

WENXIAN JIANSUO YU ZHISHI FAXIAN ZHINAN

文献检索与知识发现指南

主 编 / 吉久明 孙济庆

副主编 / 曾 媛 朱世琴 霍丽萍

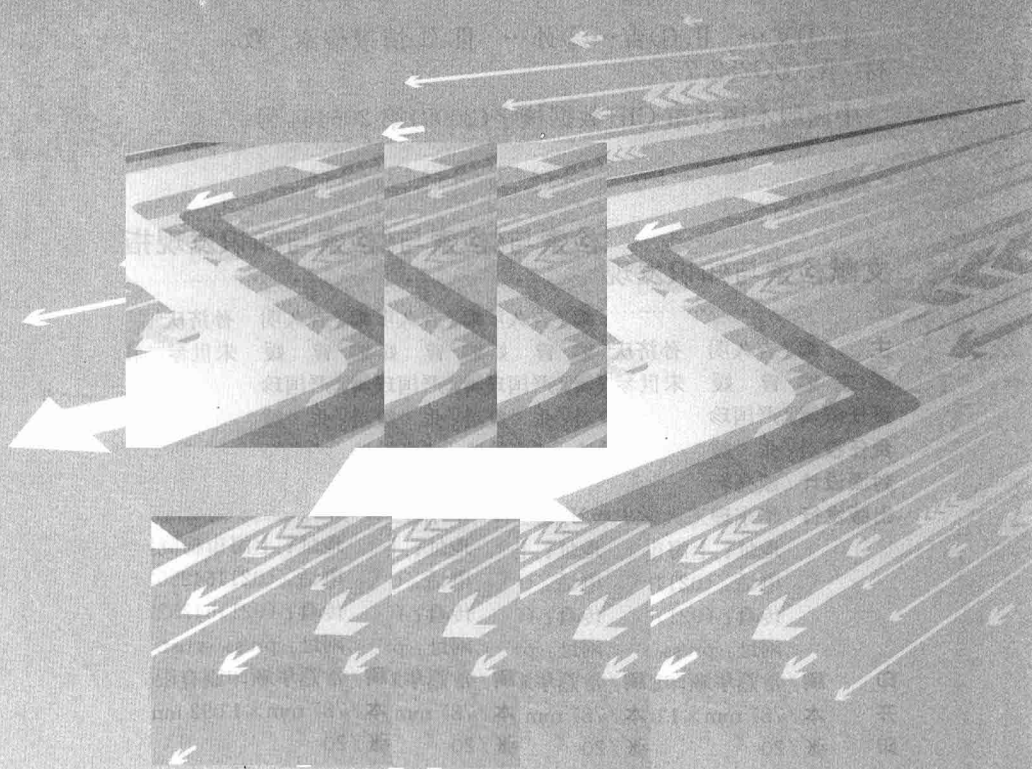


华东理工大学出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

文献检索指南

与知识发现

副主編／曾 媛 朱世琴 霍麗萍



華東理工大學出版社
EAST CHINA UNIVERSITY OF SCIENCE AND TECHNOLOGY PRESS

内 容 提 要

本教材在内容编排和实践课题的设计方面与现有教材有较大的不同,不但涉及基本的检索问题,也涉及与大学生密切相关的信息文化,还设计了操作性较强的研究性实践课题。本教材从搜索引擎的使用说起,阐述检索的基本原理和检索工具的基本要素;结合信息安全与信息道德素养的基本内容,详细介绍知识产权的内容。突出介绍学位论文文献综述阅读、学术文献分析、知识发现方法。最后,通过具体的案例对文献检索和知识发现的过程、结果作了详细的介绍。本书可以作为文献检索教师、科研人员、大学生的文献检索教材或学习检索与知识发现技能、提高信息文化素养的参考书。

图书在版编目(CIP)数据

文献检索与知识发现指南 / 吉久明,孙济庆主编.
—上海:华东理工大学出版社,2010.1
ISBN 978-7-5628-2523-4

I. ①文… II. ①吉…②孙… III. ①情报检索—教材 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2009)第 206464 号

文献检索与知识发现指南

主 编/吉久明 孙济庆

副 主 编/曾 媛 朱世琴 霍丽萍

责任编辑/严国珍

责任校对/张 波

封面设计/戚亮轩

出版发行/华东理工大学出版社

地址:上海市梅陇路 130 号,200237

电话:(021)64250306(营销部)

