



互联网环境下 信息检索与分析

主 编：张凌云 赵宏亮 张媛媛

HULTANWANG HUANJING XIA
YINXI JIANSUO YU FENXI



文化发展出版社
Cultural Development Press

互联网环境下 信息检索与分析

主 编：张凌云 赵宏亮 张媛媛

参 编：刘常青 邱慧芳 蔡 浩

HULIANWANG HUANJING XIA
XINXI JIANSUO YU FENXI



文化发展出版社
Cultural Development Press

图书在版编目 (CIP) 数据

互联网环境下信息检索与分析/张凌云, 赵宏亮, 张媛媛主编; 刘常青, 邱慧芳, 蔡浩参编.
—北京: 文化发展出版社, 2018. 7

ISBN 978-7-5142-2365-1

I. ①互… II. ①张… ②赵… ③张… ④刘… ⑤邱… ⑥蔡… III. ①互联网络—情报检索
IV. ①G254.928

中国版本图书馆CIP数据核字(2018)第145516号

互联网环境下信息检索与分析

主 编: 张凌云 赵宏亮 张媛媛

参 编: 刘常青 邱慧芳 蔡 浩

责任编辑: 张 琪

责任校对: 岳智勇

责任印制: 邓辉明

责任设计: 侯 铮

出版发行: 文化发展出版社 (北京市翠微路 2 号 邮编: 100036)

网 址: www.wenhua fazhan.com

经 销: 各地新华书店

印 刷: 北京建宏印刷有限公司

开 本: 787mm × 1092mm 1/16

字 数: 202千字

印 张: 13.75

印 次: 2018年9月第1版 2018年9月第1次印刷

定 价: 88.00元

I S B N : 978-7-5142-2365-1

◆ 如发现任何质量问题请与我社发行部联系。发行部电话: 010-88275710

前言

PREFACE

信息检索是一门门槛低但精通难的课程，它主要侧重于技术的培养，但技术的培养依赖大量的理论和背景知识的支持，信息检索是一门融合了多学科知识的交叉学科，因此学习起来会有一定的难度，并不是像很多读者认为的那样简单。

本书作者在多年教学实践中，通过对《信息检索》的课程培养任务和培养目标的研究，在实践中从学生的理解能力、课程的深度、教学效果等方面总结并形成了一套完整的教学体系和教学内容。本书既兼顾了掌握初级检索技术的低年级学生的能力，也加入了掌握较多检索技术的高年级学生在科研和学习过程中需要进一步学习的内容。

在内容安排上，由浅入深，去掉沉闷的理论，用一些简洁、易懂的事例解释原理和技术，方便读者在使用过程中，边学习边操作，侧重实用性和针对性。

本书共分为六章，由多位作者共同编写。张凌云，副研究馆员，2006年开始研究检索技术。发表专业论文数篇。负责一、二、三（部分）、五、六章撰写。刘常青，1994年3月开始从事中文图书编目，副研究馆员，发表图书馆学专业论文数篇。主要负责第四章的撰写。邱惠芳，副教授，从事图书情报和网络信息管理服务工作二十多年。撰写图书情报相关论文八篇，参与资源建设国家级和省级课题各一项，主要负责信息素养部分撰写。蔡浩，网络工程硕士，从事计算机网络与通信技术方向研究。目前，主要负责学校网络与信息化规

划、建设及运维工作，已完成了学校多项研究课题，并成功提交发明专利一项。主要负责网络安全部分的撰写。在此感谢这些作者的辛勤付出，并在本书的撰写过程中提出大量有益且中肯的意见和建议。

