息的评价和筛选。

(4) 能够有效地分析与综合利用信息

主要是指对筛选的信息进行分析和综合,概括出中心思想,得出新的结论或新观点,与学生自身的知识体系整合,产生个体的新知识或人类的新知识,并灵活运用写作技能、多媒体信息技术等将之充分表达出来,有效地与他人交流信息成果。

(5) 懂得有关信息技术的使用所产生的经济、法律和社会问题

此要求主要是指在获取、使用和交流信息,使用信息技术时能够辩证地看待言论自由与审核制度。懂得尊重信息作者的知识产权,遵守基本的信息安全法规,理解和维护信息社会的各项道德规范,具有强烈的社会责任感,等等。

3.1.7 信息素养教育与文献检索课

检索技能是现代社会必备的基础技能,是信息素养教育的必备核心素质,充分利用信息检索课开展信息素质教育,丰富和完善信息素养教育的内容,通过对信息检索课的学习,完成对学生从简单的信息检索能力培养提升到全面的信息素养的培养。

无论是社会发展的需要,还是个人自身发展的需要,信息素质都是其重要的基本素

质之一。因此,任何一所高校都必须重视大学生的信息素质教育,重视文献检索课,加强对其重要性的认识,改进课程内容、方法和质量,以适应信息时代的教育对我们提出的要求。

3.1.8 医学信息素养教育

医学是一个对信息依赖度极高的学科,要想成为一名合格的医生,具备信息素养尤为重要。医学生除了须具备通用信息素养外,还应该具备适应医学专业需要的信息能力,而这个能力又与全球医学教育最低基本要求密切相关。1999年6月,美国中华医学基金会(CMB)理事会批准资助,成立了国际医学教育专门委员会(IIME)。该委员会的任务是为制定“全球医学教育最低基本要求”(以下简称“基本要求”)提供指导。“基本要求”是指世界各地医学院校培养的医生都必须具备的基本素质,包括医学知识、临床技能、职业态度、行为和职业道德等方面。将这些领域所需的能力与信息素养内涵相对应可以看到,信息素养几乎渗透进了医学教育的所有领域。

(1) 涉及信息意识和信息道德的条目

“自我调整的能力,认识到不断进行自我完善的重要性和个人的知识及能力的局限性,包括个人医学知识的不足等”“认识有关病人文件、知识产权的权益、保密和剽窃的伦理及医学问题”。

(2) 涉及信息能力的条目

“注意倾听、收集和综合与各种问题有关的信息,并能理解其实质内容”“建立和妥善保管医疗档案”“能综合并向听众介绍适合他们需要的信息,与他们讨论关于解决个人和社会重要问题的可达到的及可以接受的行动计划”。

(3) 涉及信息分析与利用的条目

“采集包括职业卫生等在内的相应病史资料”“运用基本的诊断和技术规程,对获得的观察结果进行分析和解释,确定问题的性质”“运用循证医学的原则,在挽救生命的过程中采用恰当的诊断和治疗手段”“进行临床思维,确立诊断和制定诊疗方案”“发展独立、自我引导学习的能力,以便在整个职业生涯中更好地获取新知识和技能”。

(4) 涉及信息管理全过程的条目

“从不同的数据库和数据源中检索、收集组织与分析有关卫生及生物医学信息”“从临床医学数据库中检索特定病人的信息”“运用信息和通信技术帮助诊断、治疗和预防,以及对健康状况的调查和监控”“懂得信息技术的应用及其局限性”“保存医疗工作的记录,以便进行分析和改进”“建立和妥善保管医疗档案”。

(5) 涉及信息评价和批判性思维的条目

“在职业活动中表现出有分析批判的精神、有根据的怀疑、创造精神和对事物进行研究的的态度”“懂得根据从不同信息源获得的信息在确定疾病的病因、治疗和预防中进行科学思维的重要性与局限性”“应用个人判断主动寻求信息而不是等待别人提供信息”“根据从不同来源获得的相关信息,运用科学思维去识别、阐明和解决病人的问题”。

从上述分析可以看出,信息素养在医学教育中具有相当重要的地位。

3.1.9 人工智能背景下的信息素养

(1) AI 素养教育的重要性已获得多个重要国际组织的广泛认同

例如,联合国教科文组织(UNESCO)于 2021 年发布的《人工智能与教育:政策制定者指南》,率先从全球战略层面明确了 AI 核心知识体系,为各国制定相关教育政策提供了方向指引。美国国家人工智能咨询委员会在 2023 年发布的《建议:提高全美人工智能素养》中,则将 AI 素养的提升正式纳入国家教育发展规划。作为全球教育技术领域的重要风向标,美国高等教育信息化协会(EDUCAUSE)在《2024 地平线报告:教与学版》中敏锐地捕捉到 AI 对教育的颠覆性影响,首次将其单独列为关键趋势维度,强调 AI 正以超出预期的速度重塑教学形态与学习体验,进一步凸显了 AI 素养培育在现代教育中的紧迫性。

(2) AI 素养水平评估框架的构建

为有效推进 AI 素养教育并衡量其成效,开发科学的 AI 素养评估指标、量表和工具已成为关键研究领域。学者 Lintner 系统梳理了现有 AI 素养量表,从信效度检验、适用群体特征到测量维度划分进行了多维剖析,为精准评估不同人群的 AI 素养水平提供了理论支撑。在教学体系构建方面,研究者积极开发创新框架与模型,试图打破教育阶段的藩篱,将 AI 素养培育有机融入基础教育、高等教育及职业培训的全链条中。例如,“脚手架式 AI 素养教学模型”针对学术图书馆员设计了分层递进的研讨体系,引导学生从基础概念认知逐步进阶到 AI 伦理思辨与批判性应用,促进能力的螺旋式上升。

(3) AI 技术在信息素养教育中的应用

Carroll 等学者考察了学生在研究相关任务中使用 AI 工具的熟练程度,以及信息素养教学是否会影响他们对这些工具的看法,认为图书馆员主导的信息素养培训可以有效促进学生生成式人工智能和大语言模型的更深入理解。Chatikobo 等利用人工智能就绪指数(Artificial Intelligence Readiness Index, AIRI)评估了将人工智能整合到宾杜拉科学大学教育图书馆信息素养教学中的潜力。其研究表明,虽然教师对 AI 概念有基本的了解,但在对 AI 工具的熟悉程度方面存在显著差异。

3.2 信息意识

3.2.1 信息意识基本概念

国外对信息意识(information consciousness)的研究,主要存在两个视角:一是在信息行为过程描述时所用的“信息意识”,它与信息意识的觉醒重叠,即“意识到自身存在信息需求”,因此应该理解为信息的觉知或觉醒;二是作为信息素养培养目标的信息意识,主要指对信息的主动性、敏感性和持久性等。当我们在讨论如何培养“信息意识”时,则应该理解为信息的意识指向、意识内容及意识情感。

中世纪的后期开始,被宗教理念长期禁锢着的人内心的自由理念和信息意识在文艺复兴的熏陶下开始明显地出现于文字当中。16 世纪,法国科学家笛卡尔第一次将内在灵魂和信息意识的同一性理性客观地用自己的语言表示出来,通俗地说就是信息意识的表现形式

有很多种,比如人的知觉、大脑的思维、内心的愿望等。到了17世纪,思想家洛克把人类所有的精神世界的表现归集起来统一称为信息意识,也就是说信息意识包含于人类的所有想象、创造、感情、信念和肢体动作当中。

国内对“信息意识”的研究存在多样性,由于意识本身的复杂性,在信息意识的发展过程中,对于信息意识的定义及构成要素的认识并不统一。随着信息社会的不断发展,对信息意识培养的研究也在逐步引起我们的注意。在学术刊物上也出现了越来越多研讨高校信息意识培养的文章。已有的研究主要从三种视角来界定信息意识的概念:

(1) 从意识的内涵角度,借鉴哲学中意识的概念

部分学者借鉴哲学意识形态的概念,认为信息意识是指信息主体对客观存在的信息现象的能动反应,即人作为信息主体对信息交流活动在社会中的地位、价值、功能和作用的认识。学者张斌在《影响信息意识形成的因素及信息意识的培育》一文中认为信息意识是对信息活动的认识过程,强调意识在认知过程中的动态特征。常正霞在《大学生信息伦理现状及对策研究》文中又以上述概念为基础增加了信息主体对信息需求的认识。王蕾在《提高信息意识是实施大学生信息素质教育的关键》一文中指出信息意识中还应包括对信息现象的感受、理解和评价。曾德良等在《大学生信息素质调查与分析》中强调对信息的控制、超越,和对自身行为的控制、约束。此外,李华奇在《论信息素质教育与创新人才培养》一文中认为,信息意识是指对知识信息的本质及功能的认识,是对知识信息重要性的认识和对知识信息的敏感程度,即人们从信息的角度对社会中各种现象和各种行为的感受、理解和评价这种自觉地对信息的心理反映。

(2) 从心理学意识的角度出发,借鉴心理学意识概念

从心理学意识的概念出发,部分学者认为信息意识是意识的延伸,或者是意识的子系统。解敏认为“信息意识是一种扩展的意识,是信息主体在与信息有关的认知活动中产生的感受,并在感受积累的基础上形成的对信息活动的一种觉知能力”。严丽在《信息素养析义》一文中认为“信息意识是指人脑对信息在社会发展中的性质、地位、价值和功能的认识 and 反应”。赵波和喻丽在《我国高校信息素养教育创新路径探讨》中指出,信息意识一方面体现在人们是否能主动寻求信息的获取,主要指人们在用信息的层面来对自然现象、社会上的各种现状、形式等去体会和理解,同时从中甄选和判断出对自身有利用价值的信息。王锦贵在《论信息素养》中指出,信息意识即人们对身边信息的灵敏程度,换句话说也表示为人们主动获取信息的一个内在过程,人们在用信息的层面来对自然现象、社会上的各种现状、形式等去体会和理解,主要有人们对信息的内在需求、自身分析、主动获取、筛选和吸收的心理活动。

(3) 通过信息意识的外部特征,界定信息意识概念

吴娱和李晓新在《中外信息素养教育比较研究》中指出,信息意识是指人们对实际生活中存在的信息主观能动需求和获取能力,也指对信息的内在灵敏程度和对未知信息的感应及预估能力,对已知信息的持续获取能力。

通过综合分析众多学者对信息意识定义的阐释,我们不难从中得出这样的结论:信息意识是一个具有多种含义的概念,不能只是简单地从某一方面来解释,而应从不同层次、不同角度来把握信息意识这个词汇。我们认为信息意识至少应具有以下几种含义:

第一,是信息在人脑中的集中反映。即人在信息活动中产生的认识、观点和理论的总和。这一点是就其作为社会意识形态的一种而言的。马克思主义哲学认为意识是物质世界长期发展的产物,是客观世界在人脑中的反映,包括认识的感性阶段和理性阶段。信息意识属于意识的一种,自然应具有意识的本质属性。也就是说,在对信息意识下定义时应牢牢把握它的哲学含义。在信息社会中信息充斥于社会生活的各个方面,人们的各项社会活动都离不开信息,信息活动融于各项社会活动之中。信息意识就是在这种情况下产生于人们的社会实践,是社会发展和信息产业兴起的必然产物。感性认识是对信息表象的认识,是理性认识的基础,而理性认识则是对信息的本质、内在联系的认识,是感性认识与理性认识的结合,是人们在信息活动中产生的认识、观点和理论的总和。

第二,是人们捕捉、判断、整理、利用信息的意识,即人脑作为高级神经系统在生理上对知识转化为信息具有兴奋性,同时人脑有特有的兴奋点去促进信息的转化及利用。从生理学角度来看,意识是人脑的机能,是人所特有的对客观现实的反映。那么,信息意识就是人所特有的通过脑生理机制对信息的反映,这与人们的观察力、注意力、联想能力、分析能力等有一定关系。通过观察捕捉到可能成为有效信息的“源信息”,再经过人脑的判断、分析、整理,从纷繁芜杂的“源信息”中提取出有效信息并加以利用。

第三,是对信息与信息价值所特有的敏锐的感知力、感悟力和较强的亲和力。换句话说就是对信息所特有的自觉反映。意识在心理学上一般指有意识的、能自觉意识到的心理活动。信息意识是一种心理上的潜意识。因为信息爆炸即信息过多会降低人们对信息的敏感度,这就要求个体有自身调节意识,强调有意识的、自觉的反映和思维。另外,人们对信息的心理需求越强,意识就越明确,自觉性、能动性就越大。实际上,信息意识的强弱经常表现为人们捕捉、判断、整理、利用信息的自觉程度的高低。

第四,是对现代技术的快速的认知力。也就是说,有意识地关注信息技术的最新发展并主动地加以利用。需要指出的是社会发展的各个时代都有反映该时代的基础设施。农业社会的基础设施是犁和拉犁的牲畜,工业社会的基础设施是引擎和供给引擎的燃料。在信息社会,信息技术的迅速发展尤其是计算机和通信技术的发展及其结合形成了信息基础设施。目前,信息技术的广泛应用正在改变人们的生活、学习、工作方式。今后人们获取信息将在很大程度上依赖信息技术。因此,在对信息意识定义时,应对信息技术加以强调。

基于上述四个方面的考虑,我们对信息意识的定义为:信息意识是信息在人脑中的集中反映,即社会成员在信息活动中产生的认识、观点和理论的总和,是人们凭借对信息与信息价值所特有的敏感性和亲和力,主动利用现代信息技术捕捉、判断、整理、利用信息的意识。

3.2.2 信息意识基本组成

信息意识主要包括:

(1) 能认识到信息在信息时代的重要作用,确立在信息时代尊重知识、终身学习、勇于创新的一些新的观念。

(2) 对信息有积极的内在需求。每个人除了自身有对信息的需求外,还应善于将社会对个人的要求自觉地转化为个人内在的信息需求,这样才能适应社会发展的需要。

(3) 对信息的敏感性和洞察力。能迅速有效地发现并掌握有价值的信息,并善于从他人看来是微不足道、毫无价值的信息中发现信息的隐含意义和价值,善于识别信息的真伪,

善于将信息现象与实际工作、生活、学习迅速联系起来,善于从信息中找出解决问题的关键。

3.2.3 信息意识在高校信息素养中的重要作用

在日本,信息水平测试每年举行一次,已成为仅次于高考的全国第二大的大型考试。日本本土公司也常年开办培训班进行信息意识培养,提高员工的信息意识水平和信息能力。在法国,政府提出要把对信息意识培养普及到高中。

可以看出,在很多发达国家,在各个方面信息意识都有明显的体现,也在国民经济、体育文化建设、教育科学领域之间搭建了沟通平台。信息意识的培养在其信息素养教育中占有重要的位置。

在信息技术为主导的社会网络环境中,大学生与网络信息接触频繁。信息意识是有效获取网络信息和利用网络信息的基础。信息意识能够帮助大学生对信息的重要性产生足够的认识,对信息自觉形成敏锐的感知能力以及正确的洞察力和判断力。引导大学生充分利用网络上的信息资源,把网络上有价值的信息吸收、转化为自身的知识体系,以达到信息素养的各方面的共同提升。因而,信息意识是高校大学生信息素养教育的基础环节,在大学生的信息素养培养中具有重要的作用,主要表现为以下几点:

(1) 信息意识能使大学生充分认识到信息的重要性

科学技术的发展,特别是计算机和互联网的发展与运用,使社会进入到信息时代。信息以前所未有的迅猛态势渗透于社会生活的方方面面,越来越成为一种不可轻视的重要社会资源,对信息的重要性有充分的认识是信息意识的首要表现。

信息全球化的社会环境中,人们获得信息的方法在不断增加。互联网被广泛使用,更是打破了信息的垄断。由于网络信息在发布上自由和方便性等,网络上的信息资源相当丰富,共享程度极高。各类信息已经广泛渗透到人类生活、工作、学习的各个方面,并给人们带来深刻的影响,人们与信息的关系已经密不可分。当信息发展到生产经营活动中时,信息同能源和材料等成为三大资源之一,信息的开发、利用则随之成为信息化体系建设中的核心内容,其目的则是为了更好地实现信息价值,发挥其效用。

网络信息非常丰富,来源极广,有着更新及时、获取方便、利用成本低等优势,这些网络信息和网上知识对高校中的大学生来说,成为课堂授课外重要的信息和知识来源。给大学生的学习、日常生活等诸方面都带来了很大的帮助,成为高校学生创造性地完成学习目标,进行创新思维,实现终身学习的必要因素,也为大学生在已经拥有的知识基础上了解前沿知识、创造新知识提供条件。在这个信息成为重要资源的时代,拥有信息量的多少,对信息的准确程度地把握,处理和利用信息的能力大小对当代大学生的社会适应能力有着决定性的影响。

(2) 信息意识能促使大学生养成敏锐的信息感知能力

信息意识本质上是对信息价值自觉认知的意识行为,能够使人在信息搜索过程中产生信息敏感。信息敏感是对目标信息进行定位的前提,体现为一种自觉的信息感知意识。

信息感知能力是指通过信息感知的渠道对信息使用者产生刺激,从而刺激信息使用者对信息产生的主观认知能力。敏锐的信息感知能力是信息意识的基本能力,是高校学生利用好网络信息和掌握知识的基本条件。网络信息层出不穷,鱼龙混杂。大学生在利用校园

网和其他的网站,获取各种学习资料以及知识信息的过程中,敏锐的信息感知能力能够起正确的引导作用,帮助大学生迅速地从纷繁复杂的网络信息中,找到与自己的需求目标相符合的信息或是具有重要价值的相关信息资料,摒弃过时的和错误的信息资料。

信息意识具有能动性,良好的信息意识对大学生养成敏锐的信息感知能力有促进作用,会时刻提醒大学生认识到自身的信息需求,密切关注网络上的信息资源,善于发现网络中的各种有价值的信息资料,根据信息的有用性和可用性迅速有效地从获取的网络信息中汲取自己需要的知识,把这些有价值的信息资源整合到自己的知识体系中,以丰富自己的信息储备量和知识积累,从而在实现全面发展上取得卓越的成果。

(3) 信息意识有助于大学生形成有效的信息判断能力

网络信息的价值是通过对信息使用者的有用性大小来判断的,信息判断能力的形成需要有严谨的态度、缜密的思维和必要的信息常识,信息意识是形成信息判断力的基础。

对信息价值的判断必须在一定的思想意识指导下进行,这个思想意识就是信息意识。网络上信息环境复杂,网络信息良莠并存,有时甚至杂乱无章。对任何信息都不能盲目地、无根据地妄下定论,而应进行理性的评价和判断。鉴别出信息的真假,掌握其实质,判断信息真假是对目标信息进行有效确认的首要步骤。在对网络信息进行评价和判断中,需要抛开信息的表象,找出信息的本质所在,对信息的实用价值和适用性有所把握。信息要真实可靠,同时还必须与实际需要相符合,从而确定出哪些信息更具有针对性、更合乎目的性。这往往要经过对信息的概括、归纳、比较才能实现,其过程也同时体现作为信息使用主体的人对信息认识的过程,必然离不开信息意识。在信息意识基础上的信息评价和信息价值判断,是高校学生能够有效地、高效地从网络上获取信息并加以利用的前提。

从信息意识的本质来看,它是人脑的基本机能,是人们对信息所产生的认识的发展过程,因而具有发展性和过程性。当对某些信息一时不能做出有效判断的情况时,则需要通过对有效信息进一步积累的方法来提高自己的判断能力。有效的信息判断能力是做好信息处理的基础,也是学生的信息意识养成的重要体现。而且,良好的信息判断能力和信息分析能力能增强大学生对网络信息真假的判断力,有利于提高高校学生的认知水平,有利于加强自身的道德修养和诚信意识。

(4) 信息意识能指导和控制大学生的信息行为

网络信息使用者对信息的使用过程,简单说来就是一个不断从网上获取信息,不断对信息进行加工处理,并与自身的知识相融合、相转化的过程。在这个过程中,信息意识指导并控制着大学生信息活动的整个过程。

信息需求是指人们在实践活动中为解决各种实际问题而对信息的不满足和必要感。信息需求是引发信息行为的原动力。信息使用者的信息搜索和获取行为是受信息需求支配的。在当前的信息环境下,高校学生从网络获取信息的目的主要是为了增加知识储备和全面提升自己以更好地适应社会,其信息需求也明显地呈现广泛性、知识性、专业性、现实性、前沿性等特点,信息需求成为高校学生信息利用过程中必备要素和基本需求。处在形形色色的网络信息中,清楚地知道自己需要怎样的信息资料,显然是受信息意识的清晰程度控制的。信息意识明确的大学生,其信息需求越强烈,其信息行为的目标性和动力就越大,就会越积极主动;信息意识弱则信息需求就不会很强,所进行的信息搜索和信息获取的动力也就

越少,其信息行为也会缺少自觉性和主动性。

信息行为是指人们在信息需求驱使下,在适当的信息环境中从事信息活动的行为过程,合理地利用信息技术从网络中获取有用信息,是有效使用网络信息的必要步骤。在信息意识引导下,大学生能形成对网络信息的特有感知力,引导大学生从信息海洋中及时地发现有价值的信息资源,做出对目标信息的有效定位。信息意识在大学生根据自己的需求目标准确地获取学习资料和信息的过程中发挥着指导作用。信息意识指导大学生对收集到的信息资料进行分类整理、分析评价,在现有信息的基础上转换创造出新的信息等一系列的信息行为。并且,高校学生在信息利用的过程中,信息意识始终控制和调动着其他各项能力的协调运用和发挥。比如,大学生在运用各类信息工具对网络信息进行访问、阅读、检索、收集、评价、利用、创新的一系列过程中,信息意识都是一直参与其中并发挥着积极的能动作用。

信息意识作用下的大学生信息行为更具有科学性和高程度的主动性,其信息支配能力也越强。因此,信息意识在影响高校学生的信息需求和信息行为的同时,信息意识还会直接影响高校学生对信息利用情况,并对大学生的信息利用效果产生决定作用。对网络信息的利用是高校学生信息行为的最终目标。信息利用首先受信息支配能力的影响,信息的支配能力大小决定了信息行为的结果。明确的信息意识会使大学生的信息活动的目标更加清楚,针对这一目标所进行的信息搜索和信息获取行为就越稳定越高效,对信息的支配能力也就越灵活,就能够做到更好地掌握选择使用信息,拒绝负面信息,充分发挥网络信息的价值。

信息技术犹如一把双刃剑,它在为人们提供极大便利的同时,也产生了一些与传统伦理道德相悖的现象,如信息的滥用和泛滥、计算机病毒肆虐、电脑黑客、网络安全、网络信息共享与版权等问题,这就对人的道德水平、文明程度提出了新的要求。作为信息社会中的现代人,应认识到信息和信息技术的意义及其在社会生活中所起的作用与影响,要有信息责任感,能抵制信息污染,遵循一定的信息伦理与道德,规范自身的信息行为活动,主动参与理想的信息社会的创建。

3.3 信息伦理

3.3.1 信息伦理基本概念

信息伦理(Information Ethics)是由计算机伦理和网络伦理发展演化而来的。简单来说,计算机伦理是用来调整与计算机活动相关的人与人之间关系的规范和准则,网络伦理是用来调整与计算机网络有关的人与人之间关系的规范和准则,信息伦理则是调整信息活动中人与人之间关系的规范和准则。从三者的概念进行区分,可以发现信息伦理是立足于一般的信息或信息技术层面,而网络伦理中的互联网络是基于计算机的,且网络伦理一般不涉及人工智能、专家系统等方面的伦理问题,但计算机伦理却不能回避这些问题。所以,网络伦理和计算机伦理都包含在信息伦理之内,信息伦理涵盖计算机伦理和网络伦理。

目前学术界对信息伦理的概念尚未形成一致的看法,这是由于学者们站在不同的角度研究信息伦理的结果。本书以沙勇忠在《信息伦理论纲》中对信息伦理的定义为准,即信息活动中以善恶为标准,依靠人们的内心信念和特殊社会手段维系的,调整人与人之间以及个人与社会之间信息关系的原则规范、心理意识和行为活动的总和。