传真:(021)64252707

网址:press.ecust.edu.cn

印 刷/常熟华顺印刷有限公司

开 本/787 mm×1 092 mm 1/16

印 张/20

字 数/495 千字

版 次/2010 年 1 月第 1 版

印 次/2010 年 1 月第 1 次

印 数/1-5 000 册

书 号/ISBN 978-7-5628-2523-4/TP·170

定 价/32.00 元

(本书如有印装质量问题,请到出版社营销部调换。)

编辑委员会

主 编 吉久明 孙济庆

副主编 曾 媛 朱世琴 霍丽萍

编 委 (以姓氏笔画为序)

吉久明 朱世琴 任福兵 孙济庆 严素梅

李 楠 李 慧 汪人山 陈 荣 俞 璇

唐永林 曾 媛 霍丽萍

前言

在信息爆炸的时代,熟练掌握从信息的海洋中快速定位相关的、有用的且有价值的信息的技能,是 21 世纪人才的基本素质要求之一。

随着信息技术的不断发展,信息检索工具也发生了巨大的变化,从早期的文献检索发展到当前的信息检索,还将步入知识检索的高级阶段。对应不同阶段的信息检索工具,国内有近 800 种信息素养培养配套教材。这些教材包括专论和通论两类。如《信息存储与检索》、《信息检索》、《化学化工文献与信息检索》、《科技文献检索与利用》、《中外专利信息网络检索与实例》、《信息检索通用教程》等。同一时期两种类型的教材在内容上的区别都不大,主要以介绍文献检索的基本原理和工具的使用方法为主,对信息文化与知识发现的内容涉及不多,这类教材对于那些不熟悉搜索引擎的大学生起到了很好的指导作用。随着搜索引擎的不断普及,大学生已具备了一定的信息检索基础技能,而这类教材,也不能满足大学生的学习需求,采用研究性学习的教学方法进行文献检索、研究与分析的技能培养已迫在眉睫。另一方面,由于计算机网络的普及,国内外已经开始出现信息文化饥渴的现象,以计算机病毒为首的各种信息安全和道德危机,给全世界人民的的生活和工作造成了巨大的损失,也迫切需要在文献检索课程中增加信息文化的内容。

与现有同类教材相比,本教材在内容编排和实践课题的设计两个方面作了较大的改革。不但涉及基本的检索问题,也涉及与大学生密切相关的信息文化,还设计了操作性较强的研究性实践课题,为引导学生带着问题学习,提高研究性学习的能力服务。

第一章的主要内容为认识检索。从搜索引擎的优点说起,以搜索引擎的基本原理介绍检索的基本原理和要素。由于结合了具有相对认知度的搜索引擎,增加了可读性和兴趣性。

第二章的主要内容是信息文化。大学生是未来社会的栋梁,当今社会是信息的社会,对大学生普及信息文化知识具有十分重要的意义。因此,针对日益严重的信息安全与道德危机,在介绍信息素养能力标准语培养方法的基础上,进一步介绍信息安全与信息道德的内容,这在同类教材中并不多见。

第三章的主要内容是检索策略与技巧。重点介绍常用的检索策略与技巧,包括检索策略的制订、关键词的选择、各种算符(包括布尔逻辑算符、截词符、邻近符)的运用、原文的获

取、文献检索结果的保存与利用等知识。

第四章的主要内容是文摘型数据库检索基础。重点介绍 Chemical Abstract(CA, 美国化学文摘)、Engineering Index(EI, 美国工程索引)、Science Citation Index(SCI, 科技引文索引)、PubMed(包括 MEDLINE 在内的美国国立医学图书馆的医学文献检索服务)等文摘型数据库。

第五章的主要内容是全文型检索工具的检索基础。重点介绍全文型检索工具与文摘型检索工具的区别与联系以及部分常用全文型检索工具, 如维普中文科技期刊数据库、EBSCO 数据库, 以及专利、学位论文、会议图书报纸等全文数据库等。