文中一些楷体文字是事例或引用的规程、表格等内容，因此特别标示出来。

带*号的部分属于相对难度较高的部分，可进行选择性学习。

目录

PREFACE

第1章 / 导论 / 1

第一节 信息检索的理论意义和价值 / 1

第二节 信息检索课的源流 / 4

第三节 信息检索课的定义和特征 / 5

第2章 / 基础知识 / 9

第一节 信息的基础知识 / 9

第二节 情报学基础知识 / 15

第三节 检索基础知识 / 31

第四节 文献信息检索系统的评价 / 55

第3章 / 网络信息检索 / 63

第一节 信息素养 / 63

第二节 搜索引擎的原理与使用 / 87

第三节 网络安全与信息安全 / 108

第4章 / 图书馆功能 / 119

第一节 图书馆概述 / 119

第二节 图书馆的性质、职能和类型 / 123

第三节 高校图书馆功能 / 129

第四节 《中国图书馆分类法》简介 / 137

第五节 中国现行的其他分类法简介 / 151

第5章 / 中外文数据库的检索与利用 / 159

第一节 文献数据库 (Database) / 159

第二节 文献评价系统 / 174

第三节 知识发现系统 / 178

第四节 跨平台检索系统 / 186

第6章 / 课题检索规范流程 / 189

第一节 课题检索 / 189

第二节 课题检索结果的管理与利用 / 204

第三节 课题检索报告设计 / 209

参考文献 / 211

第1章 导论

信息检索，是一门综合性学科，结合了文献学、情报学、数据库知识、计算机基础知识、信息管理学等多学科内容，因此，学习这门课程需要一些必要的知识扩展，具有一定的知识面，才能充分理解和掌握信息检索这门课程。

第一节 信息检索的理论意义和价值



一、我们为什么要查找信息？

信息需求以信息为对象，是人们在社会实践活动中对信息、知识的欲求。人类既要通过信息来认识世界、改造世界，也要根据所获得的信息来组织知识，因此，信息需求是一种普遍存在着的心理现象，即用户以方便自己的形式表达出来的及时获取问题所需要的完整的信息要求。查找信息是为了了解、学习，是一个终生进行的社会活动，也是一个复杂、烦琐的心理活动，是社会生活中人与人之间的相互沟通、思想交流、情感传递，也是人与自然之间的交流与对话，是人类认识自然、利用自然、适应自然与改造自然的必要手段。信息

的查找与利用是人类需求的一个侧重面，一个客观的信息包含着多种含义，不同的人对它的理解与运用会有差别，并影响最终行为的结果。

二、我们如何查找信息？

随着人类社会的不断进步，信息量也从贫乏进化到巨量，甚至海量，世界上的信息已经浩如烟海，是各种良莠不齐的信息影响了我们，还是我们影响了信息，我们查找到的信息是不是我们需要的信息？

人类有着悠久的文明历史，历代流传下来的以及今天出版的众多文献是巨大的知识宝库，随着网络技术的发展，每一个人都成了一个微媒体的信息发布源，每时每刻都在产生数量巨大的信息。用科学的方法系统地开发和利用这一丰富的信息资源，使我们的生活、学习和研究有一个较高的起点，对发展科学技术具有重要的作用。

如何查找信息，或者说，信息是如何被我们找到的？很多人并没有思考过这个问题，这就像一个人的名字一样，如果一个人没有名字，我们是无法找到他的。因此，如何查找信息，更深层次的提问出题，另外信息是如何被定义、被管理的？

例如：一本书被这样一些关键词定义：计算机网络—影响—社会生活。

它可能还被定义为：信息交互—普适计算—寻路—决策。

海量的信息是无序状态还是有序状态？有序状态的形成规则是什么？是哪些因素导致了信息检索过程和结果的不确定性？只有掌握了这些规则才能在现在的信息时代不被信息淹没，并高效地提取信息、利用信息。信息管理是一个多学科交叉的领域，涉及面之宽，是以前的学科划分中难以确切界定的，其中包含了大量的技术和规则，需要进行一定的学习和训练才能掌握其中的技巧和技术。

三、学会评价信息的方法

任何事物都有正、反两个方面，信息也如此，信息有真假之分、有有序与无序之分、有正负价值之分，海量的信息中一定有一些噪音存在，如果对接触到的信息全盘接受，不加甄别，人的心理必然承受不了这种信息的轰炸，导致信息致盲、思维混乱。而相同的信息被不同的使用者利用，也会产生不同的影响，如何评价信息的正确性，以及如何正确、合理地使用信息，是当代信息社会中，人们应该具备的基本素质和能力。

四、学习信息检索的意义

1. 掌握了信息检索的方法和技能，就可以使人们的大脑从烦琐的记忆活动中解放出来，在知识的海洋中有目的地、迅速地获取信息、吸收信息，从而进行知识创新性和成果创造性的工作。

2. 开启知识宝库，开阔眼界。“信息检索”课程的学习将把学生引入到更广阔的知识范围中去，向着未知的方向进行探索。掌握了信息检索的方法和技能，就可无师自通，不断扩展知识面，不断调整其知识结构，很快地找到一条汲取和利用大量新鲜知识的捷径，从而进入旺盛的创造期。

