



普通高等教育“十一五”国家级规划教材

21世纪
计算机
科学与
技术
实践
型
教程

丛书主编
陈明

李树青 陈亚天 主 编
蒋伟伟 郑怀丽 王玉玺 副主编

经济管理信息的 检索与利用



清华大学出版社



普通高等教育“十一五”国家级规划教材

李树青 陈亚天 主 编
蒋伟伟 郑怀丽 王玉玺 副主编

经济管理信息的 检索与利用

纪
计
算
机
科
学
与
技
术
实
践
型
教
程

丛书主编 陈明

清华大学出版社

内 容 简 介

本文主要介绍各种互联网上常见经济管理类信息资源及其检索方法,重点对各种信息检索技能和方法结合实际操作演示进行了详细说明,注重内容的实用性和易读性,并对互联网上免费资源的获取方法专门做了必要的介绍。

本书可作为经济管理类相关专业的本科生和研究生的参考用书,同时也适合对互联网信息检索有兴趣的读者。

本书封面贴有清华大学出版社防伪标签,无标签者不得销售。

版权所有,侵权必究。侵权举报电话:010-62782989 13701121933

图书在版编目(CIP)数据

经济管理信息的检索与利用/李树青,陈亚天主编. —北京:清华大学出版社,2015

21世纪计算机科学与技术实践型教程

ISBN 978-7-302-41356-1

I. ①经… II. ①李… ②陈… III. ①经济管理—情报检索 IV. ①G252.7

中国版本图书馆 CIP 数据核字(2015)第 210726 号

责任编辑:谢 琛 薛 阳

封面设计:常雪影

责任校对:焦丽丽

责任印制:杨 艳

出版发行:清华大学出版社

网 址: <http://www.tup.com.cn>, <http://www.wqbook.com>

地 址:北京清华大学学研大厦 A 座 邮 编:100084

社 总 机:010-62770175 邮 购:010-62786544

投稿与读者服务:010-62776969, c-service@tup.tsinghua.edu.cn

质量反馈:010-62772015, zhiliang@tup.tsinghua.edu.cn

课件下载: <http://www.tup.com.cn>, 010-62795954

印 装 者:三河市少明印务有限公司

经 销:全国新华书店

开 本:185mm×260mm

印 张:24.75

字 数:569千字

版 次:2015年10月第1版

印 次:2015年10月第1次印刷

印 数:1~2000

定 价:44.50元

产品编号:065721-01

《21 世纪计算机科学与技术实践型教程》

编辑委员会

主 任：陈 明

委 员：	毛国君	白中英	叶新铭	刘淑芬	刘书家
	汤 庸	何炎祥	陈永义	罗四维	段友祥
	高维东	郭 禾	姚 琳	崔武子	曹元大
	谢树煜	焦金生	韩江洪		

策划编辑：谢 琛

《21 世纪计算机科学与技术实践型教程》

序

21 世纪影响世界的三大关键技术：以计算机和网络为代表的信息技术；以基因工程为代表的生命科学和生物技术；以纳米技术为代表的新型材料技术。信息技术居三大关键技术之首。国民经济的发展采取信息化带动现代化的方针，要求在所有领域中迅速推广信息技术，导致需要大量的计算机科学与技术领域的优秀人才。

计算机科学与技术的广泛应用是计算机学科发展的原动力，计算机科学是一门应用科学。因此，计算机学科的优秀人才不仅应具有坚实的科学理论基础，而且更重要的是能将理论与实践相结合，并具有解决实际问题的能力。培养计算机科学与技术的优秀人才是社会的需要、国民经济发展的需要。

制订科学的教学计划对于培养计算机科学与技术人才十分重要，而教材的选择是实施教学计划的一个重要组成部分，《21 世纪计算机科学与技术实践型教程》主要考虑了下述两方面。