3.3.2 伦理与道德的关系

在“伦理”一词中,“伦”指的是人与人之间的关系,还指条理、顺序;“理”指物质本身的纹路、层次,客观事物本身的次序。在《辞海》中,“伦理”具有两个含义,一是事物的条理、顺序;二是指人们在处理彼此间的关系时应该遵循的道理和准则。

“道德”一词原本是分开使用的,“道德”二字连用始于荀子《劝学》篇:“故学至乎礼而止矣,夫是之谓道德之极”。由此可见,道德是一种社会意识形态,是人们共同生活及其行为的准则与规范。在西方,“morality”(道德)源自拉丁文的“moralis”一词,其复数形式“mores”有“风俗”或“习惯”的意思。之后,“morality”是指美德、道德、道德原则、道德规范等。

若要对“伦理”和“道德”两个词加以区分,可以从它们产生根源的角度加以区别。伦理产生的根源是社会利益冲突,是从社会的角度提出的处理人与人之间关系的规范;道德的根源却在于人心,是从个体的角度身体力行伦理规范而后有所得。所以,我们往往将伦理的内涵与社会层面的内容关联起来,伦理一词更具客观、外在、社会性意味,而将道德的内涵与个体层面的内容关联起来,道德一词更含主观、内在、个体性意味。而一般情况下,伦理与道德两个词语是可以互换使用的。本书对信息伦理、信息道德等术语作为同一概念使用。

3.3.3 信息伦理基本组成

信息伦理不是由国家强行制定和强行执行的,而是在信息活动中以善恶为标准,依靠人们的内心信念和特殊社会手段维系的。信息伦理构成可概括为两个方面和三个层次。

所谓两个方面,即主观方面和客观方面。前者指人类个体在信息活动中以心理活动形式表现出来的道德观念、情感、行为和品质,如对信息劳动的价值认同,对非法窃取他人信息成果的鄙视等,即个人信息道德;后者指社会信息活动中人与人之间的关系以及反映这种关系的行为准则与规范,如扬善抑恶、权利义务、契约精神等,即社会信息道德。

所谓三个层次,即信息道德意识、信息道德关系、信息道德活动。信息道德意识是信息伦理的第一个层次,包括与信息相关的道德观念、道德情感、道德意志、道德信念、道德理想等,它是信息道德行为的深层心理动因。信息道德意识集中地体现在信息道德原则、规范和范畴之中。信息道德关系是信息伦理的第二个层次,包括个人与个人的关系、个人与组织的关系、组织与组织的关系,这种关系是建立在一定的权利和义务的基础上,并以一定信息道德规范形式表现出来的。如联机网络条件下的资源共享,网络成员既有共享网上资源的权利(尽管有级次之分),也要承担相应的义务,遵循网络的管理规则。成员之间的关系是通过大家共同认同的信息道德规范和准则维系的。信息道德关系是一种特殊的社会关系,是被经济关系和其他社会关系所决定、所派生出的人与人之间的信息关系。信息道德活动是信息伦理的第三层次,包括信息道德行为、信息道德评价、信息道德教育和信息道德修养等,这是信息道德的一个十分活跃的层次。信息道德行为即人们在信息交流中所采取的有意识的、经过选择的行动。根据一定的信息道德规范对人们的信息行为进行善恶判断即为信息道德评价。按一定的信息道德理想对人的品质和性格进行陶冶就是信息道德教育。信息道德修养则是人们对自己的信息意识和信息行为的自我解剖、自我改造。信息道德活动主要体现在信息道德实践中。

总的来说,作为意识现象的信息伦理,它是主观的;作为关系现象的信息伦理,它是客观

的;作为活动现象的信息伦理,则是主观见之于客观的。换言之,信息伦理是主观方面即个人信息伦理与客观方面即社会信息伦理的有机统一。

3.3.4 信息伦理主要特征

科技伦理的观点:“技术不是价值中性的,会产生伦理效应。然而,它究竟产生善还是恶,却取决于人对于这种伦理后果的自觉,取决于人的意志。”大数据时代哪种信息是属于善的信息,哪种信息属于恶的信息,信息产生、储存、传播等过程都涉及伦理问题,信息伦理较其他伦理有着它独有的特征。孙伟平和贾旭东在《关于“网络社会”的道德思考》中提出了信息道德的三大特征:自主性、开放性、多元性。

第一,自主性。大数据环境下,信息生产速度快、变化快,信息的存储载体多、方式多,信息的传播速度快、模式多,获取信息的途径也多。每个信息的生产者,也可以是信息存储者、传播者。信息存储者和传播者也可以是信息的生产者,每个信息主体都可以在信息角色中变化和转化。他们不但是组织者,也是参与者。在立法相对滞后的信息市场化中更多时候是靠信息参与者的自我约束。他们的道德是自发自觉形成的,自主性更多,依赖性更少。

第二,开放性。大数据信息量大、内容纷繁复杂,是由很多的信息主体参与共同完成的。不同的信息主体有着不同的思维模式、价值观念、道德水平。信息主体在大数据平台,可以随意生产信息,可以自主地表达自己的意愿,也能够包容和接纳他人信息,达到共享和开放。

第三,多元性。大数据时代,可以说大数据信息犹如汪洋大海。信息来源于不同的信息主体,也来源于世界的各行各业、来源于不同角落,有着不一样的内容、不一样的特点、不一样的价值、不一样的评价标准。生产的信息可以是“善”的,也可以是“恶”的,可以是自己的,也可以是别人的,信息主体存在多元化,要保证大数据正常发展,让所有公民都能享受大数据信息带来的便利,必然要接收信息的多元化。

3.3.5 信息伦理发展史

(1) 兴起与发展

信息伦理学的形成是从对信息技术的社会影响研究开始的。信息伦理的兴起与发展植根于信息技术的广泛应用所引起的利益冲突和道德困境,以及建立信息社会新的道德秩序的需要。第二次世界大战后,电子计算机、通信技术、网络技术的应用发展促使西方发达国家率先进入信息社会。在对信息化及信息社会理论的研究进程中,西方学术界逐渐发现了在一系列新的信息技术条件下引发的伦理问题,并为此开辟了一门新的应用伦理学——信息伦理学。它最早源于计算机伦理研究,20世纪70年代,美国教授W·曼纳首先发明并使用了“计算机伦理学”这个术语。1971年G·M·温伯格在《计算机程序编写心理学》一书中,首先对信息技术对社会伦理问题产生的影响进行了研究。从20世纪80年代中期开始,大量信息伦理论文和专著的涌现,使信息伦理学的研究取得了突破性的发展。1985年,J·H·穆尔在《元哲学》杂志上发表的论文中提出了“计算机伦理”概念。同年,德国的信息科学家拉斐尔·卡普罗教授发表题为《信息科学的道德问题》的论文,研究了电子形式下信息的生产、存储、传播和使用问题,他提出“信息科学伦理学”“交流伦理学”等概念,并从宏观和微观两个角度探讨了信息伦理学的问题,包括信息研究、信息科学教育、信息工作领域中的

伦理问题,他将信息伦理学的研究放在科学、技术、经济和社会知识等背景下进行,他认为任何伦理理论都是对人的自由的反映,通信与信息领域的伦理理论也是如此。这是最早以信息科学作为伦理学研究对象的论文。1986年,美国管理信息科学专家 R·O·梅森提出信息时代有 4 个主要的伦理议题:信息隐私权、信息准确性、信息产权及信息资源存取权。

(2) 深刻的变化

20 世纪 90 年代,信息伦理学的研究发生了深刻的变化,它冲破了计算机伦理学的束缚,将研究的对象更加明确地确定为信息领域的伦理问题,在概念和名称的使用上也更为直白,直接使用了“信息伦理”这个术语。1996 年,英国学者 R·西蒙和美国学者 W·B·特立尔共同发表题为《信息伦理学:第二代》的文章,他们认为,计算机伦理学是第一代信息伦理学,其研究范围有限,研究深度不够,只是对计算机现象的解释,缺乏全面的伦理学理论。1999 年,拉斐尔·卡普罗教授发表论文《数字图书馆的伦理学方面》,该文对信息时代发生巨大变化的图书馆方面产生的伦理问题加以分析和论述。2000 年,拉斐尔·卡普罗教授又发表论文《数字时代的伦理与信息》,这篇论文的主题还是论述数字时代图书馆的伦理问题,他指出:“作为一种描述性的理论,信息伦理学揭示了一种权利结构,这种权利结构对不同文化和不同时代的信息观念和传统观念的态度产生影响,作为一种不受约束的理论,信息伦理开创了对道德态度和道德传统的批判。”随后拉斐尔·卡普罗教授又发表题为《21 世纪信息社会的伦理挑战》的论文,专门论述了信息社会的伦理问题,特别讨论网络环境下提出的信息伦理问题,他将信息伦理学从计算机伦理学中区分出来,强调的是信息伦理学。他认为,新的信息技术提出了对伦理学的挑战,在虚拟现实巾存在着对传统的伦理关系的威胁。拉斐尔·卡普罗教授的信息伦理学观点的变化,反映出信息伦理学理论的发展和变化。

(3) 工业化信息化结合

在全球化的信息浪潮中,我国必须把工业化和信息化结合起来,充分吸取西方发达国家信息化的成功经验,力争跳跃式地实现向信息社会的转型。而要顺利地完我国信息化的任务,构建一个有序的信息社会,除了要加快信息技术的发展,信息资源的开发之外,构建适合我国国情的信息伦理体系也势必成为当务之急。中国是一个有着悠久历史的文明古国,本土文化资源极为丰富且影响深远,在这样的背景下,更需要正确地把握和处理文化传统与新型的信息伦理之间的关系。

3.3.6 信息伦理主要问题

当今社会是大数据快速发展的时代,由于大数据技术本身还需完善和使用大数据信息的主体道德素质高低不同,出现了部分破坏大数据环境的现象,使得大数据信息在使用过程中陷入伦理困境。结合目前社会现实状况,以伦理学视角分析大数据时代信息伦理问题产生的原因。

(1) 信息造假和恶意传播消减社会信任

大数据的发展,保证数据信息真实有效是首要条件,数据信息真实性直接影响信息采集的真实性,对后续大数据信息使用和传播有着极大的影响。在实际生活中,确实存在信息造假问题。有部分信息生产者为保护自身隐私或达到某种目的,有可能生产的就是虚假信息,甚至盗用他人信息进行虚假加工。信息采集员可能受到经济利益诱导和他人的强加干扰,

更改信息源,混淆信息源真伪。大数据的发展,信息传播起到极大作用,信息传播速度越快,信息更新和使用越快,对各行各业发展乃至社会进步都有益处。正是由于信息传播有着速度快的特性,一旦出现恶意传播,后果不堪设想。

信息造假不仅会对信息资源造成浪费,还会造成人与人之间的不信任、破坏社会秩序,给社会诚信造成严重冲击。社会诚信是一种社会风气,是人与人之间,人与社会之间相互关系中逐渐形成被社会和个人都广泛认可的诚实守信的规则和道德。特别是对于违背社会公德和影响社会稳定的虚假信息恶意传播,信息传播者无视社会公德,为了自己一己私欲,或者为了蹭热点,对社会关注的、影响大的数据进行捏造篡改后,甚至加上自己不当言论,再进行传播,形成网络暴力,给社会信任和传统的道德以极大冲击,并且大数据信息传播速度特别快,公民辨别能力参差不齐,也有可能跟着继续传播或者篡改信息后再度恶意传播,引起社会民众的恐慌。比如信息主体为了达到某种目的,将大数据信息修改为自己和公众期望的大数据信息,给公众错误的引导。大数据信息主体是如何运用大数据收集数据、分析数据、修改数据是公众无法了解的,公众看到的就是数据的本身。舍恩伯格,牛津大学的著名教授,业界公认的大数据时代预言家,代表作《大数据时代》,他的观点是没有信任就没有大数据。从他的观点中,我们能明白,信任是社会关系的基础,若是行为主体之间缺乏信任,会对大数据信息产生恐慌,个人的隐私被外界无休止地使用,会担心自身信息被他人非法使用,加剧人与人之间的矛盾,从而不敢生产数据信息,最后导致大数据时代的结束。所以根据舍恩伯格的告诫,只有改善和加强信息行为主体之间的信任才能进一步发展大数据。

(2) 隐私窃取和泄露对人格的损害

大数据时代,不法分子会利用大数据技术漏洞,运用非法手段,窃取他人隐私,为自身谋取利益和财富,损害他人隐私及经济利益。所以个人信息安全在大数据时代显得更为重要。不法分子会非法入侵大数据信息储存载体,手机、电脑、网站还有各种注册个人信息的软件等,获取数据库信息进行黑市交易或者造假再生产。个人隐私信息包括电话、姓名、职业、住址、喜好、购物、信贷等多种信息。隐私泄露是指个人不愿意公开的、敏感的、重要的、涉及自身利益的、机密的信息等,没有经过信息主体授权或非信息主体授权,以及运用不正当手段让信息主体无法掌控个人信息等情况下让他人知晓。隐私泄露表现在两个方面:一方面是泄露自己的隐私。有因缺乏自我保护意识,把个人信息随意告诉他人的,也有被迫泄露自己隐私的。例如有些应用软件设置了读取使用人个人信息功能,若用户不同意读取个人信息,就不能使用该软件,用户为享受大数据带来的诸多便利,被迫泄露个人隐私。另一方面是泄露他人隐私。大数据时代是个信息共享的时代,每个公民除了生产和推送个人信息给他人共享外,也能共享他人信息。大数据时代获取他人信息比通过报纸、广播电视等传统媒介速度更快、更多、渠道更广。

隐私的窃取和泄露不光损害了他人利益,还可能会对他人造成人格损害,影响极其恶劣。美国的社交网络服务网站脸书(Facebook),为了测试用户在面对各类复杂信息时所表现出的情绪和反应是否能够被 Facebook 所监控,于 2012 年对 Facebook 上的 70 万名用户进行了隐秘测试。该次隐秘测试过程中,Facebook 故意在网络信息的页面上加设了部分关于情绪反馈的关键性词语,通过 Facebook 用户在网络信息阅读过程中涉及的情感反馈内容和关键词,判断 Facebook 用户在阅读时所流露的情感反应和变化。对 70 万 Facebook 用户的情绪测试整整持续了一周,通过该隐秘测试,Facebook 不但成功地收集了 Facebook 用户

的情感反馈数据,且能够通过该数据较为精准地预测和判断用户之后对于各种信息的情绪和态度。该测试被 Facebook 用户知晓后,对 Facebook 的行为表示了强烈的反对和谴责,认为其侵犯了个人隐私。Sheryl Sandberg(Facebook 副总裁)在第一时间对 Facebook 未经同意的实验表示了歉意,但这种私自使用用户信息进行测试的做法还是受到了整个行业的谴责和批评。此次测试事件也让社会各界感到焦虑,并对拥有“大数据”资源绝对控制权的企业和社会组织,在对用户数据的使用是否需要道德和法律约束及规范提出了新的思考。

从信息伦理的角度对此案例进行分析,Facebook 未经过用户同意,私自利用“大数据”绝对优势对用户信息使用以及情感流露习惯进行实验不符基本的道德准则。这种未经允许,对用户个人信息进行分析甚至对个人情绪倾向进行预测,不仅侵害了用户的个人隐私权利,也伤害了用户的个人尊严;而且作为美国的社交网络服务的主流网站,同时也是一家上市企业,其行为不但没有履行社会责任,还危害了社会的稳定。更令人担忧的是,Facebook 作为美国主流的社交网络服务网站,其行为受到了社会各界以及企业内部的监督和控制,从一定程度上而言 Facebook 的行为还会有所限制,但其他对大数据信息资源拥有控制力的中小型企业 and 组织能否自觉履行社会责任,是无法保证的。

(3) 信息分配不公平导致数据权利不平等

所谓信息不公平,蒋永福教授的观点是“信息不公平指不同信息主体之间配置和占有信息资源的不对等、信息使用的不均衡导致的信息主体之间经济和社会地位差距大,因为差距大又引起信息垄断、信息霸权等更多的不平等”。信息分配不公平和信息主体差异化都会导致信息不公平。

区域发展不一致,各地各行业需求信息不同,从宏观角度来看,参与信息分配者会存在分配信息不公平。因为发达地区获取信息更快更多,也会吸引更多的人口和产业参与,信息分配者在分配信息时也会有倾向性。信息分配对地域是这样,对个人或群体也是如此。这样导致有获取信息能力的能分配到更多的信息,没有获取信息能力的分配到更少的信息,两极主体差距越来越大。

在大数据发展过程中,信息主体的文化涵养、接受教育程度、个体职业等都影响着信息主体生产信息、传播信息、使用信息等。首先信息主体的文化涵养和接受教育程度的差异,直接影响信息主体对信息源的真实生产、对信息真伪的分辨、获取信息的途径、传播信息的价值观。在大数据信息瞬息万变的情况下,不同的人会对大数据信息形成不同的看法、不同的理解,对信息利用形成不同的认知、不同的态度。其次是个体职业拉大了信息主体的差异。对信息资源的利用一般情况下取决于信息主体所处的信息环境,所以从事信息化职业或者与信息化产业联系密切的,接受信息化教育和知识越多,获取的信息量越快,信息捕获能力越强,如计算机工作人员、大数据技术操作员等。

在大数据时代,信息分配不公平,主体差异会导致部分大数据信息只属于某个人、某团体、某地区等,其他人员享有机会较少。大数据信息分配过程中信息主体的职业、经济、性别等作为决定要素,而这些要素又不是个人能够决定和控制的,这样直接导致了公民获取信息不平等,也就是权利的不平等。地区差异,带来的信息分配不公平,让地区差异越来越大,地区文化和经济会形成自我保护,导致地区和地区之间经济、文化差别更大。这种不公平的信息分配还会造成信息垄断,会让部分信息主体依法并合规地拥有信息专属权。信息垄断可以分为市场竞争中的信息垄断和非市场竞争中的信息垄断,信息分配不公平形成的是非市

场竞争的信息垄断,会刺激甚至扰乱市场竞争中信息垄断行为。

(4) 信息分化影响社会公平

信息不公平,导致贫富差距拉大,随之又进一步加大信息不公,这是个恶性循环,更是个动态循环。大数据越发展,信息不公平现象可能越明显,这个现象最明显的体现就是信息分化。谢俊贵教授在当代信息分化问题调查研究中指出“人们之间经济地位和经济能力的不同,必然会造成人们之间一般生活资料消费水平、一般生产活动投资能力以及一般生存方式的巨大差别,而且会造成人们信息消费水平、信息投资能力和信息化生产方式的巨大差别。”在大数据主体主动获取信息上也是有差距的,要获取大数据信息,前提是掌握使用互联网。谢俊贵教授在《城乡信息分化的新态势及其因应策略——基于 CNNIC 互联网普及率统计数据》一文中提到我国经济越发达、城市化建设和发展越高的地区,大数据信息越富有;经济越落后、城市建设和发展越落后的地区,大数据信息越贫瘠。城市居民占有大数据信息高于农村居民占有大数据信息。同时,信息作为一种商品在市场上流通售卖,价格高低代表了信息价值大小,拥有大量信息的或者处于信息核心区域的便能用信息取得利润。社会经济越发展,大数据信息市场越成熟,社会和相关政策越会倾向于提高信息经济;但是信息经济越发达,信息分化也会随之越严重。

在大数据时代,信息公平是社会公平的重要表现。而信息分化是信息不平等的明显体现,也是影响社会公平的因素。信息分化反映了主体对信息资源获取机会、占有信息资源和信息分配过程中的不平等,对于信息资源的占有和利用,会影响和决定信息富有者更加富有、信息贫瘠者更加贫瘠。信息主体的职业、教育,以及其他机会都会随着信息分化而加剧。例如,就业机会,部分信息主体对信息资源嗅觉敏锐、信息处理能力强能够获得更多的就业机会,能选择高收入职业;而部分信息主体自身条件受限,文化层次不高,获取信息少,获得就业机会也少。信息分化导致的不公平还体现在城市农村的信息差距、贫富差距、收入和消费的差距等。

3.3.7 信息伦理主要成因

(1) 伦理道德滞后于大数据发展

大数据依托网络为载体,大数据行为主体之间的交流不是面对面的交流,而是通过网络交流,面对的都是没有感情的机器。面对机器,行为主体之间道德关系很冷漠,相互不信任。大数据发展越快,社会会发生越多的相关规范偏离现象,信息行为主体在道德观念、道德意识、行为规范支配下,做出的自觉行为,部分是在破坏大数据环境,破坏社会的公平公正。随着社会经济、政治、文化多方面的进步,伦理道德却没有跟上大数据发展的步伐,相对滞后。

(2) 经济利益的影响

马克思主义理论中较为重要的一部分就是经济因素。“在《德意志意识形态》这本著作里,马克思和恩格斯提出了由‘物质生产——交往形式(生产关系)——理论和意识形态’序列组成的科学的社会结构理论。正是马恩最早发现了‘经济基础决定论’的观点。”大数据时代所产生的信息伦理问题一个根本的因素就是经济因素。

林德宏教授于《“双刃剑”解读》一文中指出:“求利的本性使得技术成为人们谋取物质利益的物质手段,因为人的利益问题,技术应用往往会带来负面效应。”在大数据时代,大数据

带来了巨大的经济效益,有部分信息行为主体因为眼前的经济利益或者让经济利益最大化,非法收集他人信息,侵犯他人隐私,大肆篡改他人信息进行黑市交易等,无视政府监管、国家法律法规和大数据信息环境,借助大数据技术为收集信息带来的快捷,利用大数据信息速度快的特性,不顾他人利益和社会公德,一味追求最大经济利益,在给自己带来经济利益的同时却造成他人信息被侵害,加剧社会矛盾,导致大数据信息环境混乱。例如,2017年4月《财经杂志》报道了《追踪“数据堂”:特大侵犯个人信息专案,震动大数据行业》,引起人们极大关注。这一数据信息大案件涉及在国内非常有名的大数据公司数据堂(北京)科技股份有限公司的多名相关工作人员,警方抓获的其中5人涉嫌传播敏感公民个人信息20余万条,涉案资金高达20万元,警方还发现这5人并非数据交易的源头,背后还涉及其他上游买家。最后据检方起诉书上称数据堂(北京)科技股份有限公司向买方提供了包含公民个人信息数据60余万条,买方又向需求客户发送个人信息168万余条。该数据信息大案涉及泄露他人隐私,信息量大,金额高,在大数据行业内引起巨大震动。

随着经济的发展,有些人变得越来越浮躁,认为追求经济利益就是在追求幸福,金钱无所不能,越来越觉得获得经济利益,内心的满足感就越强。长此以往,对幸福的理解会产生偏差,实际上,物质价值不能替代也不可能超越精神价值,再充足的物质条件,若不追求精神价值,内心也是空虚的,精神也是没有灵魂的。

(3) 隐私意识与保护能力的缺乏

大数据要发展,数据信息共享是必须要满足的。伴随着社会经济的发展,个人独立性越来越强,自我权利保护和自我财产占有意识也越来越强。为了社会发展,个人信息是否应该被共享,就成为隐私认识的一大挑战。每位公民都有隐私权,是从出生便应享有的一种自然的权利。隐私权是公民诸多权利中的一种,与之息息相关的是每位公民的独立和自由。隐私的最高价值就是个人的人格尊严和自由独立,所以可以说隐私是无价的。美国是重视隐私的典型和先驱国家,制定了《个人隐私法》,德国及其他一些欧盟国家也将隐私权列为个人权利并写入宪法文件中,让隐私权利得到法律保护。如美国《个人隐私法》规定:“个人隐私权是受合众国宪法保护的一项个人基本权利。”隐私权是个人权利,隐私也是属于个人受保护的对象,隐私对自己而言是无价的,例如家庭情况、信贷情况、购物情况等,这是个人的权利,他人若非法获取,是侵犯信息主体权利。但在大数据时代存在部分公民不知道什么是隐私,隐私的价值是什么,哪种情况侵犯了隐私,甚至都不知道自己泄露了自己及他人隐私。