第六章的主要内容是基于文献的知识发现方法。本章包括常用的知识发现方法原理与实践两部分内容, 其中的学位论文文献综述阅读、科研能力评价分析、基于非相关文献的知识发现等内容在同类教材中并不多见, 本章所设计的具有相关操作性的实践性研究性课题对相关教师和学生有参考价值。

第七章的主要内容为文献检索与利用案例。通过具体的案例对文献检索与利用的目的、过程、结果作了详细的介绍。

本教材由吉久明编写教学大纲, 第一、二章由吉久明撰写, 第三、第六章由霍丽萍撰写, 第四章由朱世琴撰写, 第五章由曾媛撰写, 第七章由吉久明、朱世琴、霍丽萍、曾媛共同撰写, 附录由朱世琴、曾媛、吉久明共同撰写。

在教材编写过程中, 编者参阅和引用了许多参考文献, 这为本书的完成提供了帮助, 也丰富了本书的素材, 在此对相关作者表示真诚的谢意。华东理工大学图书馆副馆长孙济庆教授、华东理工大学图书馆文献检索教研室的教师们对本书的编写给予了极大的支持和帮助, 在此也表示感谢。

此外, 本书在编写方面试图贯彻研究性学习方法, 并增加了知识发现方法的介绍, 因为是初次尝试, 疏漏之处在所难免, 敬请使用本书的读者批评指正。

编 者

2009 年 10 月

目 录

第1章 认识检索	001
1.1 初识检索	001
1.1.1 搜索引擎入门	001
1.1.2 检索的含义	004
1.1.3 检索的要素	005
1.1.4 搜索引擎进阶——基本原理	008
1.2 检索进阶(一)	011
1.2.1 信息、情报、知识、文献	011
1.2.2 文献检索工具类型、原理及其要素	017
1.2.3 计算机文献检索工具的评价	025
1.3 搜索引擎进阶(二)——专家检索与学术检索	025
1.3.1 专家检索	026
1.3.2 学术搜索引擎	026
1.3.3 搜索引擎与文献检索工具	030
第2章 信息文化	033
2.1 信息社会与信息文化	033
2.1.1 信息社会	033
2.1.2 信息文化	034
2.1.3 信息素养	035
2.2 信息的评价	041
2.2.1 信息评价的目的和意义	041
2.2.2 信息评价方法	042
2.2.3 信息评价的标准	044
2.3 信息安全	046
2.3.1 信息泄露及其防范	046
2.3.2 信息篡改及其防范	048
2.3.3 信息不可用及其防范	049

2.3.4 信息消失及其防范	050
2.3.5 计算机病毒的防范	051
2.4 知识产权	051
2.4.1 知识产权与知识产权法	052
2.4.2 专利及其应用	054
2.4.3 商标、域名及其应用	062
2.4.4 版权及其应用	065
2.5 竞争情报方法	067
2.5.1 竞争情报的产生和发展	067
2.5.2 竞争情报分析方法	068
第3章 检索策略与技巧	075
3.1 制订检索策略的基本步骤	075
3.1.1 Big6 信息问题解决方案	075
3.1.2 检索语言对检索策略的影响	078
3.2 灵活运用检索技巧	081
3.2.1 单项检索	081
3.2.2 组合检索	081
3.2.3 特殊检索技术	090
3.3 原文获取	096
3.3.1 原文的线索	097
3.3.2 获取全文的途径	102
第4章 文摘型数据库	110
4.1 概述	110
4.2 ISI Web of Knowledge	113
4.2.1 精通 Web of Science	114
4.2.2 INSPEC 数据库(科学文摘数据库)	121
4.2.3 BIOSIS Previews 数据库(美国生物学文摘)	124
4.3 Engineering Village (EI, 美国工程索引)	126
4.3.1 检索规则	127
4.3.2 检索界面	127
4.3.3 个性化服务	131
4.4 Chemical Abstracts(CA, 美国化学文摘)数据库	133
4.4.1 文摘的著录格式	133