3. 通过学习，可以在一定程度上消除自然语言和学科专业语言的障碍。我们在学习过程中有三方面的语言障碍，即自然语言障碍、学科专业语言障碍和文献信息系统中的检索语言障碍。

4. 随着科学技术的飞速发展，新理论、新观点、新技术、新产品层出不穷，人们日常的生活和科研活动已经离不开信息的查找，通过信息检索，可以了解国内外科学技术发展的脉搏和全局，及时了解最新研究成果。

5. 少走弯路，节省时间。掌握了信息检索的方法和技能，将大大提高用于思考和创造性劳动时间的比例，使学生节省科研和非创造性劳动时间，学习更多的知识，创造更多的成果。

6. 培养信息能力。信息能力指感受、获取、利用信息的能力，是当代人的基本能力。能力是成功地完成某种活动所必需的个性心理特征。能力分一般能力和特殊能力，前者指都必须具备的基本能力，如观察、记忆、学习、信息等；后者指处理某些专业活动所必须的能力，如音、体、美、工程技术等。

第二节 信息检索课的源流



信息检索是一门适用性很强的课程。主要是培养学生的检索技能和情报意识，是一门国家教委统一命名的“科学方法课”，是一门多元复合性质的综合课程。

我国高校信息素质教育起步于1981年开设的文献检索与利用课，1981年7月和8月，《人民日报》先后刊载了刘毅夫、潘树广的建议在高校开设文献检索课的文章，同年教育部颁发了《中华人民共和国高等教育图书馆工作条例》，第一次将文献检索课规定为高校图书馆的工作任务之一。1983年全国高校图书馆工作委员会秘书处筹办了理工、医药、农业3个文献检索课师资培训班，10月召开了“全国高校《文献检索与利用》课专题讨论会”。1984年2月，教育部下达了“关于在高等学校开设《文献检索与利用》课的意见”。1985年，教育部下发了65号文件“关于改进和发展文献课教学的几点意见”。通知中说：研究生更应该补上这门课。各高等院校教务处应当为文献检索课安排必要的时间，要和其他课程一样安排教师进行教材编写与印刷，并尽可能在运用现代化教学手段上予以协助。组织开设文献课是高校图书馆义不容辞的任务，应当把它看作是图书馆工作改革的一部分并列入日常工作计划。专职或专职从事教学的人可以纳入教师编制，他们承担的部分参考咨询、情报服务工作应计入教学工作量，应按规定给他们足够的备课和编写教材的时间，并安排他们进修和调研。1992年5月，原国家教委再次印发了《文献检索课教学基本要求》，对文献检索课的课程性质、教学目的要求等作了细致而全面的规定，

从而使文献信息检索课从形式到内容更加规范化、系统化。1996 年成立了全国文献检索课教学指导小组。2015 年 12 月 31 日, 教育部下达了教高〔2015〕14 号文件, 第六章服务, 第三十一条中规定: 图书馆应全面参与学校人才培养工作, 充分发挥第二课堂的作用, 采取多种形式提高学生综合素质。图书馆应重视开展信息素质教育, 采用现代教育技术, 加强信息素质课程体系建设, 完善和创新新生培训、专题讲座的形式和内容。

第三节 信息检索课的定义和特征



一、信息检索课的定义

信息检索课依据的文献检索学, 是一门科学, 它有自己的研究对象和理论体系。国家技术监督局在 1992 年 11 月颁布了《国家标准学科分类代码》, 将文献检索学列为类目, 成为一个独立的三级学科, 学科代码 8702030, 属于图书馆、情报与文献学下属类目文献学的文献检索学, 同时在情报学下有情报检索学及情报技术等专门的三级学科类目; 文献检索学作为独立的学科, 是研究(情报) 存储和查找全过程的一门新兴学科, 其研究对象为检索语言、信息源、检索系统、检索策略、检索服务、用户教育。

1. 检索语言

检索语言包括信息检索语言和计算机检索系统的查询语言。

检索语言是根据检索的需要而发明的一种专门的人工语言, 是在文献检索领域中用来描述文献和表达信息检索提问的一种专用语言。检索语言与检索效果之间有着密切的关系, 在检索过程中起着极其重要的作用。