隐私意识是指“人们在如何对待自己隐私而形成的一种思想意识,表现为尊重自己隐私并保护他人隐私的一种心理态度”。隐私问题是大数据信息伦理中常见的伦理问题,隐私保护意识薄弱是大数据信息隐私问题的重要原因之一。例如,在下载日常生活常用软件时,需要填写个人信息,为了生活便捷,大多数公民会选择如实填写个人信息,这些个人信息包括个人姓名、家庭成员情况、电话、住址等,但是却忽略了软件还设置了让个人主动授权使用信息的小功能,大多数人没有拒绝的意识,就直接勾选授权,软件就能够轻松获取填写的个人信息,甚至获取公民在第三方的记录(购物情况、信贷情况等)。这些信息如果被非法利用,后果严重。

公民隐私被窃取,还有部分原因是缺乏隐私保护技能。在我们实际的生活中,手机和电脑是常用的网络工具,在大数据环境下,无论是使用手机还是电脑,都应有良好的使用习惯。一是有些人没有习惯安装杀毒软件,或者对杀毒软件提示的高危风险不予重视,不及时修

复;二是对病毒软件及网站没有分辨和防范意识,随意登陆,对风险提示也没有重视,对高危漏洞认识不足,不重视扫描和杀毒;三是网络应用技术差,对某些应用软件应该设置用户密码的而没有设置,或者没有设置相应保护力度强的密码;四是有些人对云端信息的安全风险认识不足,将个人隐私信息上传云端。

(4) 大数据信息过程复杂化

大数据信息生产是大数据发展的前提。大数据信息生产内容范围广,形式变化快,主体差异性大,生产的有些信息真假难辨,大数据信息生产难以掌控。

大数据信息分布广泛,流动性强,信息生产主体各不相同,生产主体的职业、道德、文化层次都不同。目前大数据信息存储技术发达、存储方式多,存储载体形式多样,手机、电脑、U盘、光碟、邮箱、网站、信息平台等均可随时存储大数据信息,也正是因为存储范围广、形式多样,信息安全就变得不可控。每个公民都有一个或多个信息存储载体,而这个信息存储载体随时都在接收和发放大数据信息,并且还会从一个存储载体转移到另一个存储载体,在用信息存储载体接收、发放或者传输大数据信息时,就存在空白空间,信息很有可能会在这空白空间被他人窃取。特别是以信息系统或者信息平台作为信息存储载体的,安全风险更大,需要配套安全软件实时监控。

大数据信息使用方式灵活多变,大数据信息的使用方式是由信息使用人决定的,大数据信息的使用主体有个人、企业、社会团体、国家机构等,使用主体的道德有高低,使用方式也有不同,大数据信息产生的价值也各异。大数据发展的前提是信息共享,国家机构运用大数据信息,分析方向,做出各行各业政策引导;企业可使用大数据信息获取正当利益,有坚持大数据信息伦理,维护大数据市场环境的;也有因法律法规不健全,大数据信息伦理意识淡薄,对大数据信息进行“垄断”的,甚至在黑市上交易,获取不正当利益的。个人使用大数据信息,个体有贫富和地域差异,有尊重大数据信息,合理利用,正常发挥大数据信息作用的;也有使用大数据信息谋求私利,损害他人利益的。信息使用主体使用大数据信息方式多样,不易掌控,也是大数据信息伦理问题的原因之一。

(5) 大数据发展与法律监管的逐渐完善

大数据技术促进社会发展,也带来诸多的信息伦理问题,并且这些伦理问题还没有被纳入法律法规治理范围之中,没有相关配套的法律法规予以制止和惩戒。虽然早在2013年我国便颁布了《信息安全技术公共及商务服务信息系统个人信息保护指南》,该指南对信息安全技术涉及的个人信息保护给予明确指导,指出在使用大数据进行个人信息采集时,需要得到信息主体的授权,特别是使用个人信息时,更要得到信息主体的授权,才是合法有效。但对于大数据信息伦理的治理还是不够的,特别是隐私保护的治理。加之该指南只是引导,没有法律量刑,没有强制性,很难得到信息使用者的配合。其他的信息伦理问题也缺少针对性的法律法规。

直至2017年6月1日,《中华人民共和国网络安全法》才得以正式实施,这是我国第一部全面规范网络空间安全管理方面问题的基础性法律。《网络安全法》明确了网络空间主权的原则、网络产品及服务提供者的安全义务和网络运营者的安全义务,完善了个人信息保护规则,建立了关键信息基础设施安全保护制度。然而《网络安全法》并不完美,有些规定比较粗糙,部分细节还需要慢慢完善。2021年9月1日,《中华人民共和国数据安全法》正式实施,该法是我国数据安全领域的基础性法律,与《网络安全法》《个人信息保护法》等共同构成

了中国数据治理和保护的核心法律框架,以规范数据处理活动,保障数据安全,促进数据开发利用,保护个人、组织的合法权益,维护国家主权、安全和发展利益。

3.4 信息法律

3.4.1 信息法律基本概念

关于信息法律,目前最完整的定义是:信息法律是在调整信息活动中产生的各种社会关系的法律规范的总称。其调整对象是在信息活动中产生的各种社会关系,如信息获取关系(或称采集、收集)、信息加工处理关系(或称整理)、信息传播关系(或称传递、传输和信息存储)等。信息法律的内容包括知识产权法、信息安全法、信息公开法、电子商务法、新闻出版与传播法、有关计算机犯罪的法律等。

3.4.2 信息法律与信息伦理的关系

比较信息伦理与信息法律的关系,可以发现两者最根本的区别是自治与他治的问题。

“自治”,意味着对自己的行为进行自我约束和自我调整,即内在约束。“他治”,则意味着有针对外在行为统一而明确的管理标准和决策,即外在强制。信息法律与有组织的国家强制密切相关,专门机构、暴力后盾、程序设置、行为对象、物质结果等构成其外在强制标志。它以权利、义务为实质内容,有明确的行为模式和法律后果。信息伦理并非依靠专门机构权威性的力量保证,它在本质上是良心和信念的自由,主要依靠社会舆论的力量、人们的内心信念、良知、传统和教育的方式来维系,无特定、具体的表现形式。通常会体现在一定的舆论、传统和典型行为。

另外,信息法律与信息伦理之间是相互关联的,它们相互组合有四种情况。

首先,伦理且合法和不伦理且非法这两种情况。信息法律和伦理是作为维系信息活动秩序的两大支柱,它们所调整的社会关系在内容中有一定的重合。由上述两种情况可以推出,在这些关系中信息法律要求和鼓励的行为,也是信息伦理培养和倡导的行为;而信息法律禁止和制裁的行为,也是信息伦理所禁止和谴责的行为。也就是说,信息法律和伦理有共同的目标,即共同约束、规范信息行为,保障信息领域的正常秩序,最终达到使信息社会良性运转和协调发展。

其次,伦理但不合法和不伦理但合法这两种特殊情况。可以理解为对于符合信息伦理但不合法的行为,信息法律可以加以制裁。对于合法却不道德的信息行为,信息伦理可以予以规范。由此,可以看出信息伦理与信息法律之间具有相互补充的关系。相对于社会信息问题的具体多变和千态万状,信息法律在全面性和适应性方面存在一定的不足,而伦理的调整范围几乎覆盖社会生活的每个领域。所以对法律无法涉及的信息行为,信息伦理可以通过社会谴责、舆论力量等形式来加以规范。同样的,由于信息伦理的软约束性,又可依赖信息法律的强制力来惩罚和遏制那些恶意的信息行为。由此,两者之间相辅相成,强化和巩固了它们对社会的规范调控作用。

最后,信息伦理在一定条件下可以实现法律化。信息伦理为信息法律提供精神支柱和道德基础,诸如正义、平等、公平等信息法律原则,其本身就是人类伦理观念的组成部分。因

此,一些具有稳定性的信息伦理能够上升成为信息法律。虽然信息法律在其立法进程上有明显的滞后性,常是在某种信息行为造成了严重的后果后,国家才会开始着手研究相关的立法和制裁问题,但这一过程对于弥补信息法律的不足以及加强信息法制的建设完善具有重要的意义。

3.4.3 信息法律主要内容

从世界各国信息立法的进展以及社会信息化秩序建构的需要来看,信息法律应包括以下一些基本内容:

(1) 信息作品著作权

信息作品是指具有信息特征和作用的作品。它对一定事物进行描述,反映事物的状态或特征。信息作品具有资源性、共享性、传播性、商品性、时效性和无形性等特征,能够带给人们精神上的享受,也带来经济效益和社会效益。众多信息作品都是智力活动的成果,是知识财产。同时,信息作品又极易被复制、传播,极易给作者带来损失。著作权法就是为了调整作品作者、作品使用者以及公众之间的利益矛盾而产生的法律制度。

随着网络的普及和信息技术的飞速发展,网络环境中信息作品的著作权问题日渐突出。尤其是计算机程序、数据库、多媒体等多种形式的新型电子信息作品,它们与著作权法所保护的传统作品相比具有一些新的特点,也带来一些新的问题。因此,需要对现有的著作权进行适当的修改和补充,以适应对网络信息作品的法律保护。

(2) 信息传播法律制度

信息传播对于信息社会,犹如血液流动于人的身体。信息传播的法律规范,在我国主要体现在对信息传播主体的组织规范方面和信息传播主体的权利义务规范方面。这些法律包括大量的办法、条例,是信息传播主体的行为依据。到目前为止,我国颁布的这类法律规范涉及著作权保护、新闻出版及广告、国家安全、网络等各个方面。

近年来,我国网络方面的立法进展迅速,诸如《中华人民共和国计算机信息系统安全保护条例》《中华人民共和国计算机信息网络国际互联网管理暂行规定》《电子出版物管理规定》《中国公众多媒体通信管理办法》《计算机信息网络国际互联网安全保护管理办法》《中华人民共和国网络安全法》等。其中《中华人民共和国网络安全法》作为我国第一部全面规范网络空间安全管理方面问题的基础性法律,是我国网络空间法治建设的重要里程碑,是让互联网在法治轨道上健康运行的重要保障。

(3) 信息获取和信息消费法律制度

信息获取和消费是信息利用的重要形式。信息获取权是一个仍存在不同认识的概念。一般认为,信息获取权有广义和狭义的理解。狭义的信息获取权是指为保证公民政治权利的实现,公民有权获取政府机关依职权产生、收集、归纳、整理的信息。广义的信息获取权是指信息主体依法获得政府信息、企业信息、公共机构及公益组织信息的权利。有些学者认为,信息获取权还应包括对某些私人信息的获取。私人信息的获取包含了两种情形:一是私人有权利获取他人、政府机关和其他组织掌握的有关其本人的信息;二是指公众有权获取特定的私人信息,例如政府机关的某些工作人员的私人信息。也就是说,广义的信息获取泛指对一切可依法共享的信息的获取。

信息消费,是指人们使用信息资料满足生产和生活需要的过程,换句话说,信息消费包括信息资料的使用过程,即对信息内容的吸收和利用,以及信息需要的满足过程,即保证吸收的信息与需要相匹配。在我国信息消费的提出和研究时间较短,法律上还没有直接的规定。但是,信息消费中的生活性消费,应当适用《消费者权益保护法》;信息商品的质量管理,适用《产品质量法》等。

(4) 信息技术专利权

信息技术的高智力性、高风险性和高商业性决定了对其知识产权予以保护的必要性。而且这种保护不能仅仅停留在思想的表达形式上,而必须保护思想本身。在知识产权制度中,著作权保护是对技术思想的表达形式,例如技术报告、设计图纸、计算机软件和专著等进行法律保护的有效方式,而专利权保护则是对信息技术方案实施法律保护的最有效方式。所谓专利权,是国家专利机关根据专利法的规定,授予专利申请人对某项发明创造所享有的在法定期限内的专有权。

(5) 信息安全与计算机犯罪

信息安全与计算机犯罪是当前信息立法的热点。随着信息技术的发展和互联网的普及,计算机犯罪数量急剧增加,犯罪手段多样化、智能化,单纯用技术手段难以防范计算机犯罪的发生。因此,在保证信息系统开放的前提下,各个国家均采取法律手段保障计算机系统和信息网络的安全,预防和打击计算机犯罪。信息安全已成为信息时代综合国力、经济竞争实力和生存能力的重要组成部分。从国家安全的高度制定、实施信息安全政策与法规,以构筑完整的国家信息安全体系,已成为近年来国际上信息安全立法的一个重要趋势。

3.4.4 我国信息法律现状

我国信息化的发展速度十分惊人,单从网民数量上看,已于2018年跃居世界第一。但在信息立法方面,我国在个人信息保护、信息产权与信息交易、信息公平发展等多个领域尚有不足、有待完善。

(1) 政府信息公开立法

我国目前的政府信息公开法是2007年4月5日公布,自2008年5月1日起施行的《中华人民共和国政府信息公开条例》,共38条,是目前我国信息立法领域最显著的成果。该条例于2019年4月3日中华人民共和国国务院令第711号修订,自2019年5月15日起施行。

(2) 信息产权和信息交易立法

我国在信息产权方面的立法,先后颁布实施了《商标法》《专利法》《著作权法》《计算机软件保护条例》《计算机软件著作权登记办法》《关于制作数字化制品的著作权规定》等。同时为了与国际信息市场接轨,我国先后加入了世界知识产权组织及《保护工业产权巴黎公约》《世界版权公约》《商标国际注册马德里协定》《保护文学和艺术作品伯尼尔公约》《专利合作条约》等国际性知识产权保护公约。在信息交易方面,我国原《合同法》主要调整一般商品交易,虽然没有明确排除适用于计算机信息交易,但是其中并没有明确的规则来调整信息交易。我国《著作权法》中的著作权包括软件著作权,强调的是对任意复制的控制,并没有涉及计算机信息的内容。原《合同法》中有关技术合同的部分涉及了信息的内容,该法第349条

规定:“技术转让合同的让与人应当保证自己是所提供的技术的合法拥有者,并保证所提供的技术完整、无误、有效,能够达到约定的目标”,与美国《统一计算机信息交易法》对计算机信息的内容担保(准确性)的规定一致。但技术只是信息的一小部分,而且还不一定是计算机信息。总结是,我国现行法律中也没有能够很好调整信息交易的法律。

(3) 个人信息保护法

2004年以来,我国陆续修正或出台的《宪法》《民法通则》《合同法》《居民身份证法》《档案法》《民事诉讼法》《刑事诉讼法》《行政诉讼法》《商业银行法》《互联网电子邮件服务管理办法》《个人信用信息基础数据库金融机构用户管理办法》《短信息服务规范》等法律法规,都在其相关条款中不同程度地涉及了个人信息保护的立法问题。2009年2月28日公布实施的《中华人民共和国刑法修正案(七)》,在刑法第二百五十三条后增加一条,作为第二百五十三条之一:“国家机关或者金融、电信、交通、教育、医疗等单位的工作人员,违反国家规定,将本单位在履行职责或者提供服务过程中获得的公民个人信息,出售或者非法提供给他人,情节严重的,处三年以下有期徒刑或者拘役,并处或者单处罚金。窃取或者以其他方法非法获取上述信息,情节严重的,依照前款的规定处罚。单位犯前两款罪的,对单位判处罚金,并对其直接负责的主管人员和其他直接责任人员,依照各该款的规定处罚。”2021年8月20日,十三届全国人大常委会第三十次会议表决通过了《中华人民共和国个人信息保护法》,自2021年11月1日起施行。个人信息保护法明确了个人信息收集、处理和使用的规则,对个人信息处理者的义务等内容进行了明确的规定。《个人信息保护法》的颁布在我国个人信息保护立法上有着里程碑式的作用。

(4) 信息安全立法

我国在1997年刑法285条设立了非法侵入计算机信息系统罪,《刑法修正案(七)》中又增加了“违反国家规定,侵入前款规定以外的计算机信息系统或者采用其他技术手段,获取该计算机信息系统中存储、处理或者传输的数据,或者对该计算机信息系统实施非法控制”,“提供专门用于侵入、非法控制计算机信息系统的程序、工具,或者明知他人实施侵入、非法控制计算机信息系统的违法犯罪行为而为其提供程序、工具”两种行为作为构成侵入计算机信息系统罪的法定情形,清除了以往司法实践中打击网络“黑客”的法律障碍。刑法286条规定了破坏计算机信息系统罪。此外,1994年2月18日由国务院颁布施行的《中华人民共和国计算机信息系统安全保护条例》、1997年12月30日由公安部发布《计算机信息网络国际联网安全保护管理办法》、2000年3月30日由公安部颁布实施的《计算机病毒防治管理办法》等法律法规对侵害信息安全的计算机犯罪行为也有具体规定。2016年11月7日,第十二届全国人民代表大会常务委员会第二十四次会议通过《中华人民共和国网络安全法》,自2017年6月1日起施行。《中华人民共和国数据安全法》已由中华人民共和国第十三届全国人民代表大会常务委员会第二十九次会议于2021年6月10日通过,自2021年9月1日起施行。

3.4.5 国外信息法律现状

现代信息法律以研究信息社会中现代信息技术带来的新问题为中心。以计算机和网络技术为代表的现代信息技术发源于西方,信息社会概念也首先由西方学者提出。国外关于

信息法律的研究由来已久,在部分国家已经形成了比较有代表性的立法成果。信息社会在一定意义上是全球社会,它大大增强了世界各国之间的信息交流,加强了政治、经济和文化联系,各国在信息社会面临的法律问题具有较强的共通性。

(1) 美国信息法律

美国是世界上信息技术最发达和信息化程度最高的国家之一,在信息立法领域也是颇多建树,形成了一些比较有代表性的信息法律法规,影响较大的有《信息自由法》和《统一计算机信息交易法》。美国《信息自由法》(Freedom of Information Act, FOIA)主要规定人们在获得行政情报方面的权利,于1966年6月通过。制定该法的目的是督促政府机构向公众公开所持有的档案、文件等信息,以实现对政府权力的监督和制衡。随着社会情况的不断变化,该法历经1974、1976、1978、1984、1986、1996、2002、2007、2016年等数次修改,不断扩大公开程度或缩小例外范围,逐步趋于成熟。为充分实现信息公开的目的,该法在立法形式上,并非罗列何种信息应予公开,而是明确规定不予公开信息的例外情况,除此之外一切信息都必须公开。这样既可保证公众获知绝大部分政府信息,又可避免信息公开对其他权益造成损害,内容也更为简明、清晰。其规定的例外范围有:加密信息、内部人事务、其他法律授权不予公开、机密信息、内部决策、个人隐私、法律执行、金融业及其规章相关信息,以及矿井方面的地球物理信息(石油、水等)。随着现代信息技术的高速发展,增修后《信息自由法》更加适应这些技术发展所带来的许多挑战。首先,越来越多的信息平时就可以在机构的网页上获得。随着执法分支管理程序的规范化,公民获取信息已经极为简便。其次,增修条文明确了“文档”一词包括电子文本文档,只要申请者愿意,他们可以获取所需信息的电子文本格式;机构必须尽其合理的努力来检索其电子数据库,满足申请者。美国《信息公开法》对政府透明做出了极大贡献,其成功主要归功于其三个基本的特征:① 涵盖范围的普遍广泛性,这既表现在可以请求查阅的文档类型方面,也体现在可以申请查阅文档的个体的多样性;② 申请过程相对简单、廉价和迅捷;③ 对是否公开信息决定的独立审查。

《统一计算机信息交易法》(Uniform Computer Information Transactions Act, UCITA)由美国统一州法全国委员会(NCCUSL)起草,于1999年7月通过并于2002年进行了修订,作为示范法供美国各州立法采用。UCITA是美国最重要的一部关于网上信息交易的法律,它首次将以计算机信息为交易标的的直接电子商务纳入了法律调整范围,它的许多概念界定也具有开创性。UCITA将计算机信息界定为“是指以电子形式存在的信息,此种信息系由于或通过计算机的使用而取得,或是以计算机可以处理的形式存在。该术语包括上述信息的拷贝,及与此种拷贝相关联的任何文件或文件包”。计算机信息交易被定义为“就计算机信息或其中的信息权力的生成、修改、转移或许可使用而达成的协议,以及对此种协议的履行”。UCITA最大的贡献是提出的“信息权”的概念,它包括调整专利、版权、掩膜作品(mask works)、商业秘密、商标、公开权的法律创设的信息上的所有权利,或根据授予某人不依赖合同,基于权利持有人的该信息上的权益,控制信息或阻止其他人使用或访问信息之权利的任何其他法律创设的所有权利。信息权的概念对于大陆法系也非常有借鉴价值,它可以弥补物权法、知识产权法和债权法对信息保护的不足,对于构建完整的大陆法系财产权体系具有补白作用。UCITA的立法目的是建立电子合同的实体法律框架,以推进计算机信息交易,为此UCITA规定了一系列关于电子合同的程序性和实体性规范,包括法律选择条款、管辖法院的协议选择条款(有效解决网上交易双方匿名和无址情况下的管辖问

题)、电子合同成立方式的条款、当事方确立电子合同条款的方式、点击合同的可适用性、支配访问合同的条款,以及对电子自助的限制条款等,使得合同能以电子形式订立和履行。UCITA 可用于软件许可或购买合同、软件开发合同,以及通过因特网访问数据库的合同等内容。UCITA 中的许多规范是任意性规范并存在一系列违约规则。计算机信息交易的当事方可以自己拟定许可证的基本条款,如使用范围、许可费等,而将未约定的内容,如担保等条款由该法规定的违约规则来调整,或者仅做些必要的修改或选择。UCITA 中仅有少数规则,如善意行事、消费者保护等规则是强制性的,不得通过当事方的协议变更,其基本原则仍是合同自由。从这个意义上来看,UCITA 是一部合同法。

(2) 俄罗斯信息法律

俄罗斯信息立法最重要的成果是 1995 年 5 月 25 日通过生效的俄联邦《信息、信息化和信息保护法》。该法于 2006 年 7 月 8 日进行了修订。该法确立了一些信息立法领域基本的法律规则,包括对经济发展的信息保障、国际信息交流、信息资源管理、部门和地区信息化以及信息安全,是俄联邦信息领域的基本法。其主要调整的社会关系范围包括:组织信息资源并提供利用,建立和应用信息系统、信息技术及保障手段、保护信息和信息化环境中的主体权利。

(3) 德国信息法律

1997 年 6 月 13 日,德国联邦议院通过了《规定信息和通信服务的一般条件的联邦法令——信息和通信服务法》(德文简称 IUKDG),通常被称为《多媒体法》,于 1997 年 8 月 1 日开始实施。该法由三个新的联邦法律和六个将现有法律适用于新媒体的附属条款所组成。三个新的联邦法律分别为:《远程服务法》(The Tele-Service Act)、《数据保护法》(The Data Protection Act)和《数字签名法》(The Digital Signature Act)。该法的立法目的是为建设自由的信息和通信服务市场提供一个可靠的法律基础,即通过为信息产业创立一个稳定而可靠的法律环境刺激人们在信息产业方面的投资,同时通过市场竞争保护该领域的创新和创造。另外,该法还希望保护用户的公民权利和公共的利益。该法的内容极为广泛,包括了对 ISP 责任的规定,使用数据保护的方法保护个人隐私,为安全数字签名的签署和管理构筑官方框架。加大了对违禁出版物和网络犯罪的打击力度,保护未成年人远离不良信息,数据库保护等内容,是一部全面的综合性法律。

(4) 欧盟信息法律

欧盟信息立法的最重要成果是 1996 年通过,于 1998 年 1 月 1 日实施的《欧洲议会与欧盟理事会关于数据库法律保护的指令》(EU Database Directive,以下简称“指令”),在立法方面的重要贡献是创设了对数据库的二元保护机制。“指令”将数据库定义为“是指经系统或有序的安排,并可通过电子或其他手段单独加以访问的独立作品、数据,其他材料的集合”。包括我国在内的多数国家将数据库作为汇编作品,利用版权法加以保护,而作品具有独创性是版权法保护的前提。按照版权法的观点,汇编作品的独创性体现在对汇编材料的选取和编排上。随着信息社会信息量的爆炸性增长和信息处理能力的提高,现实生活中出现了大量不具有独创性的数据库。这些数据库多数是对从公开渠道获得的信息堆叠,但不可否认的是这些数据库同样是制作者付出劳动和资本形成的,并在现实生活中发挥着巨大的作用。但按照版权法的观点,这些数据库却因缺乏独创性而无法获得版权法的保护。“指令”在承

认具有独创性数据库的汇编作品地位和可受版权法保护的同时,创造性地利用“特殊权利”对未达到独创性标准的数据库进行保护。数据库受特殊权利保护的条件是制作数据库的投资或成本必须达到一定数额,特殊权利的主要内容是制作者有权禁止他人提取和再利用数据库的全部内容或实质性部分,鉴于数据库在现代社会的重要作用,创设特殊权利的主要目的是保护制作者为数据库付出的巨大投资,对其投资成果进行确认,以达到鼓励数据库持续投资的目的。这与产权的权利目标极为相似,可以看出无独创性数据库享有的特殊权利,也是信息产权的一种。

由于在信息时代国际间的信息交流增多,导致一些关系到国际间信息安全的国际信息法规出台。1971年联合国提出了《世界知识产权组织关于保护计算机软件的示范条例》,经济合作与发展组织也在1985年出台了《跨国数据流宣言》,地区间的相关信息安全条例和法规可以说不胜枚举。

3.5 思考题

1. 信息素养的内涵及作用是什么?
2. 具有良好信息素养的表现有哪些?
3. 当前医学人才培养对信息素养提出了哪些要求?
4. 信息法律与信息伦理之间的关系是什么?