4.4.2 检索规则	135
4.4.3 检索界面	136
4.4.4 美国化学文摘数据库小结	151
4.5 PubMed 检索系统	153
4.5.1 精通 PubMed	153
4.5.2 文献记录的显示、存盘、打印	159
第5章 全文数据库	161
5.1 综合类全文数据库	161
5.1.1 万方数据资源系统	161
5.1.2 中国知识资源总库(CNKI)	165
5.1.3 EBSCO 数据库	170
5.2 期刊全文数据库	177
5.2.1 常见的期刊全文数据库	177
5.2.2 维普中文科技期刊数据库	179
5.2.3 Elsevier(ScienceDirect)	186
5.3 专利全文数据库	192
5.3.1 专利文献概况	192
5.3.2 常见国内外专利数据库	192
5.3.3 中国国家知识产权局专利数据库	194
5.3.4 欧洲专利局数据库	198
5.3.5 美国专利数据库	202
5.4 学位论文全文数据库	206
5.4.1 学位论文概况	206
5.4.2 ProQuest Digital Dissertations(PQDD)	206
5.5 会议全文数据库	209
5.5.1 会议文献概况	209
5.5.2 常见会议全文数据库	210
5.6 图书报纸全文数据库	213
5.6.1 常见的国内图书全文数据库	213
5.6.2 常见的国外图书全文数据库	216
5.6.3 报纸全文数据库	218
5.7 标准全文数据库	220
5.7.1 标准文献概况	220
5.7.2 常见标准全文数据库	220

5.7.3 标准文献检索	221
5.8 利用全文文献数据库获取全文	222
第6章 基于文献的知识发现	224
6.1 基于相关文献的知识发现	225
6.1.1 文献计量法	225
6.1.2 内容分析法	227
6.1.3 聚类分析法	228
6.1.4 专利技术分析	230
6.2 基于非相关文献的知识发现	234
6.2.1 非相关文献知识发现理论的理论基础	234
6.2.2 基于非相关文献的知识挖掘	235
6.3 知识发现实践	237
6.3.1 文献计量方法用于发现领域研究趋势	237
6.3.2 基于非相关文献的知识发现实践——Arrowsmith 系统的运用	238
第7章 文献检索与利用案例	244
7.1 生活中的检索	244
7.1.1 氯霉素眼药水	244
7.1.2 食品添加剂: 苯甲酸钠	246
7.1.3 “人肉搜索”	248
7.1.4 终身学习	250
7.2 个人文献管理	251
7.2.1 创建个人 EndNote Web	251
7.2.2 精通 EndNote Web	252
7.2.3 NoteExpress 简介	258
7.3 期刊评价与论文投稿	260
7.3.1 期刊评价	260
7.3.2 利用期刊评价信息, 确定投稿刊物	261
7.3.3 分析出版物发文情况, 确定投稿刊物	264
7.4 利用引文数据库评价学科科研能力	265
7.4.1 中国引文数据库的统计结果	265
7.4.2 利用 SCI 数据库统计华东理工大学论文发表与引用情况结果	267
7.4.3 案例总结	271
7.5 课题调研综述	272

7.5.1 课题综述的特点和内容	273
7.5.2 文献调研及技巧	274
7.5.3 撰写课题综述的注意事项及案例	275
 参考文献	 277
 附录 1 文献检索综合报告	 281
附录 2 研究性学习	293
附录 3 各大类学科综合数据库	297

第1章 认识检索

由于编写教材的需要,某教师急需一本1999年出版的图书,但是图书馆并没有收藏该书,于是他向图书馆的采编部提出了购买该书的建议。图书馆员发现该书已经无法购买了,但却在Google上找到了该书的电子版全文。这位教师非常惊奇,急忙向图书馆员请教。

你是否也有在网上查找资料的经验?

你一般使用哪些网上工具查找资料?

这些工具是如何帮助你找到所需资料的?

1.1 初识检索

本节将在搜索引擎功能介绍的基础上,引入检索的含义与要素。本章将频繁使用“资料”、“信息”、“信息资源”等概念,这些概念对读者来说已经不陌生了,为了学习文献检索,在本章的第二节将继续讨论这些概念。

1.1.1 搜索引擎入门

搜索引擎也叫检索引擎,“搜”与“检”的意思接近,但更显得通俗些。从20世纪90年代至今,搜索引擎在技术、资源种类及数量、提供的服务方面都得到了飞速发展。通过搜索引擎,用户在很短的时间内能找到大部分的生活资料,也能找到学术论文、电子图书等专业资料,还能享受即时通信、免费邮箱、共享空间、知识问答、发表博文等服务,多数在线图书馆都无法与之媲美。