查询语言则是依据计算机系统对信息的标引、著录规则, 对信息进行分析、组织、存储所使用的语言, 只有当检索语言与计算机查询语言相匹配, 才能在计算机检索系统中将所需信息调取出来, 如图 1-1 所示。

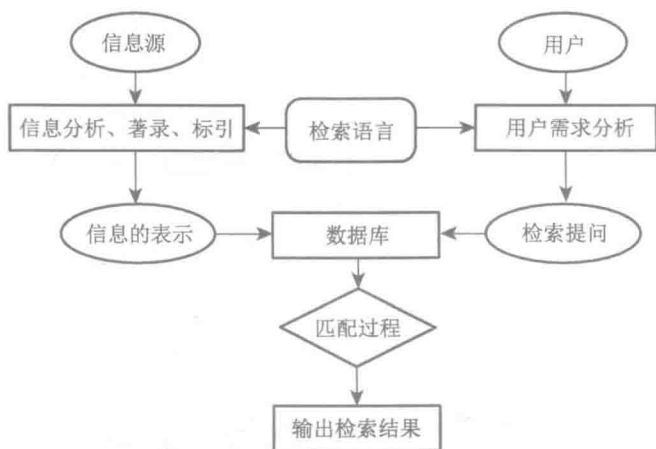


图 1-1 检索原理框架图

2. 信息源

信息源是由可以产生和持有信息的个人或机构以文字、符号、声波、光波等形式记录信息并形成信息集合。信息源有同构信息源和异构信息源两种，以同构信息组成的信息源一般称为数据库；异构信息源则以网络信息为主。

数据库：特指以文献资料为主要内容的文献书目记录、事实和数据，以及文献全文的集合等，印刷型检索工具书、图书馆卡片目录和计算机检索条件下的机读型目录也是数据库的主要建设内容。包括建库方面对文献的选择与描述，根据一定的检索语言对文献进行的标引，文档的结构与更新，数据库的评价标准与评价技术等。

3. 检索系统

检索系统包括手工检索系统、穿孔卡片检索系统、光电检索系统、电子计算机检索系统、网络检索系统和缩微胶片检索系统等。

4. 检索策略

检索策略包括处理情报检索提问的逻辑和查找步骤的科学安排。

信息检索是依据一定的需求或课题而进行的，当一个需求被唤醒或课题被确定下来后，需要对需求和课题进行分析，并进行筛选，明确哪些信息是需要的，哪些信息是不必要的，再决定采取何种方式和手段进行检索。在检索过程

中会碰到大量的信息噪音,需要调整检索策略并进行多次检索以确认检索结果。

5. 检索服务

检索服务包括对用户需求的调查与分析,各种服务方式及其实际效果的评价。

6. 用户教育

用户教育包括对用户检索知识的调查、检索能力的测试和教育内容的选择,相应的教学方式、方法等。

文献信息检索学的任务是通过文献的收集、筛选、存储、检索和利用规律的研究,充分发挥检索系统的效能。

在日常的信息检索活动中,由于所掌握的知识以及对事物思考方式的不同,人们提出问题的方式也就千差万别,同一个信息的查找,不同的人会使用不同的提问方式。从表象上看,检索过程是信息检索提问与信息集合标识之间的匹配运算,但实际的问题却要复杂得多,它不仅涉及用户认识结构为基础的信息需求唤醒、提问表达与转换、检索标识的形成和检索结果的相关性与适用性判断,而且涉及检出信息的理解与吸收利用。读者从事信息检索活动时,从需求唤醒到提出问题即将自己的信息需求转化成一定的信息标识,并利用这一标识进行信息查找,是一个复杂的过程,影响因素很多,涉及信息需求者的知识结构、信息源的知识结构、二者的匹配程度,以及对信息的利用。学习和掌握信息检索技术,在查找的过程中才不会被众多的影响因素所左右,尽量让获取的信息趋向于信息需求。

二、信息检索课的特征一

信息检索也是一门方法课,是学习获取信息知识、情报技术的学问,它有自己的知识内容体系,需要专门的学习才能掌握。信息检索,是一门综合性学科,结合了文献学、情报学、数据库知识、计算机基础知识、信息管理学等多学科内容,因此,学习这门课程需要有一些必要的知识扩展,具有一定的知识