【微信扫码】

本章相关资源

南京大学出版社
版权所有

第4章

信息检索基础

4.1 信息与信息源

4.1.1 信息的概念

从自然界角度看信息表征物质的属性:地球昼夜的变化是一种信息,它反映地球自转的运动特性和状态;山的高度是一种信息,它反映山的空间特性;树干的年轮是一种信息,它反映树木成长的时间特性——树龄;闪电是一种信息,它反映云层中所含能量的不均匀性;花的香味也是一种信息,它反映花分子结构的化学特性,等等。

一般认为,信息是物质存在的一种方式,是事物运动的状态与方式的反映。

关于信息的定义目前不下百种,人们从不同的学科和不同的角度进行了不同的阐述。经济学家认为:信息是可以与材料、能量相提并论的用以维系人类社会存在和发展的三大要素之一,是为管理和决策提供依据的有效数据。心理学家认为:信息是存在于意识之外的东西,它存在于自然界、印刷品、磁盘以及空气之中。新闻界认为:信息是对事物运动状态的陈述,是物与物、物与人、人与人之间的特征传输。哲学家认为:能被其他事物感知的,表征该事物特征的信号内容即为该事物向其他事物传递的信息。图书信息领域则把信息定义成是可以以各种形式进行传播、记录、出版及发行的观念、事实及论著。

信息的定义之所以呈现多样化,主要原因有三个方面:第一,信息本身具有复杂性,它是一个多元化、多层次、多功能的综合物;第二,信息科学是一门新兴学科,它的许多分支学科仍在随着社会、经济和科学技术的发展而发展,其内涵和外延不是很确切;第三,人们出于不同的研究和使用目的,从不同的角度或层次出发,对信息概念就会作出不同的解释。

因此,许多学者又建议将信息分为不同的层次来阐述。其中,最有影响力的是关于两个层次的论述:一是没有任何约束条件的本体论层次,即信息是一种客观存在的现象,是事物的运动状态及其变化方式的表征,不受主体意志的影响,不停运动着的事物不断产生本体论意义上的信息;二是受主体约束的认识论层次,即信息就是主体所感知或所表述的事物运动状态及其变化方式,受主体意志的影响,是主体大脑反映出来的客观事物的属性。

综合起来,可以作如下概括:一切事物,包括自然界和人类社会都会产生信息。信息是

客观世界各种事物特征和变化的反映,以及经过人们大脑加工后的再现,消息、信号、数据、资料、情报、指令均是信息的具体表现形式。从信息的本质和重要性来看,信息是区别于材料与能量的第三种资源;从信息的作用来看,信息能消除人们对所描述事物的不确定性;从信息的内容来看,信息描述了事物的运动状态与方式。

4.1.2 信息的特征

1. 普遍性和客观性

信息是事物存在方式和运动状态的表现。事物的存在和运动无处不在,因而信息是普遍存在的。信息与物质、能量一起,构成了客观世界的三大战略资源。客观性表现为:信息不是虚无缥缈的东西,它的存在可以被感知、获取、存储、处理和利用。而且,事物的客观存在先于认识主体的存在,信息由其本身特性及运动而产生,不依主体感知而改变。

2. 依存性和传递性

信息产生于事物的存在和运动,信息的产生对事物及其运动有依存性。但信息一旦产生,便可以独立于其发生源而相对独立,由其物质载体携带在时间或空间上传递。同样内容的信息可以在不同物质载体间转换。信息的传递对载体也有依存性,必须依附于一定的物质形式(如声、光、电、电磁波、纸张、磁性材料等)。信息在时间上的传递即是信息的存储,在空间上的传递就是通信。信息在时间和空间上传递的性质十分重要,让人类社会不仅能够进行有效的信息交流和沟通,还能够进行知识和信息的积累与传播。

3. 可知性和共享性

信息作为客观事物的反映,可被人类通过各种手段和方式所感知与认识。世界是可以被认识的,在此前提下,才有随之而来的对信息进行存储、加工、传递和利用,信息也由此成为人类认识世界的基础与前提。但由于人们的观察能力、认识能力、理解能力和利用信息的目的不同,抑或由于人们观察事物时选择的角度不同,关注点不同,人们从同一事物中所获得的信息内容和信息量也各不相同,由此形成了人类对客观世界丰富、多样、各异的认识。

4. 时效性和价值性

从信息产生的角度看,信息所表征的是特定时刻事物存在的方式和运动状态,由于所有的事物都在不断变化,过了这一时刻,事物的存在方式和运动状态会改变,表征这一“方式”和“状态”的信息也会随之改变,即所谓时过境迁。从信息利用的角度看,某些信息仅在特定的时刻才能发挥其效用。从信息的产生到被利用的过程,时间越短,其利用价值越大,反之越小。利用价值大的商业类信息,往往能够带来巨大的经济效益和社会效益。

4.1.3 信息、知识、情报与文献

1. 知识

知识(knowledge)是信息接收者通过人脑对信息的提炼、推理、总结和归纳而获得的结论,是人的大脑通过思维在认识世界和改造世界过程中对自然界、人类社会以及思维方式与运动规律的重新组合,并在此过程中形成对客观事物的现象、本质及规律的认识和经验。从这个意义上讲,知识也是信息,是信息的高级形态。简单而言,知识是人类认识的结果或精

华。运用知识,实际上就是遵循了事物的客观规律,这是进一步认识与改造客观世界的最基本的前提和基础,知识蕴含着推动社会发展、人类进步的巨大力量。

知识的魅力在于能够给予人类以智慧和力量。然而,知识所具有的价值及其实现不仅取决于知识本身,更有赖于人类对知识的有效利用。适合需要的知识才有价值,利用得当的知识才能称其实现了应有价值。

2. 情报

情报(intelligence)一词首先产生于军事领域。十九世纪后,人们将情报概念引入科技领域,将情报定义为“人和人之间传递的一系列符号”“在特定时间,特定状态下,对特定人提供的有用知识”“情报是判断、意志、决心、行动所需要的能指引方向的知识 and 智慧”“被激活了的知识”等。

情报具有知识性,因为只有知识才能真正发挥作用;情报具备有用性,情报特指被指定用户利用的那部分知识,人们创造情报、交流情报的目的在于充分利用其价值来开阔眼界、改变知识结构、提高认知能力,帮助人们认识和改造世界;情报具有动态性,该属性往往是指那些在指定时间和范围内发挥作用的知識,知识若不进行传递交流,供人们利用,就不能构成情报。

因此,可以将情报定义为:用不同的方法或手段,在有效的时间内所获取的有针对性的,符合预期价值要求的知识或信息内容。具有知识性、有用性和动态性,而动态性是情报区别于知识的根本属性。

3. 文献

在我国,文献(literature、document)一词有两千多年历史,最早见于《论语·八偷》:“夏礼,吾能言之,杞不足征也;殷礼,吾能言之,宋不足征也。文献不足故也,足,则吾能征之矣。”宋代朱熹解释为:“文,典籍也;献,贤也。”“文献”一词被认为是宿贤和典籍的合称。随着科学技术的发展,文献一词的内涵和外延还在进一步扩展。比如,《文献情报术语国际标准(草案)》(ISO/DIS 5127)将文献表述为:在存储、检索、利用或传递记录信息的过程中,可以作为一个单元处理的,在载体内、载体上或依附载体而存储有信息或数据的载体。而《国际标准书目著录(总则)》[ISBD(G)]则将文献定义为“以任何实体形式出现的,作为标准书目著录的书目文献实体”。由这两个权威定义可知,文献包含四个要素:

- (1) 记录知识的具体内容;
- (2) 记录知识的手段;
- (3) 记录知识的物质载体;
- (4) 记录知识的表现形态。

其中,知识构成了文献的内容,是文献的本质特征;知识的记录手段指的是记录文献时所用的技术形式,如文字、图像、符号、视频、音频等技术形式;知识的物质载体则是指文献的形态,如纸张、光盘、录像带等;知识的表现形式则是指文献的形式,如图书、期刊、磁盘等。简单地说,文献就是记录有知识和信息的物质载体。

从定义来看,文献与信息 and 知识明显的区别是其物质载体,文献载体形式有石头、甲骨、竹简、各种材质的器物、布帛、纸张、胶片、磁带、磁鼓、光盘等,而文字、数字、图形、符号、音频、视频等表现知识和信息的方法或技术的发展与应用,使得无形的信息和知识能够转变为有形的物质形式,以文献形式记录的信息与知识方便了人的交流,并且使人类在生产实践中

所积累的认识经历和认识成果的传递与保存成为可能。文献的特点为数量庞大、增长迅速、分布不均、内容交叉重复、载体类型多样、更新速度快、语种多元化、交流速度快等。

4. 信息、知识、情报与文献的关系

信息、知识、情报、文献四者之间既有联系又有区别。

信息广泛存在于自然界和人类社会,是一种客观存在的事物现象,是事物发出的信号和消息总和,其涵盖范围最大最广。知识是经过人脑加工和提炼的信息。因此,知识是系统化、理论化的信息,是信息的一部分。不是所有的信息都是知识。信息具有客观性、动态性,而知识则具有主观性、静态性。情报则是实现特定效用和利用价值的动态性知识。但是在具体情报活动中所获取的情报资料,不一定能保证其真实性、正确性,所以情报在理论上应当是知识的一部分,但实际上其中既有知识,也有信息。文献则是记录信息、知识、情报的载体,是物化了的、信息、知识和情报,但不是所有的信息都会以文献形式保存下来,如图 4.1 所示。

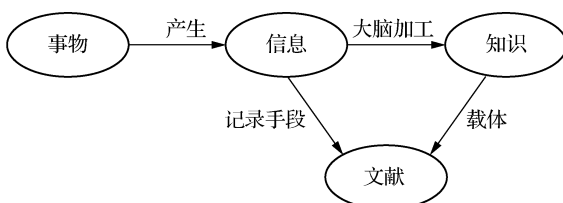


图 4.1 信息、知识、文献的形成示意图

文献记载的内容中,既有正确的部分,也有错误的部分;有我们特定活动所需的内容,其余大量的可能是我们并不关心或不需激活的内容。因此,文献是信息范畴的一部分,同时,文献与知识、情报都有重合的部分,如图 4.2 所示。

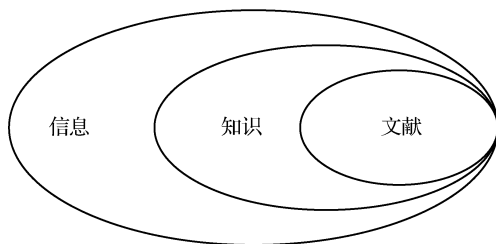


图 4.2 信息、知识、文献的包含关系示意图

综上所述,通过表征客观世界的客观信息的获取、加工等一系列思维过程,形成了反映客观事物本质和规律的具有主观性色彩的知识,将知识以某种方式系统化地记录于某种载体之上,形成文献。

4.1.4 信息源

1. 信息源的概念

联合国教科文组织 1976 年出版的《文献术语》将“信息源”定义为个人为满足其信息需要而获得信息的来源。一切产生、存储、整理、组织、传播信息的源泉都可以看作是信息源,也称为情报源。信息源内涵非常广泛,不仅包含各种信息载体,也包括各类信息机构;不仅

涵盖传统纸质版文献,也囊括现代电子文献;不仅包括信息生产机构,也包括信息储存、信息组织和信息传递机构。

2. 信息源的类型

(1) 按信息源产生的时间顺序划分

信息源按产生的时间顺序分为先导信息源、即时信息源、滞后信息源。先导信息源产生于社会活动之前,如天气预报;即时信息源产生于社会生产过程之中,如工作记录、实验报告等;滞后信息源是指已经失去利用价值的,如过期的报刊、报纸等。

(2) 按信息的加工深度及功能作用划分

零次文献(zero documents)也称灰色文献,是指未经加工而直接记录在载体上的原始信息,如实验数据、观察记录、调查材料、书信、手稿等。这些未融入正式交流渠道的信息,往往能够反映出现有的研究工作取得的新发现、新问题、新想法等,是启发科研人员的思路,形成创造性思维的优良素材。

一次文献(primary documents)也称原始文献,是人们以社会生产过程中形成的经验,科学研究活动中产生的科学和规律性认识,社会活动中获得的可靠、有价值的素材为依据撰写出来的文献。常见形式有专著、期刊论文、科技报告、专利说明书、学位论文等。该文献是目前数量最大、种类最多、内容最新、使用最广、影响最大的一类文献。

二次文献(secondary documents)也称检索型或工具型文献,是对大量零散无序的一次文献进行收集、整理、加工和组织,提炼出一次文献的外部形式(如题名、作者、出处等)和内容特征(如分类号、主题词等),并按一定的方法将以上形式和特征进行编目,使零散无序的一次文献有序化,形成用来报道和检索一次文献的检索工具或检索系统,为用户查找一次文献提供线索和途径,为科研人员节约宝贵的查阅文献时间。常见的有索引、目录、题录、文摘等。

三次文献(tertiary documents)也称参考型文献,是专业人员对零次文献、一次文献和二次文献的内容进行系统性的整理,综合分析,经过浓缩和提炼而编写的新的文献类型。具有信息量大、综合性强、系统性和参考价值高等特点,具体包括综述、述评、进展、现状、发展趋势等期刊文献类型,以及百科全书、年鉴、手册等参考工具书,其中,综述和述评是三次文献最基本的两种形式。

综上所述,各级别文献的形成及其相关关系可用图 4.3 表示。

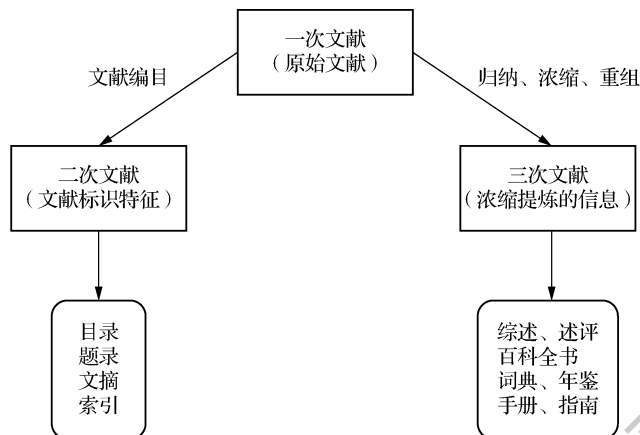


图 4.3 各级文献的形成及其相关关系

(3) 按信息源的传播形式划分

信息源按传播形式分为口传信息源、实物信息源和文献信息源。口传信息源,是非正式性交流,通过讨论、报告会等方式进行传播;实物信息源存在于自然界和人工制品中,通过社会生产、实践、劳动等方式进行传播;文献信息源存在于具体的文献中,通过阅读、视听、观看、触摸等方式传播。

4.1.5 科学活动中的信息源分类

对于科学活动中常见的信息源,可按下列标准进行分类。

1. 根据载体形式划分

(1) 书写型。一般指以龟甲、青铜器、竹简、绢、帛、纸张等为载体,通过人工抄写、刻写或铸造而形成的文献类型。

(2) 印刷型。主要以纸张为载体,通过印刷、刻录等技术手段而产生的文献类型,如传统的图书、期刊、报纸等。该类文献能够直接阅读、触摸,也是人们知识交流和文化遗产最重要、最常用的媒介。

(3) 缩微型。是以感光材料为载体,利用缩微摄影技术制成的文献,如缩微胶卷等,其特点是体积小、容量大、易于复制、携带和保存。

(4) 视听型。是以磁性或感光材料为载体,借助特殊技术直接记录声音和图像而形成的一种文献形式,该文献类型能够直接记录历史,保留历史影像,进行场景再现。

(5) 数字型。利用数字化技术把信息存储在磁盘、光盘等介质上,通过计算机及现代通信手段传播利用的一种信息类型,如电子图书、数字化期刊等,其特点是存储密度高、读写速度快、出版与传播速度快、检索便捷、能够突破时空限制。

2. 根据出版形式划分

此种划分方式一般将文献分为十大类,也称为十大科技情报源。其中除图书、期刊公开发行易于获取外,其他八种在出版发行和获取途径上相对特殊,常统称为特种文献。特种文献特色鲜明、内容广泛、数量庞大、参考价值高,也是非常重要的信息源。

(1) 图书(book)是对某一领域的知识或已有的研究成果及生产经验等做系统性论述,具有形式规范和内容可靠、全面、系统、科学、完整以及稳定等优点,因其通常目的是传授知识和传承文化,而不是传递最新的情报,所以出版周期相对较长、更新速度较慢。

(2) 期刊(periodical)是指具有相对固定的刊名、编辑出版单位、出版周期和报道范围,有统一的出版形式,按一定顺序编号逐期刊行的连续出版物。其特点是种类多、数量大、出版周期短、报道速度快、传递内容新颖、信息量大(约占整个信息量的70%),因此它是获取最新的科技动态和科技情报的重要来源。科技期刊与图书相比,能更快速地反映当前的研究前沿与科技发展水平,是重要的信息源,也是学术交流和传播活动中最重要和最稳定的媒介。判断学术期刊质量的重要方法就是看是否为核心期刊。

核心期刊(core periodical)是针对某一学科或专业领域,刊载大量专业论文和利用率较高的少数重要期刊。该类期刊能够反映该学科或领域的最新成果和研究动态。

(3) 会议论文(conference literature)是指在学术会议上发表、宣读或交流的论文、报告、会议纪要等资料,其形式可以是会议前预印本、会间文献和会后论文集,或者以大摘要方

式,以增刊或专辑形式刊发于某些学会刊物上(国外期刊中常见),其特点有内容创新程度高、专业性强、出版速度快、时效性强等。

(4) 学位论文(dissertation)是指按国家标准(GB/T 7713.1)定义为作者提交的用于获得学位的文献。学位论文包括学士论文、硕士论文和博士论文。一般而言,硕士论文、博士论文具有选题新颖、论述系统、实用性强、篇幅不受限制、拥有详尽的参考文献等特点,且涉及某一研究主题的研究背景、技术路线、实验方法、数据获取、分析结论,深入地探讨和论述该领域的现状、问题及对策,具有重要的阅读和参考价值。

(5) 专利文献(patent literature)是指受到法律保护的技术发明,是知识产权的一种具体体现形式。专利文献是实行专利制度的国家及国际性专利组织在审批专利过程中产生的官方文件及其出版物的总称,是受专利法保护的有关技术发明的法律文件。专利文献记载着发明创造的详细内容及被保护的技术范围,是集技术、法律、经济信息于一体的特殊类型的科技文献。其特点是内容新颖、涉及技术领域广泛、实用性强等,能反映科技的最新水平。

(6) 科技报告(scientific and technical report)根据国家标准(GB/T 7713.3)的定义,是指进行科研活动的组织或个人描述其从事的研究、设计、工程、试验和鉴定等活动的进展或结果,或描述一个科学或技术问题的现状和发展的文献。科技报告呈送给科学技术工作主管机构或科学基金会等组织或主持研究的科研工作者,以便有关人员审查和评价,并对报告中的结论和建议提出修正意见。科技报告旨在提供系统、翔实的信息,不以发表为目的,是科研历程及其成果的完整记载,其内容比期刊论文新颖、详尽和专深。科技报告在一定范围内发行,具有一定的保密性。

(7) 标准文献(standard literature)是指由政府或有关部门以及权威机构批准下发的标准化、技术性、法规性文件,简称标准,是从事生产、设计、管理、产品检验、商品流通、科学研究共同依据,在一定条件下具有某种法律效力,有一定的约束力。其特点具有适用范围明确、时效性强、有法律约束力等。

(8) 科技档案(technical records)是指在科研活动中所形成的一系列技术性文件,包括规划、方案、任务书、协议书、工程图纸、图表、实验记录、总结等,是科研活动的真实记录,能反映科研活动的全过程和最终结果。其内容真实可靠、详尽具体,具有较高的参考价值。

(9) 产品资料(product data)是厂商为其产品的性能、构造、规格、用途、使用方法等所做说明的相关技术资料,包括产品样本、产品说明书、产品目录和厂商介绍等。

(10) 政府出版物(government publication)是各国政府部门及所属机构发表各类文件的总称。一般分为行政性文献和科技性文献。前者主要包含政府法令、方针政策和统计资料等,涉及政治、法律、经济等多个方面;后者主要包含政府部门的研究报告、标准、专利文献、科技政策、公开的科技档案等。政府出版物对了解各国的方针政策,经济状况和科技水平具有较高的参考价值。

4.2 信息检索基本原理

最早的信息检索主要依靠信息分类,如两千多年以前,我国汉代的《七略》一书,是最早的图书分类法。随着现代信息技术的飞速发展,产生了海量信息,传统的手工分类已完全不能满足用户的信息需求,尤其是科技期刊的大量出现,已不可能将刊登在相关领域内的期刊

上的所有文献都阅读或浏览一遍。文献索引便应运而生,读者可以按照一定的方法和需要查找相关文献。信息检索人员根据文献的外部形式(如题名、作者、出版社等)和内容特征(如分类号、主题词等)编制二次文献(如书目、索引、文摘、目录等),为用户查找一次文献提供检索工具和检索途径。

随着文献的激增,用户需求的增长和信息技术的迅猛发展,传统的线性检索(先组式索引检索、穿孔卡片检索、缩微胶卷检索、脱机检索)很快被非线性的超媒体检索(联机检索、光盘检索和网络检索)所代替。

4.2.1 信息检索的概念

信息检索(Information Retrieval, IR),其定义有狭义和广义之分。广义的信息检索既包含信息存储,又包括信息检索或信息查询,即信息的“存”和“取”两个部分。狭义的信息检索只包含信息检索,也被称为信息查询。