超级链接: OCLC 调查

OCLC^①在2005年开展了《Perceptions of Libraries and Information Resources (2005)》的调查,共收到3300多名来自6个国家OCLC成员单位读者的反馈意见。该调查从图书馆和资源使用情况、到馆或在线访问图书馆情况、对图书馆资源的评价与建议4个方

① OCLC,即联机计算机图书馆中心(Online Computer Library Center)创立于1967年,其总部坐落于美国西部的俄亥俄州。是一个不以营利为目的、提供计算机图书馆服务的会员制研究组织,其宗旨是为广大用户拓展对全世界各种信息的应用以及减低获取信息的成本。它已经发展成今天世界上最大的图书馆合作网络,向全球112个国家和地区的71000所图书馆(其中包括中国大陆的近150所图书馆)提供服务,其中编目服务是OCLC全部服务的核心。自2003年以来,OCLC已经为全世界图书情报机构提供了近十份免费的调查报告,旨在帮助图书情报机构了解信息环境,从而制订更好的服务策略。

面开展了调查。主要获得了以下调查结果:

首先选择搜索引擎查找信息的人占 84%,而首先选择在图书馆的首页上查找资料的只占 1%。20%的大学生和 13%的美国年轻人表示他们对在线图书馆非常熟悉,而非常熟悉或基本熟悉搜索引擎的大学生达 98%,只有极少数的大学生从没有听说过搜索引擎。

这份报告凸显了搜索引擎在大学生的生活中占有重要的地位。

通俗地讲,搜索引擎是一种查找资料的中介平台,用户在这个平台上可以快速地获得符合搜索条件^①的资料概要及资料的来源地址(网址或 URL^②),图 1.1-1 显示以“文献检索精品课程”为搜索词^③在 Google 上的搜索结果的第一页。搜索结果页面由多条记录^④组成,每条记录包含带链接的标题和概要两部分,点击标题可以直接浏览该条信息的页面。

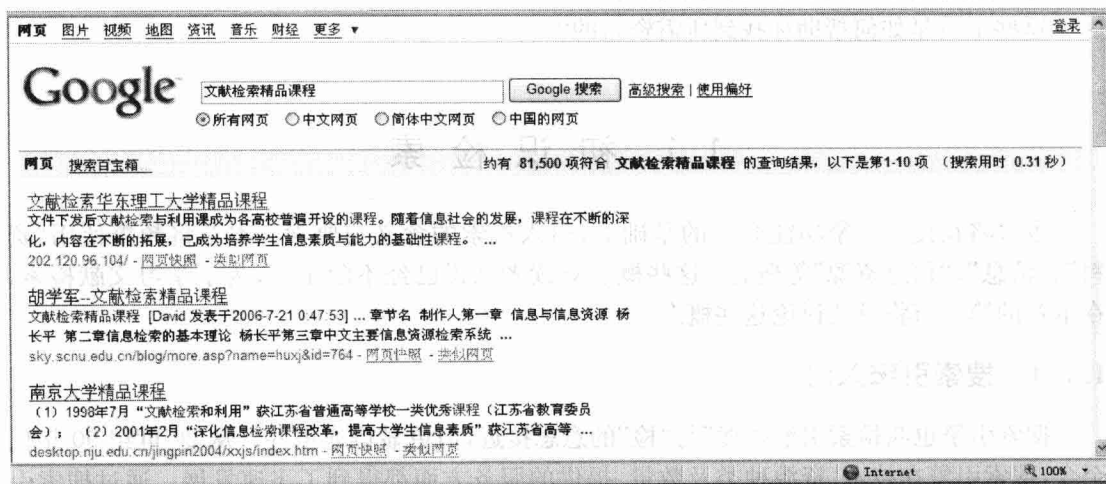


图 1.1-1 在 Google 上搜索标题或概要中含有“文献检索精品课程”的网页信息

点击图 1.1-1 所示的网页上方的“图片”、“视频”、“资讯”、“音乐”、“财经”等链接,则获得标题或概要中含有“文献检索精品课程”的图片、视频、资讯、音乐、财经等信息,点击“更多”右边的倒三角图标,弹出一个菜单,可以对希望获得的信息的文件格式或内容进行进一步的限制,而选择这个菜单的最后一个,“更多”,将显示更多的 Google 产品和服务。而分别选择“所有网页”、“中文网页”、“简体中文网页”、“中国的网页”则显示符合指定语种、地域的网页或文件。