面，才能充分的理解和掌握信息检索这门课程。

面对浩瀚的信息资源，利用恰当的方法，准确地找到自己想要的信息，并将信息进行合理的分类、管理，利用信息产生价值，是大学生在学习阶段应该具备的基本能力。

信息检索依赖于信息管理系统，信息管理系统对信息的管理通过特定的语言和组织方式对信息进行收集、整理、加工、组织、转化和传递。学生通过学习，掌握了这些语言并了解其组织方式，才能在信息集合中查找到自己所需要的信息或资料，提高读者检索与利用信息的技术和能力。

三、信息检索课的特征二

信息检索课是一门信息素质课，是培养学生信息生理、心理和文化素养，强化信息意识，提高信息科学素养、具备信息道德和信息鉴别能力、提高信息劳动的效率、增强信息交流水平的课程，其目的是为学生增长知识、培养全面能力、提高综合素质。

第2章 基础知识

第一节 信息的基础知识



一、信息的概念

（一）信息（information）的定义

信息到底是什么，目前，学术界还没有形成一个一致的、明确的、完整的定义。

《辞海》中，信息是指客观的消息、情况、情报等，是指对消息接受者来说预先不知道的报道等。

《新华字典》将信息表述为：信息是指音信、消息。

《韦伯斯特词典》将信息表述为：信息是知识和情报的通信和接受，是通过调查、研究或要求而得到的知识，是情报、消息、新闻、事实和数据等。

信息论中，信息是指用符号的报道，报道的内容是接受者预先不知道的，是事物的运动状态和关于事物运动状态的陈述。

在上述定义中，信息论中的信息定义是最大范畴的定义，是泛指事物的运动状态和关于事物运动状态的陈述。《辞海》中的定义与此类似。

《新华字典》和《韦伯斯特词典》中对于信息的定义则增加了信息使用的主体，也就是对信息产生感应的接受者。

我国科学家钱学森对信息有过这样的阐述：信息是激活了、激化了的知识，是为了解决一个特定的问题所需要的知识。这一阐述将信息进一步明确化为情报的一种，具有产生价值和意义的特点。

本书讨论的信息更接近于钱学森先生对于信息的阐述，是在日常生活中用以帮助我们更好的学习和生活所使用的信息。充分了解信息存在的形式，有利于在检索过程中准确把握信息的提问策略，提高信息检索的效率。

中国国家标准 GB/5271.1—2000《信息技术词汇第一部分：基本术语》中对信息的定义是：关于客体（如事实、事件、事物、过程或思想，包括概念）的知识，在一定的场合具有特定的意义。

信息是一个复杂的概念，不同的研究者和研究领域对信息都有着不尽相同的定义，在如下这个定义中：“是认识主体所感知或所表述的事物运动的状态与方式”，信息除了事物特征外，还加入了认知主体这一侧面，因为信息存在依据不同的主体是有不同认识的。由于认知主体的不同，信息也具有多重性。对同一事物的认识，会根据认知主体的不同，以及同一认知主体在不同环境和条件下，而产生不同的内容。

在文献学学科中，信息与知识和文献等概念的关系是从信息数据而来，表达事物运动的形态；有用的信息形成知识；将知识记录在载体中形成文献。因此，信息到知识的进化是一个金字塔形状的发展进程，其结构如图 2-1 所示。

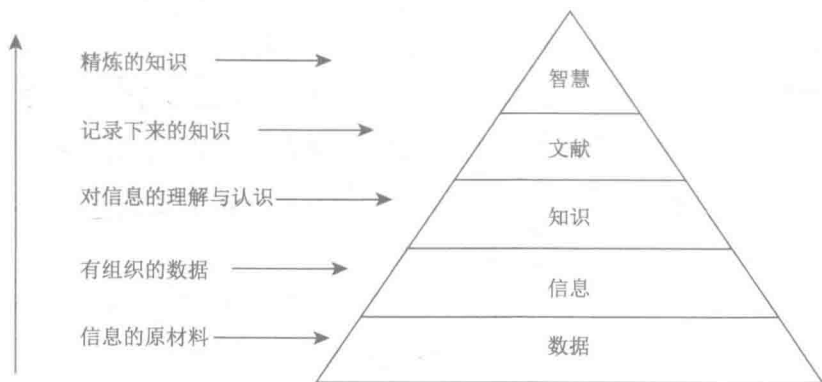


图 2-1 信息的结构