广义的信息检索是指信息工作人员将信息按一定方式组织和存储起来,并根据信息用户的需要找出有关信息的过程,包括信息的存储和检索两个过程(information storage and retrieval),所以也被称为“信息存储与检索”。信息存储和信息检索两方面的内容是由两个方向相反而又相互依存的工作过程构成,即存储是检索的基础,检索是存储的目的。而狭义的信息检索仅仅指的是信息检索或信息查询,即信息用户根据自身需求查找相关信息的过程。对于信息用户而言,更加关注如何快速地获取所需的信息,狭义的信息检索对信息用户而言因而更为重要。

信息存储过程是指由信息工作人员将大量无序的、零散的信息集中起来,根据信息源的外部形式和内容特征,将信息进行整理、组织、分类、标引等处理,使之有序化、系统化和集中化,并按相关信息要求构建具有检索功能和反馈功能的检索工具或检索系统,以使用户检索和利用信息。

信息检索过程是指信息用户根据自己的信息需求,并依据一定的方法和策略,借助相关设备和工具,从已构建好的信息检索工具或检索系统中,查询所需要的信息,即信息查询。信息检索的本质是将用户的信息需求与检索系统或检索工具中的文献标识特征(外部形式和内容特征)进行匹配的过程。

4.2.2 信息检索的发展历程

1. 手工检索阶段(1876—1954 年)

信息检索起源于对传统纸质文献进行的手工检索和图书馆参考咨询。文献信息日益多样化,用户信息需求也愈发个性化,手工检索在文献编目等方面取得了丰硕成果,主要表现在以下两个方面:

一是—批高质量的文献型检索工具的出现。如 1884 年美国工程信息公司的《工程索引》(Engineering Index, EI), 1898 年英国电气工程师学会的《科学文摘》(Science Abstracts, SA), 1907 年美国化学文摘社的《化学文摘》(Chemical Abstracts, CA), 1926 年美国生物科学信息服务社的《生物学文摘》(Biological Abstracts, BA), 1961 年美国科学信息研究所的《科学引文索引》(Science Citation Index, SCI)等。

二是手工检索分类方法和分类语言的相继出现。在手工检索分类方法方面,如 1876 年

美国学者杜威创造的《图书馆图书小册子排架及编目适用的分类法和主题索引》，也就是后来的《杜威十进分类法》(Dewey Decimal Classification, DDC)，该分类法也是目前世界上应用最为广泛的分类法；1901年在时任美国国会图书馆馆长普特南的主持下编制而成的《美国国会图书馆分类法》(Library of Congress Classification, LL. C)；1907年由比利时学者奥特莱和拉封丹在《杜威十进分类法》的基础上改编而成的《国际十进分类法》(Universal Decimal Classification, UDC)。在手工检索分类语言方面，先后出现了先组式语言、后组式语言、叙词语言等。以上成果为计算机信息检索的发展奠定了坚实的理论基础。

手工检索虽然具有操作简单、费用低廉、查准率高等优点，但也存在着检索效率低、查全率低等缺点，并且随着科学技术的迅速发展，对传统的纸质文献进行手工检索的方式早已不能满足用户个性化的信息需求，该检索方式很快被计算机检索所取代。

2. 计算机检索阶段(1954年至今)

20世纪40年代，诞生了世界上第一台计算机，随着计算机在信息检索领域的应用，传统的手工检索阶段进入了计算机检索阶段。计算机凭借着强大的信息存储能力、快速的检索能力，稳定可靠的信息输入能力，使人们能够快速、准确地从海量信息中获取满足自身需求的特定信息。

计算机信息检索是将信息组织和存储在磁盘或光盘上成为数据库，用户利用计算机从数据库文档中找出所需的信息。广义的计算机检索包括信息存储和信息检索两个过程，狭义的计算机信息检索仅包含检索过程。不论是广义的还是狭义的，其检索原理都是：将用户检索提问标识输入计算机检索系统，后者将前者与数据库中的存储特征标识进行匹配，若用户提问标识与数据库中的存储特征匹配，即为命中信息，并由计算机检索系统通过一定形式反馈给用户。

计算机信息检索具有其独特的优势：

① 检索速度快。传统手工检索方式在查找一个课题时往往需要花几天甚至十几天时间，而计算机检索只需几秒或几分钟即可完成。随着计算机处理技术和网络传输性能的提升，检索速度会更快，效率会更高。

② 检索内容新。目前论文发表均提供电子文档，已被录用的论文需要首先录入数据库进行标引供用户检索，然后再以纸质版形式形成检索工具。从录入数据库进行标引到纸质版形式的检索工具，其时间较长，通常情况下是两个月到半年。

③ 检索途径多。限于篇幅，手工检索工具一般提供索引、目录、文摘、年鉴、手册等检索方式，而计算机检索系统可以对所有外部形式和内容特征进行标引，并且后者还提供布尔逻辑检索、截词检索(或通配符检索)、字段检索、词组检索、加权检索等多种技术方法。

④ 资源共知共享。借助网络，用户可以不受任何时空限制，联机查询网上信息资源，检索世界上任意主机上的免费数据库，获取OA资源，实现全球资源的共享。

⑤ 输出方式多样。对于检索结果，计算机可提供屏幕浏览、打印、存盘等多种输出方式，还可以对结果进行再次筛选和排序，对每一条记录也可以限定输出某些具体字段。

随着计算机技术、通信技术以及存储介质的不断进步，计算机检索同样经历了由低级阶段到高级阶段的发展历程，即脱机检索、联机实时检索、光盘检索、网络化检索四个阶段。

(1) 脱机检索阶段(1954—1964年)

脱机检索系统是计算机检索初期使用的一种检索系统，利用独立计算机的输入和输出

装置进行检索,用磁带作存储介质。在使用该系统检索文献时,计算机只能按照一定的顺序检索磁带上记录的信息,每一次检索都必须从头到尾读取磁带上面的信息,耗费了大量时间。因此,检索工作人员不得不以批处理方式来进行查找。

首先,需要脱机检索用户先填写检索申请单,再由检索工作人员将之转换为计算机可读形式,批量输入计算机中进行查找。因此,脱机检索通常也被称为脱机批处理检索。20 世纪 50 年代中期到 60 年代中期,具有代表性的脱机检索系统有:1954 年美国海军建成的世界上第一个试验性计算机检索系统、1959 年美国人卢恩利用 IBM650 电子计算机建成的世界上第一个定题检索系统、1961 年美国化学文摘社第一次利用计算机来处理书目信息编制而成的《化学题录》。脱机检索的编印索引、书目,回溯检索和定题检索服务虽然在各个学科领域得到了广泛使用,但由于早期的计算机数据需要批量输入命令,须用穿孔卡片或纸带,存储介质主要是磁带,采用遍历式顺序的检索技术,缺少信息的输出终端,导致人机交互性不强、受时空限制等弊端,这些弊端极大地限制了脱机检索技术的发展。

(2) 联机实时检索阶段(1965—1975 年)

20 世纪 60 年代后期,随着计算机集成电路技术的发展,尤其是具有高密度存储等特点的磁盘、突破了时空限制的通信和操作系统等信息技术的相继出现,信息检索由此从脱机批处理阶段进入了联机实时检索阶段。人们可以在磁盘上建立起直接读取和存储的随机文件,并开发了一台主机通过通信线路连接到多个终端的联机检索系统。用户可借助于命令式的检索程序通过用户终端以“即问即答”的方式与计算机交互,从而实现跨时空的检索,最终满足自己的信息需求。

联机检索系统主要由用户终端、通信网络、计算机及数据库组成。用户利用计算机终端设备,通过通信线路与存储数据库的中央计算机相连,直接进行人机对话。联机检索直接面向最终用户,检索过程通过人机对话交互方式可随时修改检索策略,能及时获取检索结果。该阶段的主要成就有:1969 年,NASA 开发的第一个大规模联机检索系统 RECON 全面投入运行;1970 年,美国洛克希德公司和美国系统发展公司分别开发了 DIALOG 系统和 ORBIT 系统,同年,美国国立医学图书馆将 MEDIARS 系统发展为医学文献分析与联机检索系统;1973 年,欧洲宇航局建成的 ESA—IRS(Europe Space Agency Information Retrieval Service)系统等投入运行,是欧洲最大的联机情报检索系统。

联机实时检索系统虽然取得了丰硕的成果,但存在很多缺点,如指令复杂、检索工作必须依赖专业人员完成、联机检索收费高、国际通信费用昂贵等。该检索系统随着信息技术的飞速发展,尤其是集成电路技术的不断发展,逐渐发展为光盘检索系统。

(3) 光盘检索阶段(1976—1990 年)

20 世纪 70 年代中后期,出现了以娱乐为主的光盘(Compact Disc,CD),但直到 80 年代中期,相关研究人员才开发出第一种有关信息检索的光盘,并逐步应用到信息检索领域,逐渐形成了光盘检索系统。该检索系统在早期发展阶段只是单机驱动和单用户检索,为满足多用户同时检索的需求以及同一数据库同时检索多张光盘的要求,陆续出现了光盘塔、光盘库及光盘网络等技术,这些技术的发展使信息检索发生了革命性的改变。

光盘检索是以光盘作为存储介质,通过光盘驱动器来查找信息的一种检索方式,将检索软件及数据存储光盘上,任何一台安装有光驱的计算机,都能在光盘数据库的支持下,成为光

盘检索系统。该检索系统具有以下优点:存储容量大但体积微小,无须通信联网,不受时空限制,使用简单,易于操作,无须专业知识即可安装和检索,设备投资少,检索价格低廉(一次购买,无限次检索),节省大量通信费用,易存储,使用寿命长,内容丰富(可将文本、图像、音频等存储在一起)。其缺点主要有:信息获取比国际联机慢(回溯检索需多次换盘),信息更新不及时。

(4) 互联网检索阶段(1991 年至今)

自 20 世纪 90 年代起,随着互联网技术的飞速发展,基于互联网技术的信息检索系统开始出现,特别是世界各国的计算机网络普遍采用 TCP/IP 通信协议,使得不同主机之间的信息交换成为可能。同时许多检索系统相继进入网络,各检索系统的主机成了网络上的节点,而每个节点又连接着许多终端,构成了一个庞大的网络化检索系统。

互联网检索是指利用公用信息网或卫星微波系统,借助相关专用设备,将终端设备与计算机网络相连,用户可不受时空限制访问远程信息资源。网络上的用户既是网络资源的索取者,同时也是网络资源的提供者,真正实现了信息资源的共享。目前,用户不仅可以利用自己的终端访问世界上任何已接入互联网的数据库,还能够检索网上数字图书馆的信息资源,信息检索的广度和深度有了极大的飞跃,互联网使计算机信息检索进入全新的发展阶段。

与手工检索相比,计算机检索的优点主要表现在:① 检索速度快,仅几秒或十几秒即可从成千上万条,甚至上亿条记录中找出所需信息;② 检索范围广,可以更加全面地浏览任何学科的信息,可以同时检索到各学科的各类型信息;③ 检索途径多样化,除了手工检索阶段的目录、题录、文摘等途径,还可通过题名、作者、关键词、主题词、机构、基金、摘要等多种途径进行检索;④ 交互性好,拥有友好的交互页面,可以随时随地实现互动和反馈;⑤ 更新速度快,通信技术、高密度存储技术和计算机技术的高度融合,实现了实时更新。但其同样存在着较为明显的弊端:① 检索准确率低,信息垃圾较多;② 回溯性文献少,尤其是对于较早的文献,不能及时检索到;③ 用户需要掌握一定的信息检索方面的知识,计算机检索尤其是互联网检索,不仅要求用户具备一定的信息检索相关理论知识,还要求熟练掌握检索策略,这些都会增加用户的负担;④ 投资相对较高,需要用户拥有一定的检索设备,如网络、个人电脑或者公共电脑等,会额外增加用户的经济压力。

4.2.3 信息检索的原理

从信息存储与信息检索过程可以看出,信息检索就是将检索需求与存储标识进行匹配,选出符合要求的文献信息。若想获取最佳的信息,首先需要对相关文献信息进行收集、整理、组织、标引等处理,通过相关规则和策略将之进行集中存储,使之从无序到有序、从零散到集中,从而形成检索工具或检索系统(通常也被称为数据库);其次需要对用户提出的文献信息需求进行内容分析,提取主题概念和属性;最后利用与文献信息集合相同的标识系统来表达用户信息需求中的主题概念和属性,将后者与检索工具或检索系统中的信息进行匹配,若匹配一致,则检索工具或检索系统返回相关信息,即检索结果。

如图 4.4 中所示,信息检索的基本原理是:将信息存储中的存储标识与信息检索中的用户提问标识进行相似度计算,相似程度最高的结果最有可能符合用户的信息需求。为保证用户检索需求与文献存储标识两者之间尽可能相关,以提高匹配成功率,要求文献信息存储与文献信息检索遵循相同的原则和方法,检索语言和名称规范也须尽可能相同,只有这样用户才能获得符合需求的检索结果。

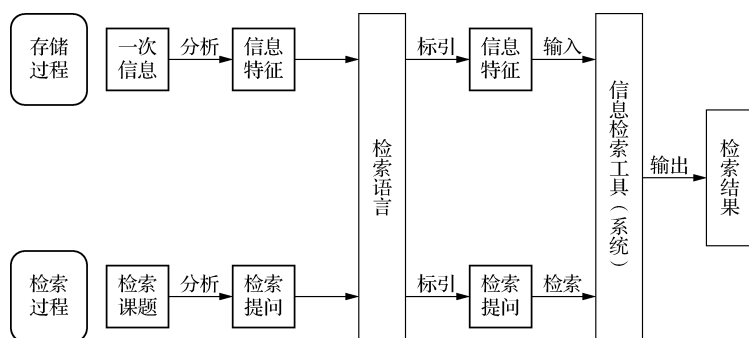


图 4.4 信息检索原理示意图

4.3 检索语言

4.3.1 检索语言及其作用

1. 检索语言的概念

信息检索语言是信息存储与检索过程中用于描述信息特征和表达用户信息提问的一种专门语言,是将标引语言和检索用语进行相符性比较的人工语言。这种人工语言就是从自然语言中精选出来并加以规范化的一套词汇符号,是概括文献信息内容或外在特征及其相互关系的概念标识体系,是沟通文献信息存储和检索两个过程中信息标引人员与信息检索人员双方思路的桥梁,是编制检索工具的各种索引的依据。

当信息存储时,标引人员将搜集到的信息按其外表特征和内容特征用一定的语言加以描述,并赋予一定的标志,如题名、著者、关键词等。将其整理、加工、存储于检索系统中用户进行信息检索时,首先要对检索课题进行分析,用同样的语言抽取几个能代表检索课题要求的检索标志,通过与检索系统中存储的标志相匹配,获取所需信息。这种在信息检索中用来联系文献信息和用户需求的“语言”就是信息检索语言。所以,信息检索语言是适应信息检索的需要,并为信息检索特设的专门语言。

2. 检索语言的作用

用户要检索到适当的信息,关键是要让自己使用的提问标识与检索系统使用的检索标识一致。二者达到统一的关键在于使用相同的检索语言,因此检索语言在信息检索中起着至关重要的作用。检索语言是沟通信息存储与信息检索的桥梁。在信息存储过程中,用来描述信息的内容和外部特征,从而形成检索标识。在信息检索过程中,用来描述检索提问,从而形成提问标识。当提问标识与检索标识完全匹配或部分匹配时,结果即为命中文献。

检索语言的主要作用如下:① 标引信息内容及其外表特征,保证不同标引人员表征信息内容的一致性;② 对内容相同及相关的信息加以集中或揭示其相关性;③ 使信息的存储集中化、系统化、组织化,便于检索者按照一定的排列次序进行有序化检索;④ 便于将标引用语和检索用语进行相符性比较,保证不同检索人员表述相同信息内容的一致性,以及检索人员与标引人员对相同信息内容表述的一致性;⑤ 保证检索者按不同需要检索信息时,都能获得最高查全率和查准率。

4.3.2 检索语言类型

目前,世界上的信息检索语言有几千种,依其划分方法的不同,其类型也不一样。常用的划分方式是按照表述文献特征、结构原理、组配方式。

1. 按照表述文献特征进行划分

按照表述文献特征进行划分,信息检索语言可划分为表达信息外部特征的语言和表达信息内容特征的语言,如图 4.5 所示。

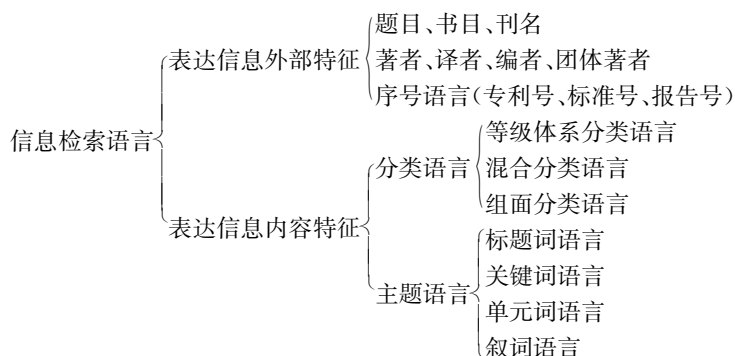


图 4.5 按照表述文献特征进行划分

(1) 表达信息外部特征的语言

表达信息外部特征的检索语言主要是指文献的篇名(题目、书名、刊名等),著者语言(著者、译者、编者、团体著者等),序号语言(专利号、标准号、报告号等)。

(2) 表达信息内容特征的语言

文献内容特征主要指文献正文所论述的主题、观点、简介、结论等。主要是分类语言(等级体系分类语言、混合分类语言、组面分类语言等),主题语言(标题词语言、关键词语言、单元词语言、叙词语言等)。这类检索语言在标引中提供分类语言和主题语言的检索途径。

2. 根据检索语言的结构进行划分

根据检索语言的结构划分,信息检索语言可以划分为分类语言、主题语言、名称语言、代号语言和引文语言,如图 4.6 所示。

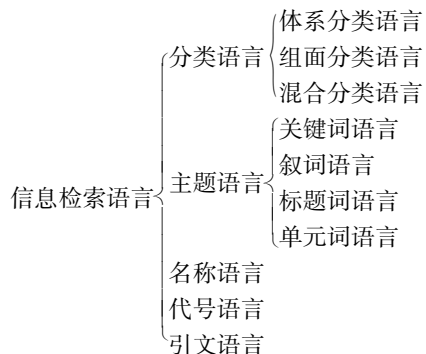


图 4.6 根据检索语言的结构划分

(1) 分类语言

分类语言是将表示各种知识领域(学科及其研究问题)的类目按知识分类原理进行系统排列并以代表类目的数字、字母符号(分类号)作为信息主题标识的一类信息检索语言。分类语言的主要特点是按学科、专业集中信息,并从知识分类角度揭示各类信息在内容上的区别和联系,提供从知识分类检索信息的途径。它犹如一张知识地图,能够使检索者沿着大大小小知识领域的隶属并列关系,找到所要到达的目的地——记载着所需要的知识的文献信息。

分类语言是信息工作中运用特别广泛的信息组织方法,最近几年在网络信息资源组织中的作用也逐渐得以体现,其中体系分类语言是最常用的。它的最大优点是具有按学科或专业集中地、系统地揭示信息内容的功能,可使检索者鸟瞰一个学科或专业信息的全貌,并可触类旁通。这对于系统地掌握和利用一个学科或专业范围的信息是很方便和有效的。由于人们一般都是在某个专业范围内从事科研、生产、教学、管理等活动的,比较习惯于从学科、专业出发获取信息,所以它的这一优点特别重要。

目前,国内外最常用的分类语言包括《中国图书馆图书分类法》《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。

分类语言是用分类号和相应的分类款目名称来表达信息主题概念,并将信息按学科性质分门别类地系统组织起来的一种检索语言。分类语言能反映事物的从属派生关系,便于按学科门类进行检索。它又分为体系分类语言、组面分类语言和混合分类语言三种。

① 体系分类语言

体系分类语言是一种直接体现分类等级概念的标志系统。它以科学分类为基础,以信息内容的学科性质为对象,运用概念的划分与概括的方法,按照知识门类的逻辑次序,从上到下、从总到分,进行层次划分,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。这些类目层层隶属,形成一个严格有序的等级结构体系。如《杜威十进制分类法》《中国图书馆分类法》(简称《中图法》),使用的就是典型的体系分类语言。分类表则是这种语言的具体体现。体系分类语言的主要特点是:按学科、专业集中信息并从知识分类角度揭示各类信息在内容方面的区别和联系,提供从学科分类检索信息的途径。

② 组面分类语言

组面分类语言是用科技术语进行组配的方式来描述信息内容。这些科技术语按其学科性质分为若干组,称为“组面”。组面内各个术语都附有相应的号码。标引信息时,根据信息内容选择相应的组面和有关术语,把这些术语的号码组配起来,构成表达这一信息内容的分类号。例如,印度阮冈纳赞的《冒号分类法》,对“牙医外科”的分类号为 L124:4:7,其中字母 L 代表医学,数字 124 表示牙齿,数字 4 表示疾病,数字 7 表示外科,这些字母和数字通过冒号组配就形成一个分类号。

组面分类语言是体系分类语言的发展,组面分类的分类标志是散组式的、组合的、可以分拆的,其中诸因素是可以变换位置的,这样给分类语言带来了很大的灵活性,克服了体系分类标引能力差的弱点,以及“集中和分散”的基本矛盾,在提高检索效率上前进了一大步。

③ 混合分类语言

它是组面分类和体系分类语言的结合,两者有所侧重,因而又有组面体系分类语言和体系组面分类语言之分。《中图法》是我国第一部集中了全国图书馆和信息部门的力量共同编

制的一部综合性大型文献分类法。目前广泛应用于各类型图书馆。《中图法》主要是从科学分类和知识分类的角度来揭示文献内容的区别与联系,按学科和专业集中文献,提供从学科和专业出发检索文献的途径。

《中图法》主要由类目表、复分表和索引三部分组成。

A. 类目表。类目表是《中图法》最主要的组成部分,它由基本部类、基本大类、简表及详表构成。

a. 基本部类和基本大类。《中图法》的分类体系是指基本部类与基本大类的构成及其序列,以及所有类目相互联系与相互制约形成的等级结构。基本部类与基本大类的构成及其序列是《中图法》最基本的分类体系。

《中图法》的五大基本部类主要为:马克思主义、列宁主义、毛泽东思想、邓小平理论;哲学、宗教;社会科学;自然科学;综合性图书。这是整个分类表最先确定、最本质、最概括的区分。

《中图法》在基本部类的基础上,根据当前学科状况区分形成了一组具有独立体系的纲领性类目,即一级类目,这是在基本部类基础上展开的知识分类体系框架,是传统的、稳定的、较为概括的学科或知识领域。《中图法》在五大部类的基础上设置了22个基本大类的分类框架。

b. 简表。简表是在基本大类的基础上进一步区分出的类目。它是整个分类表的骨架,起着承上启下的作用。简表一般包括三级类目,可以反映出整个分类表的概貌。

c. 详表。详表是在简表的基础上扩展而成的。这是分类表最主要、最本质的组成部分,是分类表的主体,也是文献分类标引和分类检索的主要依据。

《中图法》的详表是由分类号和类目组成的集合。类目是类的名称,是具有共同属性的组概念。分类号是类目的标志符号,其主要作用就是简明、系统地表示每个类目在分类体系中的位置,以便组织分类目录和图书排架。类目和分类号一一对应,相辅相成。

《中图法》的编号制度采用基本的层累制。层累制是根据类目的不同等级,配以相应不同位数号码的编号方法,类目的等级与其号码位数是相对应的。层累制的号码可以无限纵深展开,可充分满足类目体系层层展开配号的需要,同时又有良好的表达性。

《中图法》的标记符号采用汉语拼音与阿拉伯数字相结合的混合号码。其中字母表示大类,数字表示其大类下进一步细分的小类(下位类)。一般情况下,数字位数的多少代表其类目划分的级数。