① 搜索条件指在查找信息时要求命中信息必须满足的条件,常以检索式表示搜索条件,检索式的含义及应用详见 1.2.3 节。

② URL(United Resource Link),俗称网址。

③ “搜索词”表示搜索结果中必须包含的信息,也称检索词。检索词的类别和形式可以多种多样,可以是一个或多个包含逻辑关系的字符串、特定的物品名称、地点、人名及其他专有名词,也可以是各种特定的号码,如图书或期刊的国际标准编号、化学物质登记号、专利号、标准号等,还可以是一篇文章中的一个句子、一段话等。

④ “记录”是数据库管理系统的术语之一,每条记录由不同的字段组成,并有一个唯一区别于其他记录的字段。如图书馆借书数据库中,每一种图书或每一位读者的基本“行”信息组成一条记录,其中的书名、作者或姓名、读者证号等“列”信息为字段。信息记录的字段也称为信息的著录项或特征项,人们常常根据实际需要确定信息著录项。

点击“高级搜索”则可以对符合条件的网页做更细致的限定,如图 1.1-2 所示。

图 1.1-2 Google 的高级搜索界面

图 1.1-2 表明,用户可以指定更多的搜索限制条件,如网页的日期、字词的位置、使用权限等;改变每页显示记录的条数最多可达每页 100 项,这一设置只对正在进行的搜索起作用;搜索特定的网页,如与 202.120.96.104/index.html 类似的网页,或与该网页存在链接的网页;搜索开放源代码,如学习 C# 语言时,可以搜索用 C# 编写的开放源代码^①进行学习,在代码搜索界面上选择 C# 语言即可。搜索代码的语言与一般网页的语言不同,高级的搜索方法需要仔细查看 Google 代码搜索相关说明。

点击 Google 首页的“使用偏好”可以改变默认的 Google 首页设置。

当浏览器启用了 Cookie^② 时,可以保存每页显示记录的条数、界面显示语言、是否在新的网页中显示查询结果、是否在搜索框内提供查询建议等设置。

此外,Google 还为网站管理员提供了许多有用的工具,只要在 Google 首页中点击“Google 大全”,就可以进入包含 Google 黑板报和网站管理员中心的页面。

点击图中的“网站管理员中心”,又进入图 1.1-3 所示的页面。

熟悉了搜索引擎 Google 的基本功能、高级功能和服务后,我们来认识检索的含义及要素。

^① 开放源代码(Open Source)是指在一定的知识产权保护条件下可以免费使用的计算机程序源代码,计算机程序人员可以通过源代码达到学习或二次开发的目的。与此对应的还有开放获取(Open Access)文献资源。

^② “由于许多信息检索工具都使用 Cookie, Cookie 的相关知识和设置方法详见用以访问检索工具的浏览器的帮助”。

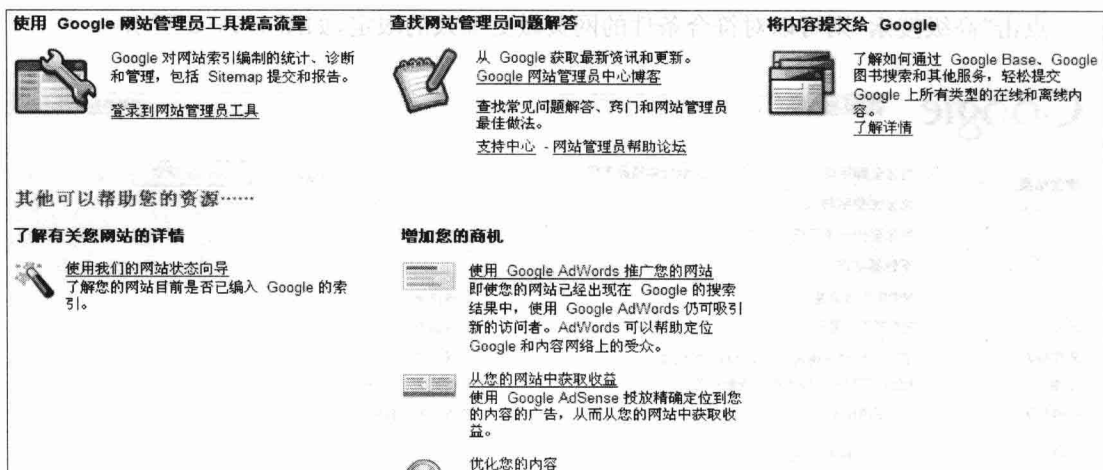


图 1.1-3 Google 网站管理员中心

1.1.2 检索的含义

信息检索就是用户^①(或检索者)以查找满足特定条件的信息线索为目的,使用某种方式(手工、计算机或其他)、借助某种工具(不同载体类型的工具书^②、计算机信息系统等)、查找信息的过程^③,如图 1.1-4 所示。



图 1.1-4 信息检索含义

搜索信息时所借助的工具称为信息检索工具,用户想查找满足特定条件的信息的需求称信息需求,满足用户设定条件的信息称为目标信息。目标信息满足的特定条件称检索条件,一次检索过程可以设定多个检索条件,不同的检索条件之间有逻辑或、逻辑与、逻辑非^④三种关系,每一个检索条件由目标信息的某个特征项及该特征项内容所应包含的检索词两

① 这里的用户是指有权使用检索工具的个体,用户可能是实名的,也可能是匿名的。

② 常见的工具书载体有纸质版、电子版。

③ 有些学者认为信息检索的含义有广义和狭义之分,广义的检索包括了检索工具的基本原理及构建检索工具的步骤。而对于一个信息的终端用户而言,这部分内容应该是透明的,因此对理解检索的含义并不构成分歧。

④ 逻辑关系常用逻辑算符表示,如果两个检索条件之间是逻辑与的关系,表明目标文献必须同时满足两个检索条件。若两个检索条件之间是逻辑或的关系,表明目标文献只要满足两个检索条件之一。若两个检索条件之间是逻辑非关系,表明目标文献只能满足逻辑算符左边的检索条件,且不能满足右边的检索条件。

部分组成,描述检索条件的表达式为检索式。

用户为达成信息需求而设定检索条件、选择合适的检索工具和适当的方式操作检索工具、识别检索工具给出的信息线索^①及依据信息线索找到的信息是否满足信息需求的方法、保存或处理检索工具给出的满足信息需求的信息的方法等统称为检索策略,这些内容中只要有一个不同就会形成不同的检索策略。

综上所述,检索由以下阶段组成:用户认知阶段、用户确立信息需求阶段、用户制订检索策略阶段、在检索工具中实施检索阶段、对检索结果进行处理阶段、分析阶段、利用阶段。尤其要指出的是,这里的用户认知阶段是指用户在一定的知识环境下产生对信息需求的阶段,信息需求与用户的认知能力有密切的关系。

以用户在 Google 上的搜索为例,如果用户在 Google 首页上的搜索条件输入框中,输入“文献检索精品课程”,这里的“文献检索精品课程”就是目标信息必须满足条件的检索词,即要求查找那些在网页标题或网页概要字段中含有“文献检索精品课程”的网页信息的线索,这里的“在网页标题或网页概要字段中含有‘文献检索精品课程’”就是检索条件,检索条件涉及两个字段:网页标题和网页概要,所涉及的检索词就是“文献检索精品课程”。

点击页面上的“Google 搜索”,随即就得到图 1.1-1 所示的页面,页面上的每条记录都包含网页的标题和网页内容概要,其中标题是一个带有该条信息来源地址(URL)的链接。对用户而言,用户利用 Google 页面上的“Google 搜索”这个按钮,得到了他想要信息的线索,并沿着线索找到该条信息。由于是在计算机上,由用户辅助 Google 这一计算机信息系统完成查找信息任务的,因此称这一过程为计算机检索信息或信息检索的基本过程,简称计算机信息检索。相应地,称用户使用纯手工的方式利用某种工具查找信息的过程为手工信息检索。计算机信息检索所使用的工具称为计算机信息检索工具,如 Google 等搜索引擎;手工信息检索所使用的工具称为手工信息检索工具,如常用的新华字典、邮局征订报刊使用的报刊目录索引、邮政编码索引等。