B. 复分表。复分表又称为辅助表、附表,是由共同性的子目构成单独编制,供有关类目进一步区分时共同使用的表。《中图法》的复分表分为总论复分表、世界地区复分表、中国地区表、国际时代表、中国时代表、世界种族与民族表、中国民族表、通用时间地点表。

C. 索引表。《(中国图书馆分类法)索引》单独出版发行。它收录了分类表中已列出的具有检索意义的概念,并将这些概念依汉语拼音字顺排列起来。它的编制目的主要在于提供从字顺查检分类表类目的途径,其次是使分类表具有一定程序的主题检索功能。

(2) 主题语言

主题语言又称主题词语言,是一种描述性语言,是用自然语言中的词、词语来描述信息内容特征,即信息所论述或研究的事物概念。换言之,不论学科分类如何,主题语言直接借助于自然语言的形式,作为信息内容的标志和检索依据,是一种以主题词字序为基本结构

的检索语言,比较直观。

主题语言就是以自然语言的语词经过规范处理(采取人工控制措施)后直接作为信息内容主题标识。主题标识按字顺排列,并用参照系统和其他方法来间接地显示概念之间的关系。主题语言的特点是按实物集中信息,用参照系统等方法间接显示概念或实物之间的关系,提供事物名称的字顺检索途径。检索者只要确知他所需要检索对象事物的名称,就可从主题检索系统的字顺中直接检索出该对象事物及其各方面的有关信息,并可通过参照系统等扩大检索范围。

一般来说,用主题语言组织与揭示信息具有直接和直观的特点,即主题语言用于组织信息不仅具有“直呼其名、依名查检”的直接性,而且其基本标识上是独立完整的事物概念。在网络环境中,主题语言检索系统得到发展与完善,尤其是关键词检索,在网络中的应用相当广泛,有相当一部分网络信息资源浏览器搜索引擎都以关键词为组织与揭示信息的重要途径和方法,如 Google、百度、Yahoo 等。

目前,国内外常用的主题语言有《美国国会图书馆标题表》《医学主题词表》《汉语主题词表》等。

主题语言可分为关键词语言、标题词语言、叙词语言、单元词语言四种。

① 关键词语言

关键词是直接从信息的标题、正文或摘要中抽取出来,未经过规范化处理,能够表达信息主题内容的关键性词汇。关键词语言是一种未经过规范化的自然语言,但具有表达信息概念直接、准确等特点,被广泛应用于手工检索和计算机检索。

关键词索引是以信息中的一些主要关键词作为检索标志,按字顺排列,并指出信息出处的一种索引。按其款目是否保留非关键词(冠词、介词、连词等)及不同的排检方法,分为单纯关键词索引、题内关键词索引和题外关键词索引。

关键词语言的主要特点是:标引完全专指,易于实现自动标引。其不足之处是:由于对词汇不经控制或少量控制,其检索质量较差。

② 标题词语言

标题词是从自然语言中选取并经过规范化处理的,表达事物概念的词、词组或短语。标题词语言是用经过规范化处理的名词术语来直接表达信息所论及的事物或主题,并将全部标题词按字顺排列起来而形成的一种检索语言。它是一种先组式的检索语言,也是最早使用的主题语言。

标题词表是根据标题词语言编制的,收录标题词及其规则的一部标题词典。它对标题词进行规范化处理和管理,通过参照系统显示词与词直接的逻辑关系,是标引和检索信息的依据。例如,美国工程信息公司编制的《工程主题词表》(简称 SHE)。

标题词语言的主要特点有:形式直观、含义明确、操作简便。其主要不足是:概念难以多向成族,无法从多个因素、多个途径检索,灵活性较差等。

③ 叙词语言

叙词语言是从自然语言中优选出来并经过规范化处理的名词术语。叙词语言是采用表示单元概念的规范化语词的组配,来对信息内容主题进行描述的后组式词汇型标志系统的检索语言,也是目前使用最广泛的主题语言。

叙词受词表控制,词表中词与词之间无从属关系,都是相互独立的概念单元。检索时可

根据需要选出相应的叙词,按照组配原则任意组配检索概念。叙词语言的主要特点是:组配准确,标引能力强;组配方式灵活,可实现多向成族、多途径、多因素检索,检索效果较好。其不足之处是:词表编制和管理难度大,对标引人员要求高,标引难度大。

《汉语主题词表》是由中国科学技术信息研究所、国家图书馆主编的,我国第一部全面反映自然科学和社会科学领域名词术语的大型综合性汉语叙词表。它是一种将自然语言转换为检索语言的叙词控制工具,是叙词语言的具体表现。

④ 单元词语言

单元词语言是以单元词作为文献内容标志和检索依据的一种主题语言。所谓单元词,是从文献正文、摘要或题目中抽取出来的最一般、最基本的,其概念不可再分的词。它一般未经过规范化,也无词表。检索时,根据检索课题的内容特征,选取恰当的单元词进行组配检索。例如,美国化工专利使用的《化学专利单元词索引》。

单元词语言的主要特点有:词表体积小,标引专指度高,概念可多向成族,可进行多因素和多途径组配检索,灵活性较大。其不足之处是:直接性较差,采用字面组配,在字面分解与语义分解不一致时,容易产生误差,概念显示不充分,难以进行相关检索。

(3) 名称语言

名称语言是以人名、地名、书名等代表信息特征的名称为检索标识。作为标引和检索的桥梁,不同的数据库中所设置的作者检索途径、机构检索途径等都是运用名称语言对信息的特征予以描述和展开的结果。

(4) 代号语言

代号语言是文献特有的顺序号(包括标准号、合同号等)。在检索时,代号语言的标引和检索都较为直接和有效,检索起来也非常准确。

(5) 引文语言

引文语言是利用文献之间引用与被引用之间的关系,来表达检索文献主题之间的相互关系,不需要标引文献,检索简单而有效,在数据库检索中用得很多。

3. 根据检索语言的组配方式划分

根据检索语言的组配方式,信息检索语言可以分为先组式信息检索语言和后组式信息检索语言,如图 4.7 所示。

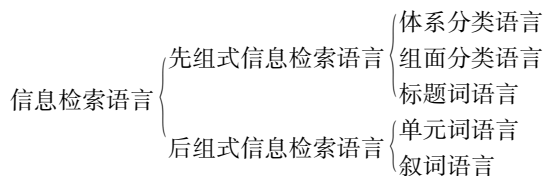


图 4.7 根据检索语言的组配方式划分

(1) 先组式信息检索语言

先组式检索语言是指在检索之前,表述文献主题概念的标识已经拟定好的检索语言。先组式语言又分为定组型和散组型两种。定组型是指表述文献主题概念的标识,在编制检索语言词表时已经预先固定组配;散组型是指表述文献主题概念的标识,在编制检索语言时并不预先固定组配,而是在标引文献时,根据文献信息的主题内容予以组配,检索阶段也通

过相同的组配与之匹配。主题检索语言中的标题词语言属于先组定组型检索语言,而叙词语言当作标题词语言使用时,则属于先组散组型检索语言。

(2) 后组式信息检索语言

后组式表达文献信息主题概念的标识,在编制检索语言词表和标引文献的时候,不预先固定组配,而是在检索进行时根据检索的实际需要,按照组配规则临时进行组配的检索语言。叙词语言属于此种类型。

4.3.3 医学信息检索常用的检索语言

国内生物医学文献数据库最常使用的检索语言工具有《中国图书馆图书分类法》、美国《医学主题词表》(MeSH)和《中国中医药学主题词表》。

1. 中国图书馆图书分类法

(1) 简介

中国图书馆分类法,简称中图法,是由北京图书馆、中国科学技术情报所等单位于 1971 年共同编制完成,1974 年出版,并经过多次修订与再版,1999 年 3 月第四次出版。

中图法是体系等级分类语言,它采用五分法,即把知识学科分成 5 个基本部类:“马克思主义、列宁主义、毛泽东思想、邓小平理论”“哲学、宗教”“社会科学”“自然科学”和“综合性图书”。在各基本部类下再展开,共形成 22 个大类,用 1 个英文大写字母表示一级类目名(工业技术类除外),并以字母的顺序反映大类的顺序。根据各学科的内容需要,大类下再依次分二级、三级、四级、五级子类目。各子类目用阿拉伯数字表示。

中图法是国内最有影响力,使用最广泛的文献信息组织工具。国内大多数图书馆、情报机构使用中图法。国内也有一些大型书目数据库使用中图法作为组织数据的工具,并提供分类检索途径,如《中国生物医学文摘数据库》(CBMdisc)、清华同方的《中国学术期刊全文数据库》、维普的《中国科技期刊数据库》等。目前图书情报界的学者们正致力于中图法在网络信息资源组织中的应用研究。

(2) 中图法的医药卫生分类

中图法用“R”代表“医药卫生”大类。该大类又分为 17 个基本类目(二级类目):

R1 预防医学、卫生学	R74 神经病学与精神病学
R2 中国医学	R75 皮肤病学与性病学
R3 基础医学	R76 耳鼻咽喉科学
R4 临床医学	R77 眼科学
R5 内科学	R78 口腔科学
R6 外科学	R79 外国民族医学
R71 妇产科学	R8 特种医学
R72 儿科学	R9 药学
R73 肿瘤学	

每一个二级类目,又分为若干三级、四级、五级类目,下面是“R2 中国医学”的类目划分:

(三级类)	(四级类)	(五级类)
R21 中医预防、卫生学	R241 中医诊断学	R246.1 内科
R22 中医基础理论	R242 中医治疗学	R246.2 外科针刺麻醉法
R24 中医临床学	R245 针灸学、针灸疗法	R246.3 妇产科
R25 中医内科	R246 针灸疗法与临床应用	R246.4 小儿科
R26 中医外科	R247 其他疗法	R246.5 肿瘤科
R271 中医妇产科	R248 中医护理学	R246.8 五官科
R272 中医儿科	R249 医案医话	R246.82 眼科
R273 中医肿瘤科		R246.83 口腔科
R274 中医骨伤科		R246.9 其他

2. 医学主题词表

医学主题词(Medical Subject Headings, MeSH)是美国国立医学图书馆(NLM)编制的医学领域内权威的主题语言,在医学信息检索系统中得到了广泛的应用,如世界上最大的生物医学数据库(MEDLINE)是由该词表索引和组织文献数据的;国内最大的中文生物医学文献数据库(CBMdisc)也用 MeSH 组织文献;美国 NLM 用 MeSH 组织馆藏;国内大多数医学图书馆和情报机构用 MeSH 对文献进行主题标引,组织主题目录;Medical Matrix 等网络医学搜索引擎用 MeSH 组织网页信息。

MeSH 表由主题词字顺表、树状结构表及副主题同表组成。

(1) 主题词字顺表

主题词字顺(permuted index)表是 MeSH 表的主体,主题词和入口词按字顺排列。

MeSH 对词义经过严格的规范化处理,主要体现在“一一对应”方面,即一个概念只能用一个语词来表达,一个语词只能表达一个概念。若一个概念存在同义词或近义词,则只选其中一个词作为主题词,所有的文献信息都归在这个主题词下面。这样,可以避免文献信息的分散,便于用户查全信息。如“肝肿瘤”在英文中可以有 liver cancer、liver tumor、liver carcinoma、liver neoplasms 等多种拼写方法,MeSH 选用 liver neoplasms 作主题词,所有的相关文献信息都归在这个词下面,其他的词只作为该词的入口词,通过入口词可以查到主题词。

MeSH 用参照系统揭示按字顺排列的主题词之间的同义、相关和属分等语义关系。如 liver cancer see liver neoplasms,说明这两个词是同义关系。

MeSH 是动态变化的,每年出版修订版,及时对内容进行修订、增补和调整,增加新术语,淘汰旧术语,以满足医学科学技术的发展和进步的需要。2004 年版的 MeSH 包括 22568 个主题词。

(2) 树状结构表

树状结构(tree structure)表也叫分类表,是将字顺表中的主题词按学科的隶属关系分成 15 个大类,每大类之下又继续细分,像树的分枝一样,最多可分至 9 级。MeSH 用这种等级结构来组织主题词,通过查找范畴大的上级词可以查到用下级词索引的范畴窄的文献,这种结构也使得检索者可以有效地浏览词表,从而找到合适的主题词。下列就是有关 gastrointestinal neoplasms 的树状结构表:

gastrointestinal neoplasms
 intestinal neoplasms
 colorectal neoplasms
 colonic neoplasms
 adenomatous polyposis coli
 gardner syndrome

(3) 副主题词表

MeSH 副主题词(subheadings)表中共收录了 82 个副主题词,如诊断、治疗、外科学、康复、护理等。副主题词是对主题词的进一步限制,用来检索更为专指的文献信息。如用主题词“脑肿瘤”和副主题词“外科学”,则检索出脑肿瘤外科治疗方面的文献,而不是脑肿瘤其他方面的文献。每个副主题词可以根据其特定含义和使用范围与不同的主题词组配。正确使用 MeSH 的副主题词有助于检索者提高查准率。

3. 一体化医学语言系统

一体化医学语言系统(Unified Medical Language System, UMLS)是检索语言的集成系统,其目的是为了解决因不同检索系统的差异性和信息资源的分散性所造成的检索困难。UMLS 由美国国立医学图书馆(NLM)于 1986 年构建,它整合了各来源词表中的生物医学概念、术语、词汇及其等级范畴。UMLS 包括 4 个部分:超级叙词表(metathesaurus)、语义网络(semantic network)、信息源图谱(information sources map)和专家词典(specialist lexicon),这 4 个部分是一个有机的整体。

超级叙词表是 UMLS 的词汇部分,2001 年版收录了 80 万个概念,共 190 万个词汇。这些概念和词汇来自 60 多个生物医学受控词表、术语表、分类表、专家系统等,其中 MeSH 表是最主要的词汇来源。语义网络是为建立概念、术语间错综复杂的关系而设计的,它为超级叙词表中所有概念提供了语义类型、语义关系和语义结构。语义网络把词汇联结成一个整体,揭示了词汇间的关系。信息源图谱是一个关于生物医学机读信息资源的数据库,用于测度信息源与特定提问的相关性,以便选取最合适的住处源。专家词典是一个包含众多生物医学词汇的英语词典。

UMLS 在计算机数据库和网络资源的组织及检索中具有广泛的应用前景。它不仅可以克服不同系统检索语言的差异性,还可以实现跨数据库检索的词汇转换,帮助用户对计算机化的病案系统、书目数据库、事实数据库、图像数据库、专家系统等各种联机信息源中的生物医学信息进行一体化检索,以获取符合用户需求的特定或综合的信息。

著名的 PubMed 数据库已使用 UMLS,现已集成了 MEDLINE、PreMEDLINE、GenBank、EMBL、SWISS-PORT、MMDB、OMIM 等 10 多个数据库,用户可以获取其中的生物医学文献信息、DNA 序列信息、蛋白质序列信息、三维结构信息以及基因组信息。PubMed 可以将用户的自然语言转换成 UMLS 中的医学主题词,并自动对数据库进行检索。此外,它还具有自动查找相关文献的功能。PubMed 的数据库的集成功能与智能检索得益于 UMLS 的成功运用。

一些医学专业搜索引擎开始用 UMLS 组织网络信息资源,提高了检索的效率,且使引擎趋向智能化。目前使用 UMLS 的医学搜索引擎包括 Medical World Search、Clini Web

International 等。

4.4 信息检索技术

4.4.1 信息检索技术的概念

信息检索技术是指利用检索系统,检索有关信息而采用的一系列技术的总称,主要包括布尔逻辑检索、截词检索、限制检索、位置检索等技术。

1. 布尔逻辑检索

布尔逻辑得名于乔治·布尔(George Boole),一位科克大学的英国数学家,在19世纪中叶首次定义了逻辑的代数系统。现在,布尔逻辑在电子学、计算机硬件和软件中有很多应用。

在实际检索中,检索提问涉及的概念往往不止一个,而同一个概念又往往涉及多个同义词或相关词。为了正确地表达检索提问,系统中采用布尔逻辑运算符将不同的检索词组配起来,使一些具有简单概念的检索单元通过组配成为一个复杂概念的检索式,用以表达用户的信息检索要求。

所谓布尔逻辑检索,是用布尔逻辑算符将检索词、短语或代码进行逻辑组配,指定信息的命中条件和组配次序,凡符合逻辑组配所规定条件的为命中信息,否则为非命中信息。它是计算机检索系统中最常用的一种检索方法,逻辑算符主要有:与(AND)、或(OR)、非(NOT)。

(1) 逻辑“与”

逻辑“与”也称逻辑乘,用关系词“AND”或“*”表示, $A \text{ AND } B$ (或 $A * B$)表示两个概念的交叉和限定关系,只有同时含有这两个概念的记录才算命中信息。

检索结果如图4.8所示,图中阴影部分即为同时包含A和B两个概念的命中信息。如,查找“胰岛素治疗糖尿病”的检索式为“insulin(胰岛素)and diabetes(糖尿病)”。

(2) 逻辑“或”

逻辑“或”也称逻辑和,用关系词“OR”或“+”表示, $A \text{ OR } B$ (或 $A + B$)表示两个概念的并列关系,记录中只要含有任何一个概念就算命中信息,即凡单独含有概念A或含有概念B或者同时含有A、B两个概念的信息均为命中信息。

检索结果如图4.9所示,图中阴影部分即为包含A或B概念的命中信息。如,查找“肿瘤”的检索式为“cancer(癌)OR tumor(瘤)OR carcinoma(癌)”。

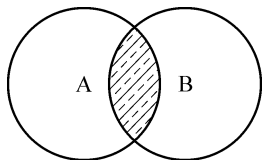


图 4.8 逻辑“与”示意图

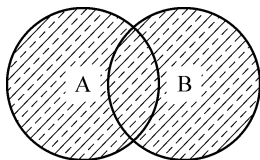


图 4.9 逻辑“或”示意图

(3) 逻辑“非”

逻辑“非”也称逻辑差,用关系词“NOT”或“-”表示。 $A \text{ NOT } B$ (或 $A - B$)表示两个概

念的排除关系,指记录中含有概念 A 而不含有概念 B 的为命中信息。

检索结果如图 4.10 所示,图中阴影部分即为包含 A 且排除 B 的命中信息。如,查找“动物的乙肝病毒(不要人的)”的文献的检索式为“hepatitis B virus(乙肝病毒)not human(人类)”。

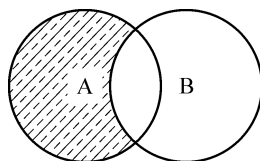


图 4.10 逻辑“非”示意图

对于一个复杂的逻辑检索式,检索系统的处理是从左向右进行的。在一个检索式中,可以同时使用多个逻辑运算符,构成一个复合逻辑检索式。复合逻辑检索式中,运算优先级别从高至低依次是非、与、或,可以使用括号改变运算顺序。

2. 截词检索

截词检索(truncation)是指用给定的词干做检索词,查找含有该词干的全部检索词的记录,也称词干检索或字符屏蔽检索。

截词检索是指在检索式中用专门的截词符号表示检索词的某一部分允许有一定的词形变化。截词符一般用“?”或“*”表示,不同的系统、数据库,其代表的含义有所不同。如美国 DIALOG 系统用“?”表示截词符。

截词的方式有多种,按截断部位可分为右截断、左截断、中间截断、前后截断等;按截断长度可以分为有限截断和无限截断。

(1) 前截断

前截断也称左截断,截去某个词的前部,是词的后方一致比较,也称后方一致检索。如由“? computer”可检索出含有“computer”“minicomputer”“microcomputer”等的信息记录。

(2) 后截断

后截断也称右截断,截去某个词的后部,是词的前方一致比较,也称前方一致检索。如由“computer?”可检索出含有“computer”“computers”“computerize”“computerized”“computerization”等的信息记录。

(3) 中间截断

中间截断也称屏蔽词、两边一致检索。一般来说,中间截断仅允许有限截词,主要用于英、美拼写不同的词和单复数拼写不同的词,如 organi?ation 可检索出含有 organisation 和 organization 的记录。由此可知,中间截断使用的符号为“?”,即用“?”代替那个不同拼写的字符。

(4) 前后截断

词干的前后各有一个截词符,截去某个词的前部和后部,也称任意匹配检索。如由“? computer?”可检索出“computer”“computers”“computerize”“computerized”“computerization”“minicomputer”“microcomputer”等的信息记录。

由上述可见,任何一种截词检索,都隐含着布尔逻辑检索的“或”运算。采用截词检索时,既要灵活又要谨慎,截词的部位要适当,它可以起到扩大检索范围、提高查全率、减少检索词的输入量、节省检索时间、降低检索费用等作用。如果截得太短(输入的字符不得少于 3 个),将增加检索噪声,影响查准率。

截词检索最早出现在西文检索中,现在中文检索也大量使用。在中文检索中,我们把一

句、一段甚至全文当成一个“词”，选择检索模式有前方一致、完全匹配、任意匹配三种。前方一致是后截断检索，任意匹配是前后任意匹配。

在多数检索系统中，检索框中输入的检索词，如没有明确选择，大多默认为任意匹配。

3. 限制检索

限制检索(range)是指限制检索词在数据库记录中规定的字段范围内出现的信息，为命中信息的一种检索技术。限制检索适用于在已有一定数量输出记录的基础上，通过指定字段或使用限制符，减少输出信息数，达到优化检索结果的作用。限制检索的方式有多种，例如进行字段检索、使用限制符、采用限制检索命令等。

(1) 字段检索

字段检索是把检索词限定在某个/些字段中，如果记录的相应字段中含有输入的检索词则为命中记录，否则检不中。在检索系统中，数据库设置的可供检索的字段通常有两种：表达文献主题内容特征的基本字段和表达文献外部特征的辅助字段。如在南京医科大学康达学院图书馆书目检索系统中，选择检索类型，有题名、责任者、主题词、ISBN、订购号、分类号、索书号、出版社、丛书名等字段。

(2) 使用限制符

用表示语种、文献类型、出版国家、出版年代等的字段标识符来限制检索范围。

(3) 使用范围符号

如 less than、greater than、from to 等，如查找 1989~1999 年的文献，可表示为“PY=1989:1999”或者“PY=1989 to PY=1999”。

(4) 使用限制指令

限制指令可以分为一般限制指令(limit，它是对事先生成的检索集合进行限制)和全限制指令(limit all，它是在输入检索式之前向系统发出的，把检索的全过程限制在某些指定的字段内)。

上述几种限制检索方法既可独立使用，也可以混合使用。不同数据库中所包含的字段数目不尽相同，字段名称也不一定相同，常见的检索字段有：

题名	title	TI
文摘	abstract	AB
责任者	author	AU
责任者单位	corporate source	CS
地址	address	AD
刊名	journal	JN
叙词	descriptor	DE
语种	language	LA
主题词	subject	SU
文献类型	document type	DT

4. 位置检索

位置检索也叫邻近检索。记录中词语的相对次序或位置不同,所表达的意思可能不同,而同样一个检索表达式中词语的相对次序不同,其表达的检索意图也不一样。布尔逻辑运算符有时难以表达某些检索课题确切的提问要求。限制检索虽能使检索结果在一定程度上进一步满足提问要求,但无法对检索词之间的相对位置进行限制。

位置检索是在检索词之间使用位置算符(也称邻近算符)来规定算符两边的检索词出现在记录中的位置,从而不但获得包含有指定检索词,而且这些词在记录中的位置也符合特定要求的记录,能够提高检索的准确性,相当于词组检索。位置算符检索是用一些特定的算符(位置算符)来表达检索词与检索词之间的邻近关系,并且可以不依赖主题词表而直接使用自由词进行检索的技术方法。

按照两个检索词出现的顺序、距离,可以有多种位置算符。而且对同一位置算符,检索系统不同,规定的位置算符也不同。以美国 DIALOG 检索系统使用的位置算符为例,介绍如下:

(1) (W)算符

(W)是“with”的缩写,还可以简写为()。这个算符表示其两侧的检索词必须紧密相连,除空格和标点符号外,不得插入其他词或字母,两词的词序不可以颠倒。

例如,检索式为“communication(w)satellite”时,系统只检索含有“communication satellite”词组的记录。

(2) (nW)算符

(nW)是“nWords”的缩写,表示此算符两侧的检索词之间允许插入最多 n 个词,顺序不可颠倒。

例如,检索式为“laser(1w)print”可检索出包含“laser printer”“laser color printer”和“laser and printer”的记录。如“socialist(1w)economy”可同时查出含有“socialist commodity economy”“socialist planned economy”“socialist national economy”的文献。

(3) (N)算符

(N)是“near”的缩写,这个算符表示其两侧的检索词必须紧密相连,除空格和标点符号外,不得插入其他词或字母,但两词的词序可以颠倒。

例如,“computer(n)network”可检索出含有“computer network”“network computer”的记录。

(4) (nN)算符

(nN)表示允许两词之间插入最多为 n 个其他词,包括实词和系统禁用词,且两词的词序可以颠倒。

例如,检索式为“computer(2n)system”可检索出含有“computer system”“computer code system”“computer aided design system”“system using modern computer”等形式的记录。

(5) (F)算符

(F)是“field”的缩写。这个算符表示其两侧的检索词必须在同一字段(例如同在题目字段或文摘字段)中出现,而它们在该字段中的相对次序和相对位置的距离不限。

例如,“water pollution(f)control”表示在同一个字段中(如篇名、文摘、叙词等)同时含

有“water pollution”和“control”的记录均可检索出来。

(6) (S)算符

(S)是“sub-field”的缩写,表示在此运算符两侧的检索词只要出现在记录的同一个子字段内(例如,在文摘中的一个句子就是一个子字段),此信息即被命中。要求被连接的检索词必须同时出现在记录的同一句子(同一子字段)中,不限制它们在此子字段中的相对次序,中间插入词的数量也不限。

例如,“high(w)strength(s)steel”表示只要在同一句子中检索出含有“high strength”和“steel”形式的均为命中记录。在检索过程中,可利用检索字段进行后缀或前缀限制。

在检索语句或检索词后加斜线(/),再加后缀代码,或者前缀后加等号(=)来限定查找范围。

例如,查找有关“彩色电视”方面的文献。

? S(colour OR color)(W)(television OR TV)/TI, DE

上式表示只在篇名和叙词字段中查找,缩小查找范围。

例如,查找意大利 Firenze 市的名叫 STABILIMENTO 公司的简况。

? S CO=STABILIMENTO AND CN=Italy AND CY=Firenze

上式利用前缀和后缀代码进行限制。

5. 加权检索

加权检索是某些检索系统中提供的一种定量检索技术。加权检索同布尔逻辑检索、截词检索等技术一样,也是信息检索的一个基本检索技术,但与它们不同的是,加权检索的侧重点不在于判定检索词或字符串是不是在数据库中存在,以及与别的检索词或字符串是什么关系,而是在于判定检索词或字符串在满足检索逻辑后对文献命中与否的影响程度。加权检索的基本方法是:在每个提问词后面给定一个数值表示其重要程度,这个数值称为权,在检索时,先查找这些检索词在数据库记录中是否存在,然后计算存在的检索词的权值总和。权值之和达到或超过预先给定的阈值,该记录即为命中记录。阈值可视命中记录的多寡灵活地进行调整,阈值越高,命中记录越少。

运用加权检索可以命中核心概念文献,因此它是一种缩小检索范围、提高检准率的有效方法。但并不是所有系统都能提供加权检索技术,而能提供加权检索的系统,对权的定义、加权方式、权值计算和检索结果的判定等方面,又有不同的技术规范。

6. 聚类检索

聚类是把没有分类的事物,在不知道应分几类的情况下,根据事物彼此不同的内在属性,将属性相似的信息划分到同一类下面。聚类检索是在对文献进行自动标引的基础上,构造文献的形式化表示——文献向量,然后通过一定的聚类方法,计算出文献与文献之间的相似度,并把相似度较高的文献集中在一起,形成一个个文献类的检索技术。根据不同的聚类水平的要求,可以形成不同聚类层次的类目体系。在这样的类目体系中,主题相近、内容相关的文献便聚在一起,而相异的则被区分开。

聚类检索的出现,为文献检索尤其是计算机化的信息检索开辟了一个新的天地。文献自动聚类检索系统能够兼有主题检索系统和分类检索系统的优点,同时具备族性检索和特性检索的功能。因此,这种检索方式将有可能在未来的信息检索中大有用武之地。

4.4.2 信息检索技术的热点

目前,信息检索已经发展到网络化和智能化的阶段,信息检索的对象从相对封闭、稳定,由独立数据库集中管理的信息内容扩展到开放、动态、更新快、分布广泛、管理松散的 Web 内容。信息检索的用户也由原来的情报专业人员扩展到包括商务人员、管理人员、教师、学生、各专业人士等在内的普通大众,他们对信息检索的方法、技术及结果提出了更高、更多样化的要求。适应网络化、智能化和个性化的需要是当前信息检索技术的研究热点。

1. 智能检索

传统的全文检索技术基于关键词匹配进行检索,往往存在查不全、查不准、检索质量不高的现象,特别是在网络信息时代,利用关键词匹配很难满足人们对检索的要求。

智能检索利用分词词典、同义词典、同音词典改善检索效果,比如用户查询“计算机”时,与“电脑”相关的信息也能检索出来;进一步还可在知识层面或者概念层面辅助查询,通过主题词典、上下位词典、相关同级词典,形成一个知识体系或概念网络,给予用户智能知识提示,最终帮助用户获得最佳的检索效果。比如用户可以进一步缩小查询范围至“微机”“服务器”,或扩大查询至“信息技术”,或相关的“电子技术”“软件”“计算机应用”等范畴。另外,智能检索还包括歧义信息和检索处理,如“苹果”,究竟是指水果还是电脑品牌,“华人”与“中华人民共和国”的区分,将通过歧义知识描述库、全文索引、用户检索上下文分析以及用户相关性反馈等技术结合处理,高效、准确地反馈给用户最需要的信息。

智能检索也称为知识检索,是在现有的信息检索技术以及模型上发展而来的。智能检索和信息检索的不同,就在于智能检索强调了语义,不会与信息检索一样,只是基于字面的机械匹配,它从文章的语义、概念出发,能够揭示文章的内在含义。做好了语义和概念层次上的标引工作,智能检索就提高了查全率和查准率,减轻了用户的负担。

智能检索技术吸取多个学科的研究成果,通过对文本、图像和视频信息的智能处理,实现信息的精确检索。

2. 知识挖掘

知识挖掘源于全球范围内数据库中存储的数据量急剧增加,人们的需求已经不只是简单的查询和维护,而是希望能够对这些数据进行较高层次的处理和分析,以得到关于数据总体的特征和对发展趋势的预测。知识挖掘最新的描述性定义是从数据集中识别出有效的、新颖的、潜在有用的,以及最终可理解的模式的非平凡过程。

知识挖掘目前主要指文本挖掘技术的发展,目的是帮助人们更好地发现、组织、揭示信息,提取知识,满足信息检索的高层次需要。知识挖掘包括摘要、自动分类(聚类)和相似性检索等方面。

自动摘要就是利用计算机自动地从原始信息中提取文摘。在信息检索中,自动摘要有助于用户快速评价检索结果的相关程度;在信息服务中,自动摘要有助于多种形式的内容分发,如发往电子信箱、手机等。

相似性检索技术基于文档内容特征检索与其相似或相关的文档,是实现用户个性化相关反馈的基础,也可用于去重分析。

自动分类可基于统计或规则,经过机器学习形成预定义分类树,再根据文档的内容特征

将之归类;自动聚类则是根据文档内容的相关程度进行分组归并。自动分类(聚类)在信息组织、导航方面非常有用。

3. 异构信息整合检索和全息检索

在信息检索分布化和网络化的趋势下,信息检索系统的开放性和集成性要求越来越高,需要能够检索与整合不同来源和结构的信息,这是异构信息检索技术发展的基点,包括支持各种格式化文件,如 TEXT、HTML、XML、RTF、MS Office、PDF、PS2/PS、MARC、ISO2709 等的处理和检索;支持多语种信息的检索;支持结构化数据、半结构化数据及非结构化数据的统一处理;与关系数据库检索的无缝集成以及其他开放检索接口的集成等。

所谓“全息检索”,就是支持一切格式和方式的检索。从目前的实践来讲,发展到异构信息整合检索的层面,基于自然语言理解的人机交互以及多媒体信息检索整合等方面尚有待取得进一步突破。

另外,从工程实践角度,综合采用内存和外存的多级缓存、分布式群集和负载均衡技术也是信息检索技术发展的重要方面。

随着互联网的普及和电子商务的发展,企业和个人可获取、需处理的信息量呈爆发式增长,而且其中绝大部分都是非结构化和半结构化数据。内容管理的重要性日益凸显,而信息检索作为内容管理的核心支撑技术,随着内容管理的发展和普及,也将应用到各个领域,成为人们日常工作生活中的密切伙伴。

4.5 信息检索的方法、途径和步骤

4.5.1 常用的信息检索方法

检索信息究竟需要采用什么方法,我们必须根据课题性质和研究目的而定,本节主要从文献的角度宏观地谈信息检索的一般方法,具体的各类型信息资源及网络信息资源的检索方式参考后续章节。

1. 常规法

常规法也叫检索工具法,是利用各种检索工具查找文献的方法,即以主题、分类、著者等途径,通过检索工具获取所需文献的一种方法,这种方法又可分为顺查法、倒查法、抽查法和引文法 4 种。

(1) 顺查法

即由远及近的查找法。如果已知某创作或理论研究成果最初产生的年代,现在需要了解它的全面发展情况,可从最初年代开始,按时间的先后顺序,一年年地往近期查找。用这种方法所查得的文献较为系统全面,基本上可反映某学科专业或某课题发展的全貌,能达到一定查全率。在较长的检索过程中,可不断完善检索策略,得到较高的查准率。此法的缺点是费时费力,工作量较大。一般在新开课题时采用这种方法。

(2) 倒查法

即由近及远,由新到旧的查找法。此法多用于查找新课题或有新内容的老课题,在基本获得所需信息时即可终止检索。此法有时可保证信息的新颖性,但易于漏检而影响查全率。

(3) 抽查法

这是利用学科发展波浪式的特点查找文献的一种方法。当学科处于兴旺发展时期,科研成果和发表的文献一般也很多。因此,只要针对发展高潮进行抽查,就能查获较多的文献资料。这种方法针对性强,节省时间。但必须是在熟悉学科发展阶段的基础上才能使用,有一定的局限性。

(4) 引文法

也称为跟踪法、追溯法、扩展法,就是利用文献后所附的参考文献、相关书目、推荐文章和引文注释查找相关文献的方法。当查到一篇新发表的文献后,以文献后面所附的参考文献为线索,由近及远进行逐一追踪的查找方法。

这种由此及彼地扩大检索范围的检索方法,往往可以查到意想不到的切题文献。在检索工具不完备的条件下,广泛地利用文献综述或述评、研究报告等文献后所附的参考文献,不失为扩大检索范围的好方法。优点:不需要利用检索工具,查找方法简单。缺点:检索效率不高,漏检率较大。

引文索引法可分为两种:一种是由远及近地搜寻,即找到一篇有价值的论文后进一步查找该论文被哪些其他文献引用过,以便了解后人对该论文的评论、是否有人对此做过进一步研究、实践结果如何、最新的进展怎样等。由远及近地追寻,越查资料越新,研究也就越深入,但这种查法主要依靠专门的引文索引,如《科学引文索引》(*Science Citation Index*)、《社会科学引文索引》(*Social Sciences Citation Index*)。

另一种较为普遍的查法是由近及远地追溯,这样由一变十、由十变百地获取更多相关文献,直到满足要求为止。这种方法适合于历史研究或对背景资料的查询,其缺点是越查材料越旧,追溯得到的文献与现在的研究专题越来越疏远。因此,最好是选择综述、评论和质量较高的专著作为起点,它们所附的参考文献筛选严格,有时还附有评论。利用引文法高效率地查找文献的最有用的工具是利用引文索引。

2. 追溯法

又称回溯法。这是一种传统的查找文献的方法。就是当查到一篇参考价值较大的新文献后,以文献后面附的参考文献、相关书目、推荐文章和引文注释为线索而查找相关文献的一种方法。这些材料指明了与用户需求最密切的文献线索,往往包含了相似的观点、思路、方法,具有启发意义。循着这些线索去查找,不仅利用了前人的劳动成果,省却了很多时间和精力,而且可能在原来的基础上有新的发现。这是一种扩大信息来源最简单的方法,一般多利用述评、综述或专著进行追踪查找。在没有检索工具或检索工具不完整时可借助此法获得相关文献。但由于参考文献的局限性和相关文献的不同,会产生漏检。同时,由近及远的回溯法无法获得最新信息,而利用引文索引进行追溯查找则可弥补这一缺点。

3. 综合法

综合法又称为循环法,它是把上述两种方法加以综合运用方法。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这两种方法。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录为线索进行查找,如此循环进行,直到满足要求时为止。综合法兼有常规法和追溯法的优点,可以查得较为全面而准确的文献,是实际中采用较多的方法。这种检索方法多在创作人员选

定了课题,制定了创作计划后才使用,或检索工具不全时使用。

4. 浏览法

浏览法是在直接浏览各类信息源的过程中获取所需信息的检索方法。就是不通过文献参考检索工具,直接凭经验浏览本专业或本学科的核心期刊和图书资料,从中直接查找到所需要的文献,这样就能及时获得最新最直接的信息和原件全文,并能基本获取本学科发展的动态和水平。但是这样检索的范畴不够宽,因而漏检率较大,检索偶然性大、查全率低、回溯检索困难。因此,在开题或鉴定时还必须进行系统的检索。

5. 即期积累法

为尽可能及时全面地掌握文献,除了要引导学生在研究课题的前后突击查找文献外,还应注意平时的日积月累,养成经常阅读的好习惯。即从现刊中查找与自己课题有关的信息,也能够懂得在科学研究要把握最新的动态,及时了解别人在做什么,有什么新的发现,以便自己少走别人已经走过的弯路,保证自己的研究处于比较领先的水平。

6. 排除、限定和合取法

这实际上是将信息加工的方法融入检索中去。使用排除这一概念,是指对查找对象的产生和存在的状态在实践和空间上加以外在否定。把这一方法移植到检索中就是在时间或空间上极大地缩小检索范围。

如要查中国网络资源建设的文章,确定1994年以前Internet未进入中国,则可排除1994年以前的报刊资料,这就是排除法。限定法是相对于排除法而言的,指对查找对象在时间和空间上加以内在此的肯定。排除的结果必然是限定,反之亦然。令人满意的答案往往不是完整地记录在某一篇文章中的,如果把不同资料中涉及所需信息的记录都截取下来,汇集到一起,再经过去粗取精、去伪存真的加工,构成一个完整的答案,这就是合取法。采用这一方法,不仅要各类工具书触类旁通、灵活运用,还要学会分析来自各方面庞杂的材料。

总之,信息检索方法多种多样,各有利弊,应以课题需要和所处的信息环境灵活采用。上述方法,各有其优缺点,查找时要结合检索条件、时间、人手的限制等因素综合考虑。除了考虑方法以外,查阅技巧也是不可忽略的。有时方法对头,检索策略也无问题,可就是查不到近在眼皮底下的答案。因此,切忌匆匆翻阅,浅尝辄止,这样做往往成为漫无目的地胡猜乱翻,结果一事无成。如果初查失败,不要急于丢弃原来的方案,前后多查几页往往会找到有用的线索,甚至是意外收获。

4.5.2 信息检索途径

检索途径依赖于文献信息的特征。文献具有外部和内容两种特征。文献的外部特征主要是指文献载体上标明、易见的项目,包括文献题名、责任者、序号、出版者、出版地、出版年等;文献的内容特征包括所属学科及所属主题等。因此,根据文献的外部特征和内容特征,可将信息的检索途径分为两大类型。

1. 文献外部特征的检索途径

(1) 责任者途径

责任者途径即通常所说的著者姓名途径。责任者是指对文献内容负责或做出主要贡献

的个人或团体,包括著者名、评者、编者等。责任者途径是根据文献著者的名称查找文献信息的途径,是外文检索工具较为重要和惯用的途径。按著者姓名字顺排列,易于利用,又便于编排,也易于机械加工。

使用著者途径检索文献信息须注意文种的不同和姓名排列方式的差异,如单姓、复姓、父母姓连写、本名、教名以及姓名中附加的荣誉称号等。欧美人的姓名习惯名在前、姓在后,而目前使用的各种著者目录和著名索引则按姓在前、名在后的方式以字序排列,因此在具体检索时应按姓在前、名在后的字顺查找。

(2) 题名途径

题名途径也称书名途径。题名是表达、象征、隐喻文献内容及特征的词或短语,是文献的标题或名称,包括书名、刊名、篇名等。文献题名有正题名、副题名和辅助题名。题名检索途径是指根据文献题名查找文献信息的途径。它把文献题名按照字顺排列起来编成索引,其排法简单易行,易于查检。但因书名和篇名较长,不宜作为检索标志,又因不同文字的形体结构和语法结构有自己的特色,字尾变化复杂,所以难以把同样意义的文献集中于一处,实际使用价值已不为人们看好而逐渐不为人所重视。

(3) 文献类型途径

文献信息检索工具收选的信息源多种多样,如期刊、图书、科技报告、专利、技术标准、政府出版物、会议录等。为满足用户不同的检索要求,如会议文献或专利文献的查找,不少检索工具也增设文献类型检索途径,如专利号索引、图书索引、会议索引、报告号索引等,以满足不同用户的需求。

(4) 代码途径

代码途径也称序号途径,是通过已知的文献专用代号查找文献的途径。代码是一些文献类型的特有标志,与文献有对应关系,如国际标准书号(ISBN)、国际标准连续出版物号(ISSN),以及索引号、专利号、合同号等。

2. 文献内容特征的检索途径

(1) 分类途径

分类途径是指按文献内容的学科分类体系查找文献信息的途径。一般来说,一种检索工具的编制都必须按学科建立自己的分类体系,其收录的文献按分类目录中的排序进行编排,这样编排的结果可将同一学科的文献集中,按学科查找文献分类目录和分类索引是普遍使用的分类检索工具。分类途径的缺点是对于较难分类的新兴学科和边缘学科来说,查找不便。利用此途径查找时须首先了解反映学科体系的分类表,再将概念变换为分类号,然后按分类号进行检索。由于概念变换为分类号的过程中易出差错,所以也会导致漏检和误检。但是很多用户希望从其熟悉的分类系统及学科概念的上下左右关系了解事物的派生、隶属、平行等关系,满足族性检索的需求。分类途径能够较好地满足这一要求。

(2) 主题途径

主题是文献所表达的中心思想、所讨论的基本问题和研究对象。主题途径指根据表达文献主题内容主题词及其派生出的关键词为标志查找文献信息的途径。其主要检索工具是主题目录和主题索引,或标题词索引、关键词索引、叙词索引等。主题目录按文献内容主

题词组织,以文献所讨论的主题直接检索,可以查到分散于各学科里同一主题的文献。主题索引是工具书辅助索引之一,它可揭示包含该主题的文献信息在文献正文中的位置。

主题途径检索文献信息的优点是:用主题词作为标志,表达概念准确、灵活、专指度高,可使同一主题的文献集中,检索效率高;又由于主题词可随科技发展增加或更新,因此便于查找新兴学科的文献信息,在各学科和其分支交叉渗透日益增多的当前,主题途径较好地适应了这一要求。其缺点是:主题索引缺少学科系统的整体性和层次性,因此难以达到很高的查全率。

(3) 分类主题途径

分类主题途径是分类途径与主题途径的结合,它能够尽量避免两者的不足,取其所长。一般来说,它比分类体系更具体一些,无明显的学术层次划分;又比主题法概括一些,但保留了主题体系按字顺排序以便准确查检的特点。

3. 其他检索途径

(1) 出处途径

出处途径是数据库系统提供的检索途径。输入原文献的刊载处,如报刊名、出版单位名称,即可检索到该刊载处出版、发表的有关文献。

(2) 时间途径

时间途径是以文献的时间范围查找文献的途径。这是数据库检索系统普遍提供的一种检索方式,输入或选择某一时间,可检索到该时间出版发表的所有文献。时间途径一般和其他检索途径配合使用,很少单独使用。

(3) 任意词途径

任意词途径也称自由词途径。它是以自然语言编制的全文检索系统所提供的文献查询方式。自由词或任意词指直接取自文献本身,是未经规范和控制的语言输入字、字符、数字、词或词组等任意字或词,可检出所有在任意一处出现该字、字符、数字、词或词组的文献。

(4) 专门术语途径

专门术语途径主要是指一些辅助检索途径,如按化学分子式排出的分子式索引,可提供一种从分子式角度查找化学化工文献的目的;另外还有化学物质索引、合金索引、地名索引等各种专门索引,以满足查检特定种类文献信息的需求。

4.5.3 信息检索的步骤

检索步骤是对查找信息全过程的程序划分。检索步骤的科学合理性非常重要。主要分为以下8个步骤。

1. 检索课题的分析

分析检索课题,明确检索的目的、要求和范围,这是制定检索策略的基础和前提。任何一个检索都是根据已知去查找未知,通过分析检索课题,明确的已知线索越多,查获所需信息的可能性就越大。

通常检索目的可分为3种:

① 科研攻关型。解决研究或生产中的一些技术难题,如某一理论、方法、设备、过程等

的具体问题,这类检索要求查准率高,只要找到合适的文献即可。

② 课题普查型。针对某一课题收集系统、详尽的资料,这类检索要求查全率高,往往要检索若干年的文献,一般采用回溯检索的方式。

③ 研究探索型。密切跟踪、了解国内外某一方面的最新成果,掌握最新的科研动态,这类检索要求信息新颖、及时性强,多采用定题检索的方式。

明确检索要求与范围,主要应搞清楚检索课题所涉及的学科、专业范围;检索的主题概念是什么;能用哪些名词术语表达;所需要的信息类型是文献还是具体的数据、事实;对检索文献的类型、语种、出版时间、地域范围等有什么具体要求;是否还有其他的已知线索,如文献名称、有关人名、机构名称、文献号码(专利号、标准号、报告号)等,将已知线索一一分析出来。

2. 检索系统和数据库的选择

选择合适的检索系统,主要是选择检索工具/数据库,要根据检索课题的内容范围和要求来决定。要了解检索工具/数据库的学科专业范围及各种性能参数,其内容主要包括:

检索工具/数据库的类型是否满足检索需要。

检索工具/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。

检索工具/数据库收录的文献类型、文献存贮年限、更新周期是否符合检索需求。

检索工具/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,是否按标准化著录。

检索工具/数据库提供的检索入口是否与检索课题的已知线索相对应。

3. 检索途径和检索词的选择

检索途径主要根据分析课题时确定的已知条件,以及所选定的检索工具能够提供的检索途径来决定。常用的检索途径有著者、分类、主题、文献题名、文献号、代码(如分子式、产品型号)、引文等,还有文献类型、出版时间、语种等。每种途径都必须根据已知的特定信息进行查找。

检索词也称检索点,与检索途径相对应,是检索途径的具体化。确定检索词就是将检索课题中包含的各个要素及检索要求转换成检索工具/数据库中允许使用的检索标识,即用所选定的检索工具/数据库的词表(如主题词表、分类表)把检索提问的主题概念表达出来,形成主题词或分类号等,也可以是关键词(视检索系统而定)、人物姓名、地名、文献名等。

检索词是表达信息需求或检索课题内容的基本元素,也是数字信息检索系统中有关数据进行匹配运算的基本单元。检索词选择正确与否,直接影响着检索结果。在全面了解检索课题的相关问题后,应提炼主要概念与隐含概念,并排除次要概念,以便确定检索词。