1.1.3 检索的要素

信息检索的含义阐明了信息检索的要素:用户的信息需求、信息资源、检索策略(检索条件、信息检索工具)、检索效果。

1. 用户是实施信息检索的主体

用户的信息需求是产生信息检索的必要条件,用户的信息需求一般通过检索条件来表达。由于用户个体的差异,不同的用户往往具有不同的信息需求,不同的用户对具有相同描述内容的信息需求的理解也会不同,用户对同一检索工具认知的程度也会影响其正确地选择检索工具,从而制订不同的检索策略。

2. 信息资源与用户的信息需求是一种互为依赖的关系

用户产生信息需求是信息资源价值的直接体现,信息资源是用户信息需求得以满足的基本保证。用户通过检索工具发现或获得信息资源,检索工具是在对信息资源进行加工整理的基础上形成的,信息资源和用户的信息需求促使人们开发了检索工具。

^① 如:信息的内容概要,它是以网页还是图像的形式存在的,从哪里可以得到它等。

3. 检索策略帮助用户最终找到所有的目标信息

检索策略是否合理取决于用户对信息需求和信息检索工具的理解和熟悉程度。如：某学生想查找含某药剂的制备方法的外文文献的全文，但他不知道该使用什么样的检索词，也不知道哪些检索工具可以帮助他找到外文文献的全文，他很难在第一时间制订合适的检索策略。

用户确定了合适的检索工具后，选择满足信息需求的检索词和检索字段构成检索条件，在检索工具中设定合适的检索条件后，由检索工具完成查找目标文献的任务。

不同的检索策略会导致不同的效果，在检索策略中起关键作用的是检索条件的设定。

例1 汉语“信息”一词最早出自哪里？给出解决这个问题的检索策略。

解析：解决本例问题首先应选择搜索引擎，使用简单搜索即可，包括百度和 Google 等，比较难的是检索条件如何确定。

如果就事论事，要求包含答案的信息线索中同时含有“信息”、“一词”、“最早”、“出自”4 个词，在 Google 中搜索^①，输入“信息 一词 最早 出自”，在第四条记录中找到了来自“百度知道”的答案；重新输入带单字节双引号的“信息”“一词”“最早”“出自”，在第一条找到了答案同样来自“百度知道”的答案。如果改输入“信息 最早 出现”，则在第一条就找到了答案，但答案来自“爱问知识人”，该条线索与“百度知道”中的线索有所不同。如果改输入带单字节双引号^②的“信息”“最早”“出现”，在前 20 页中找不到答案。可见，在其他内容相同的情况下，检索条件是决定检索需求能否满足的决定性因素。

4. 检索效果是衡量检索是否成功的重要指标

检索效果的好坏与用户的信息需求有直接关系。如果用户希望找到所有符合条件的信息线索，那么当用户找到了所有符合条件的信息线索时，表明其检索效果为 100%。如果用户只希望找到所有符合条件的信息线索中的任意一条，那么当用户找到了一条符合条件的信息线索时，其检索效果也是 100%。但是在检索实践中，我们往往会发现，一方面目前不存在能够穷尽世界上所有信息的检索工具，一般情况下，用户很难找到所有符合条件的信息线索，有时连一条也找不到；另一方面由于在创建各种信息线索时，没有严格统一的标准，人们往往以自身的习惯为准，导致许多符合用户需求的信息线索不能被找到，或者符合用户的检索条件的信息线索并不符合信息需求。为了能够在客观上把握上述两种情况对检索效果的影响，专家们提出了查全率和查准率的概念。

(1) 所谓查全率，即用户找到的符合信息需求的信息线索记录数，占其应该找到的符合信息需求的信息线索记录数的百分比。用字母 a 表示用户找到的符合信息需求的信息线索记录数，字母 c 表示用户应该找到的所有符合信息需求的信息线索记录数^③，则：

$$\text{查全率} = 100 * a / c \quad (1)$$

如果用户找到的符合信息需求的信息线索记录数等于应该找到的符合信息需求的信息

① 搜索时间为 2009-7-18。

② 如果输入的是双字节的双引号，搜索引擎则不把检索词作为词组或短语看待。

③ 这里只讨论去除重复记录的情况。