检索词应满足内容匹配和形式匹配两方面的要求。内容匹配要求,即由主题概念转化而成的检索词应能准确、完整地表达检索课题的内容,这是由信息需求决定的。形式匹配要求检索时使用的语言和检索系统中使用的语言一致,这样检索词才能被系统“认识”。

检索词主要有语词性检索词和非语词性检索词两种形式。语词性检索词是各种数据库中必不可少的基本检索项,其常用的检索词分为受控词及受控词与字段符号的组合,和非受控词及非受控词与字段符号的组合。语词性检索词主要包括主题词即单元词、关键词、标题词、叙词和自由词等。非语词性检索词包括分类号、专利号、年代号、登记号、期刊代号、书

号、语种代号等。

检索词的确定一般包含以下几种操作:

① 切分是对课题的语句以词为单位进行切分,转换为检索的最小单元。当词切分后失去原来的意思时,则不应再切分,应注意保持检索词意义的完整。专用名词,如地名、机构名等不可切分,如机器人不可切分为“机器”和“人”。

② 删除是对不具有检索意思的虚词如介词、连词、副词等或者使用频率较低或专指性太高的词,过分宽泛或过分具体的限定词、禁用词,不能表达课题实质的高频词汇存在蕴含关系的可合并词都应予以删除,使语句转换成为关键词的集合。

③ 替换是对表达不清晰或容易造成检索误差的词,用更明确、更具体的词予以替换,实际上是对词语的规范。如“绿色包装”中的“绿色”实际上它的意思是环保、无污染等含义,应替换为“环保”“无污染”等表达明确、不易造成混淆的词。

④ 聚类是对切分出来的词进行语义上的合并,使语句由词的集合转换为概念的集合,信息检索主要是对所论述的事物或主题进行检索,而语句被切分成词以后,有可能失去信息用户本身所要表达的含义,因此要从概念的层面进行重新组合(需要对检索课题所涉及的专业有所了解)。

⑤ 扩充是将课题中筛选出来的词进行同义词、近义词、相关词、相关词的扩充词形变化,缩写词、翻译名等应尽量选全,以避免检索过程中漏检,提高查全率。

⑥ 增加是针对一词多义或者在检索结果中有很多干扰信息时,可采用“限义”的手段,具体的方法是增加限义词、挖掘隐台词、提取潜在的检索词;也可用逻辑“与”或逻辑“非”的方法增加限义词。

4. 检索式的编制

编写检索式是指将选择出的检索词用布尔逻辑算符以及截词算符、位置算符、字段代码等,按检索需要进行合理组配,形成表达信息需求的具体的检索式。

手工检索时,检索过程是由人的手翻、眼看、大脑判断来进行的,检索式存在于检索者的头脑中。

计算机检索时,检索提问与信息标识之间的对比匹配工作是通过计算机进行的,检索必须通过计算机能够理解和运算的形式——检索式来体现。检索式是检索策略的核心,是信息需求的最终体现,其质量好坏将关系到检索的效果乃至成败。

合理的检索式应达到两个基本要求:一是能充分而准确地反映信息需求的内容;二是能适应所查数据库的索引体系、用词和匹配规则,即与数据库中的信息标识相匹配。只有确切表达信息需求的检索式,才有可能从数据库中检出符合需要的信息;而只有与数据库中的信息标识达到高程度匹配的检索式,才能把需要的信息检索出来。

编制检索式时须注意以下事项:

① 对于同类或并列概念的词,用逻辑“或”进行组配,特别是采用自由词检索时,要尽量选全同义词、近义词等进行逻辑“或”组配;对于有交叉关系的概念,用逻辑“与”进行组配,但应注意去掉与课题无关的概念组配,以防因限制过严而漏检。

② 为了提高检索速度,在使用布尔逻辑算符时,应把估计出现频率低的词放在“AND”左边,把出现频率高的词放在“OR”左边,同时使用“AND”和“OR”时,应把“OR”放在

“AND”的左边。

③ 应考虑哪些词利用截词算符,哪些地方要用位置算符,是否需用字段代码加以限制。综合利用各种算符,可提高检索效率,使检索结果更为理想。

④ 应避免可能产生多种逻辑判断的组配。组配的结果只能表示一种含义,如果可能产生两种或多种含义,就应采取相应的措施加以限制。

5. 检索方案的评价、调整

在检索过程中,应及时分析检索结果是否与检索要求一致,不一致时要对检索词和检索式等做相应的修改和调整,调整检索策略,直至得到比较满意的结果。

(1) 检索结果信息量过多时检索策略的调整

产生检索结果信息量过多的原因可能有两点:一是主题词本身的多义性;二是对所选的检索词截词截得太短。在这种情况下,就要考虑缩小检索范围,提高检索结果的查准率。

- ① 减少同义词与同族相关词,去除相关性不强的检索词,少用甚至不用逻辑“或”。
- ② 增加限制概念,采用逻辑“与”连接检索词或进行二次检索。
- ③ 使用字段限定,将检索词限定在某个或某些字段范围内。
- ④ 使用逻辑“非”,排除无关概念。
- ⑤ 调整位置算符,由松变严。
- ⑥ 改模糊检索为精确检索。

(2) 检索结果信息量过少时检索策略的调整

对检索结果信息量较少的结果,可进行扩检,以提高查全率。造成检索结果信息量少的原因有以下几点:一是主题词选用不正确,如使用不规范的主题词、产品的俗称等;二是同义词、相关词、近义词没有运用全;三是逻辑“与”“非”运算符用得过多;四是各种限制过多过严。应该考虑扩大检索范围,提高检索查全率。

- ① 选全同义词与相关词并用逻辑“或”将它们连接起来,增加网罗度。
- ② 减少逻辑“与”的运算,放弃一些次要的或者太专指的概念。
- ③ 取消或放宽一些检索限定,如年限、学科、文种等。
- ④ 使用截词检索。
- ⑤ 扩大或修改检索字段限制的范围,如改题名为文摘或全文等。
- ⑥ 调整位置算符,由严变松。
- ⑦ 改精确检索为模糊检索。

在上述调整中,一是要从学科专业知识出发,选择泛指词、专指词及相关词,并确定组配逻辑;二是要利用计算机检索系统的功能,从文献的类型、年代、文种等外表特征入手对命中文献集合进行调整与控制,直到获得较满意的检索结果。

6. 检索结果的输出

检索结果的输出形式有多种,包括目录、题录、文摘、全文或自定义形式等,用户要根据需要进行选择。另外,检索结果的输出格式、方式也有多种,包括显示、复制、打印、传输、下载、电子邮件等,用户要根据费用的多少来进行选择。

7. 拓宽检索

利用检出文献的信息,考察检出相关文献的出处,可能会查找到一批相关文献,如中国知网的引文网络。另外,还可根据相关文献(全文)的引用文献或参考目录再查找到一批相关文献。

8. 充分利用各种资源

为了提高查全率,在检索中可以使用各种导航工具、搜索引擎等,进行全方位的网络资源搜寻。由于检索结果得到的只是文献线索,检索结束后,还要根据所获得的文献线索获取原文。

获取原文的途径有四种:一是利用本单位图书馆馆藏文献资源获取原文;二是利用联合目录通过馆际互借获取原文;三是利用全文数据库直接下载全文;四是利用检索结果中提供的著者或出版机构与之联系获取原文。在获取原文的过程中,要注意以下问题:

(1) 识别文献类型

不同类型的文献资源来源不同,在索取原文时首先就要区别文献的类型。

(2) 将缩写刊名恢复全称

检索工具中在文献来源项的著录中,常常将期刊名称按一定的缩写规则进行缩写。

(3) 识别不同语系文字的音译

在西文检索工具中,中文、俄文、日文等的文献作者和出版物名称通常采用音译法转换成英文进行著录。

(4) 利用各种收藏目录

在索取原始文献的过程中要根据不同类型的文献查找不同的联合目录、馆藏目录、联机公共目录等,查知其原文的收藏单位,再进行借阅。

(5) 利用文献传递服务,获取远程文献

许多大型检索系统提供文献传递服务,可以根据检索结果,在线提出索取全文的申请,通过 E-mail、传真等方式获得原文。

4.6 检索效果评价

检索效果是信息检索服务反映的效率和结果,它反映检索系统的能力,包括技术效果和社会经济效益之分。技术效果主要指信息检索系统的性能和服务质量,以及系统满足用户信息需求的程度。社会经济效益是检索系统通过满足用户信息需求所产生的社会效益和经济效益,如费用、时间等。技术效果和社会经济效益不是对立的,它们是互相联系统一体。

4.6.1 检索效果评价指标

假设进行检索时,检索系统把文献分成两部分:一部分是与检索需求相匹配的文献,并被检索出来,用户根据自己的判断将之分成相关文献(命中)a 和不相关文献(噪音)b;另一部分是未能与检索需求相匹配的文献,根据判断也可将之分成相关文献(遗漏)c 和不相关文献(拒绝)d。一般情况下,检索出来的文献数量为(a+b),相对整个系统的规模来说是很

小的,而未被检索出来的文献(c+d)数量则非常大。

1. 查全率(recall ratio)

查全率是指从检索系统中检出的与检索课题相关的文献信息数量与检索系统中实际与该课题相关的文献信息总量之比。

$$\text{查全率} = \frac{\text{检出的相关文献量}}{\text{检索系统中的相关文献总量}} = \frac{a}{a+c} \times 100\%$$

对于数据库检索系统,查全率为检索出的记录数与数据库中满足用户检索需求的记录数之比;而对于互联网信息检索而言,文献总量是很难计算的,甚至估算也很困难,因为互联网上信息是瞬息万变的,今天存在的信息,明天就可能找不到了,同时还会出现更多新的信息。要按传统的方式计算查全率,就要检验检索工具反馈的所有检索结果,而检索结果的数量是极其庞大的。为此,相对查全率是一种可以实际操作指标,但从其定义来看,人为因素的影响较大。

$$\text{相对查全率} = \frac{\text{专业人员检出的相关文献量}}{\text{全部实际检索出文献集合并集的总量}} \times 100\%$$

要提高查全率往往要扩大检索范围,但扩大检索范围可能导致查准率下降。为此需要提高标引质量和主题词表质量,优化检索提问式,准确判断文献的相关性和相关程度。具体来说,就是规范检索语言,选取适当的检索方法,选择合理有效的检索技术、检索策略,加强标引工作。

2. 查准率(precision ratio)

查准率是从检索系统中检出的有关某检索课题的文献信息数量与检出的文献信息总量之比。

$$\text{查准率} = \frac{\text{检出的相关文献量}}{\text{检索的相关文献总量}} = \frac{a}{a+b} \times 100\%$$

在理想的情况下,系统检索出用户认为相关的全部文献,用户相关性估计和系统相关性判断是重合的,即 $b=0, c=0$, 则查全率为 100%, 查准率也是 100%。实际上,这样的检索结果是不可能出现的。一般情况下,查全率的计算比较困难,因为检索系统中相关文献的总量是很难估算的。

同样,对互联网信息检索而言,真实查准率也是很难计算的。因为对于命中结果数量庞大的检索课题来说,相关性判断工作量极大,很难操作。为此可以定义一个相对查准率。

$$\text{相对查准率} = \frac{\text{检索者确定为相关的文献量}}{\text{检索者在检索过程中看过的文献总量}} \times 100\%$$

这个公式与传统的查准率定义有很大的差别,受人为因素影响太大,缺乏可重复性和客观性。

查全率反映所需文献被检出的程度,查准率则反映系统拒绝非相关文献的能力,两者相结合反映检索系统的检索效果。信息检索的理想状态是查全率和查准率都达到 100%, 但这是不可能的。查全率和查准率之间的互逆相关性是由英国 C. W. Cleverdon 领导的

Cranfield 试验所发现的, Cleverdon 在 1962 年首次将之运用于实际信息检索系统的评价实验(Cranfield II)中。也就是说,在排除了人为因素的情况下,任何提高查全率的措施都会降低查准率,反之亦然。究其原因不是在检索系统本身,而是在检索对象——文献,因为文献反映的信息与各个学科知识之间的普遍联系,各种知识之间的相互渗透、相互包容是影响查全率和查准率不能同时达到 100%;而成反比关系的客观因素,称为合理影响因素,由之造成的误检和漏检称为合理误检和合理漏检。

在同一个检索系统中,当查全率与查准率达到一阈值(即查全率在 60%~70%之间,查准率在 40%~50%之间)后,两者呈互逆关系,查准率每提高 1%将导致查全率下降 3%。因此,信息检索的最佳效果是查全率为 60%~70%,且查准率为 40%~50%。

虽然用查全率和查准率可以评价检索效果,实际上它们存在着难以克服的模糊性和局限性。由于检索系统中相关文献总量是一个模糊量,无法准确估计,故难以准确计算查全率。另外,“相关文献”对不同的检索者而言,认识不一致,其中含有主观因素。因此用上述公式计算查全率和查准率是相对的,它们只能近似地描述检索效果。

3. 漏检率(omission ratio)

漏检率是未检出的相关文献数量与系统中的相关文献总量之比。

$$\text{漏检率} = \frac{\text{未检出的相关文献量}}{\text{检索系统中的相关文献总量}} = \frac{c}{a+c} \times 100\%$$

漏检率与查全率是一对互逆的检索指标,两者之和为 1,查全率高,漏检率必然低。

4. 误检率(noise ratio)

误检率是检索出的不相关文献数量与检索出的文献总量之比。

$$\text{误检率} = \frac{\text{检索出的不相关文献量}}{\text{检索出的文献总量}} = \frac{b}{a+b} \times 100\%$$

误检率与查准率是一对互逆的检索指标,两者之和为 1,查准率高,误检率必然低。

5. 响应时间

响应时间是指从用户提问到提问被接受,再到检索结果输出,平均消耗的时间。手工检索响应时间人为因素影响较多,响应时间一般较长;单机检索系统的响应时间主要是由系统的处理速度决定的;网络检索的响应时间在相当大的程度上取决于用户使用的通信设备和网络传输速度等外部因素。就是同一检索系统,在不同的时间检索同一问题,其响应时间也可能不一样。缩短网络检索的响应时间,一方面可以提高服务器和客户机的整体性能;另一方面要增加网络的带宽,控制输入网络的数据量。

6. 其他指标

除了查全率、查准率和响应时间外,传统的评价指标还有以下几种:

(1) 收录范围

一个检索系统中收录的文献是否齐全,包括专业范围、语种、年代与文献类型等,这是提高查全率的物质基础。

(2) 工作量

从检索系统中获得相关文献消耗的精力与工作时间。

(3) 可用性

按可靠性、年代与全面性的因素检出文献的价值。

(4) 显示

检索结果的输出格式等。

网络检索工具,尤其是搜索引擎,其评价有其自身的特点。目前网络检索工具主要以自动方式在网络上搜索信息,经过标引形成索引数据库,索引数据库的构成是网络检索工具检索效果实现的基础。

检索工具提供的检索功能,直接影响检索效果,所以网络检索工具除了提供传统的检索功能外,还提供了一些高级检索功能,如多语种检索功能、自然语言检索功能、多媒体检索功能和相关反馈等。

在检索效果评价方面,除查全率、查准率和响应时间外,还应将重复链接数量和死链接数量作为评价指标。

4.6.2 提高检索效果的方法

信息检索效果是评价一个检索系统性能优劣的质量标准,它始终贯穿于信息存储与检索的全过程。用户在进行信息检索时,总是希望把与检索课题相关的信息全面(查全率)、准确(查准率)、迅速(响应时间)地检索出来,获得满意的检索效果。

要提高检索效果,主要应从两方面入手:一是提高检索系统的质量;二是提高用户利用检索系统的水平。

检索系统的质量不由用户控制,要提高检索效果,主要从用户入手。

1. 提高检索人员的素质

信息检索是用户具体操作的,人为因素占支配和主导作用。在信息检索中主要依靠检索人员的大脑不断地进行思考、判断、选择和决定,如检索策略的制订、检索途径与方法的选择、检索技术的运用、检索式的构造等,检索效果与检索人员的知识水平、业务能力、经验和工作责任心密切相关。

(1) 提高检索人员的知识素质

检索人员的知识素质包括知识、技能和能力。知识是指信息学、信息组织与存储、信息检索、计算机应用、外语等知识;技能是指咨询解答、信息整序、语言与文字表达等技能;能力是指捕捉信息的能力、超前思维的能力、综合分析的能力等。只有具备一定的检索知识和广博的知识内涵,才能形成一定的检索能力,从而提高检索效果。

(2) 提高检索人员的思想素质

思想素质是关系到检索人员全面素质提高的重要因素,并影响着检索效果,主要体现在职业道德精神、检索结果的辨别分析、检索观点的公正等。因此,提高检索人员的思想素质,就是要避免人为因素的影响,进而保证检索效果的提高。

2. 优选检索系统

检索系统的质量是决定检索效果的基础,所以优选检索系统是保证检索效果的重要环节。由于检索系统类型多种多样,各具特色,同时还存在交叉重复的现象,对一般用户来说,要熟悉与其专业相关的检索系统的功能不是一件容易的事情,选择恰当的检索系统就更加困难,这就要求检索人员必须全面了解检索系统,如收录范围、标引语言、排检方式等,才能根据检索课题的要求,选择专业对口强的检索系统。不存在可以满足任意检索需求的检索系统,每一个检索系统都有自己的强项和特点。提高检索效果,必须对检索系统进行优选。

3. 优化检索策略与步骤

正确的检索策略,可优化检索过程与检索步骤,有助于求得查全率和查准率的适当比例,节省检索时间与费用,取得最佳的检索效果。

4. 精选检索词

使用检索系统进行信息检索时,检索词的选择也是一个重要环节。在选择检索词时尽量使用专指性强的词,学会使用截词,不使用常用词,避免使用多义词,避免出现错别字。

5. 巧构检索式

运用逻辑算符、位置算符、限定符、通配符及相关的检索技巧,来巧构检索式是提高检索效果的有效途径。

6. 熟悉检索代码与符号

检索代码与符号是使用检索工具的语言保证,是检索与系统相匹配的关键,选取是否恰当,将直接影响检索效果。因此,检索人员必须利用相应的分类表、词表,选取与检索工具相匹配的正确代码与符号。

7. 鉴别检索结果

检索结果的鉴别分为印刷型资源和电子资源。对印刷型资源可从版权页上的出版者、作者和序跋中的作者及相关内容介绍等进行鉴别;对电子资源主要从信息来源与出版、权威性、用户、网站内容、时效性等方面鉴别。

4.6.3 信息检索的常见误区及改进策略

1. 忽视对参考文献的追踪

参考文献是科技论著的一个重要组成部分,它明确标示出引用他人学术思想、理论、成果和数据的部分,并注明其来源,以体现科学的继承性和对他人劳动的尊重,这也表明了科学的严肃性和真实性。读者可以通过所引用的文献来评估论著的学术水平,这有助于鉴定和确认成果。参考文献方便用户进一步检索有关资料,实现资源共享。

要想准确、有效地利用参考文献,应注意以下几个原则:

(1) 有的放矢地多读几篇质量较高、相关度较大的文章(即期刊影响因子高、文章引用次数多、研究领域专业期刊上的文章),将其正文内容与它的出处关联起来,由此得到线索,继而得到原文。

(2) 统计不同文章关键点所引参考文献。如果某篇参考文献在不同文章里均出现过,则应获取其原文阅读。

(3) 统计所引参考文献的著者。若该著者在相同或不同文章里出现多次,则应获取其原文阅读,并以该著者作为检索路径检索所有的文章,判断该作者所在的课题组对该领域的贡献。

(4) 统计所引参考文献的期刊。若该期刊在相同或不同文章里出现多次,则说明它是该研究领域中十分重要的期刊,应定期浏览该刊上所有的文章。科学研究的体系虽然不同,但其理论和实验方法是相通的。若用主题词检索不到某些文章,最好的办法是浏览积累。

2. 忽视对核心期刊的追踪

在各种电子文献数据库日趋完备、论文全文容易获取的今天,多数检索者往往沉迷于利用主题词进行检索。其实,这种传统主题检索方式存在以下固有缺陷:

- (1) 主题词不易选取;
- (2) 少数的主题词无法反映全文的内容;
- (3) 无法超越自己的认知局限;
- (4) 无法检索到自己所不知道的知识。

如果定期浏览自己专业中最重要的4~6种核心期刊,则可以避免主题词检索带来的漏检。通过浏览,科研人员可以从“面”上了解学术进展和热点(主题词检索则只有一个“点”),然后根据个人兴趣和工作进展,逐篇仔细阅读新作。跟踪几种核心期刊与跟踪本领域中几个重要人物的工作也同样重要。

3. 忽视对学科带头人和科研团队的全方位了解

在检索和阅读文献时,不要忘记关注以下问题:该研究起源于什么时候、哪个研究机构;目前哪个课题组最活跃、领军人物是谁、学术思想是什么;最近发表了哪些文章、发表的文章刊登在哪些刊物上。实际上,EI、SCI、CA的网络版均有这些分析功能,目的是从对该领域一篇篇文章的平面理解发展为全方位的立体了解。

在科学研究发表文章过程中,也存在着“二八”定理,即20%的科研团队发表了该领域80%的文章。一个优秀的科研团队往往有一个比较完整的学术思想,而且它也在不断地发展和变化。由于研究时间和领域跨度大,它的很多学术思想会散落在各个时期的文章里,因此应该对该科研团队领军人物进行全方位的检索,这样得到的收获会大大超越主题词检索的结果。虽然只简单用了“著者检索”途径,但它实现了时间和领域的双重跨度检索(所谓领域跨度,是指体系不同但方法相同的跨越),这是用主题词检索无法做到的。因此,除了用主题词检索外,科研人员还应对领域中的研究领军人物及其科研团队进行全方位的追踪,将一篇篇毫无头绪的文章与该课题组相联系,把握大师们的思想脉络,形成一个由人、时间、文献三者构成的立体文献数据库。

4. 非科学性地阅读与整理检索到的文献

当今各行业、各学科的研究论文浩如烟海,而在我国由于职称晋升和研究生毕业的需要,也加剧了文章数量畸形增长、质量参差不齐的局面。即使作者认真地写作,也可能因作者认知水平、实验条件所限而导致错误的结论。好多人往往没有怀疑精神,不会批判性阅读,有些错误的研究不仅浪费时间,还会造成学术思想混乱。因此,在阅读文献的过程中切忌全盘吸收。

批判性阅读是一种科学的阅读态度和方法,指在阅读时应当保持独立思考,从立论所依

据的事实证据和收集证据时所使用的方法开始,检查推理概念的准确性和逻辑性,以及结论是否恰当准确地反映了论文或研究成果的价值,既不因循守旧,也不轻信盲从。对待古今中外的一切文献都应持这种态度,即“尽信书不如无书”。没有批判的勇气和智慧,阅读就不会深化、有所创新或突破。

批判性阅读采用以下方法比较容易入手:

- (1) 比较和对照不同作者对某一问题观点的异同点;
- (2) 对前人的理论和结果进行批判(如探讨模型或机理的前提是否合理,测量的范围是否恰当);
- (3) 思考作者提出这些说法和想法的理由是什么,是事实、理论还是信念;
- (4) 注意作者自己或者其他作者有没有对该篇文献进行过修正或评价(用 SCI 引文检索的方法);
- (5) 从现有文献寻找课题的切入点,这是科研人员创新的基础。

当今世界,面临浩如烟海的文献挑战,最重要的对策之一是养成批判性的阅读能力和选择能力。

4.7 思考题

1. 简述信息的基本概念和分类是什么?
2. 信息、知识、文献、情报之间的关系是什么?
3. 什么是主题语言? MeSH 字顺表的词汇有哪些类型?
4. 提高文献查全率的方法有哪些?
5. 提高文献查准率的方法有哪些?
6. 信息检索的完整步骤是什么?



【微信扫码】

本章相关资源