一方面，高等学校的计算机科学与技术专业的学生，在学习了基本的必修课和部分选修课程之后，立刻进行计算机应用系统的软件和硬件开发与应用尚存在一些困难，而《21 世纪计算机科学与技术实践型教程》就是为了填补这部分空白。将理论与实际联系起来，使学生不仅学会了计算机科学理论，而且也学会了应用这些理论解决实际问题。

另一方面，计算机科学与技术专业的课程内容需要经过实践练习，才能深刻理解和掌握。因此，本套教材增强了实践性、应用性和可理解性，并在体例上做了改进——使用案例说明。

实践型教学占有重要的位置，不仅体现了理论和实践紧密结合的学科特征，而且对于提高学生的综合素质，培养学生的创新精神与实践能力有特殊的作用。因此，研究和撰写实践型教材是必需的，也是十分重要的任务。优秀的教材是保证高水平教学的重要因素，选择水平高、内容新、实践性强的教材可以促进课堂教学质量的快速提升。在教学中，应用实践型教材可以增强学生的认知能力、创新能力、实践能力以及团队协作和交流表达能力。

实践型教材应由教学经验丰富、实际应用经验丰富的教师撰写。此系列教材的作者不但从事多年的计算机教学，而且参加并完成了多项计算机类的科研项目，他们把积累的经验、知识、智慧、素质融于教材中，奉献给计算机科学与技术的教学。

我们在组织本系列教材过程中，虽然经过了详细的思考和讨论，但毕竟是初步的尝试，不完善甚至缺陷不可避免，敬请读者指正。

本系列教材主编 陈明

2005 年 1 月于北京

前 言

本书主要从操作实践的角度介绍各种互联网信息资源的获取方法,尤其是对相关经济管理类信息资源做了重点说明。经济管理类信息资源是一种较为常见和使用频率较高的信息资源,与我们的生活和工作密切相关。掌握好相关的互联网信息检索方法和资源获取方法,具有非常重要的实用意义。

对于内容编排,在编纂本书的过程中,我们考虑了下面三个主要因素:

一是加强对信息检索技能和方法的介绍,不论是互联网搜索引擎,还是各种专业的信息资源数据库工具,读者只有很好地掌握各种常见的信息检索方法,才能更好地利用这些工具来完成各种信息资源的获取。因此,本书在第4章专门结合搜索引擎的检索方法,对基于关键词的基本检索方法和基于Web目录的分类检索方法进行了专门的介绍,并结合典型案例分析对各种常见检索策略进行了详细说明。同时,对于部分特殊的高级检索方法也在后文的特定数据库介绍章节中进行了补充说明。

二是注重内容的实用性和易读性。对于理论性的信息检索内容,本书介绍的并不是很多,相反,对于具体的操作方法,本书不仅非常详细地给出具体的介绍,而且还结合实际截图对关键步骤予以说明,几乎所有的截图都取自2015年三四月间,时效性很强。读者可以自行按照说明和截图演示,对相关内容予以上机实践。

三是强调免费资源的获取方法,例如对如何利用搜索引擎来获取相关信息资源做了大量篇幅的介绍,不仅有专门章节说明,而且在诸如图书、论文和其他各类信息资源介绍时,都对使用搜索引擎来获取的方法做专门介绍,同时也对互联网上的免费信息资源服务做了详细的说明,使得读者可以更为容易地获取到相关信息内容。

本书由李树青、陈亚天任主编,蒋伟伟、郑怀丽、王玉玺任副主编,卢振侠、何光明、郑爱琴、王珊珊、石雅琴、陈珍、陈凤、杨橙、陈莉萍等也参与了本书资料整理和部分章节的校对工作。事实上,本书的写作也得益于互联网所提供的各种信息资源,现代互联网确实能够为我们提供很多非常有用的帮助。同时,本书写作还得到了南京财经大学信息工程学院同仁的大力支持,清华大学出版社也给予了很大的帮助。在此一并表示感谢!

限于个人能力,本书还有很多不足之处,敬请各位读者和专家批评指正。

编 者
2015年4月

目 录

第 1 章 导论	1
1.1 信息资源	1
1.2 信息资源检索	3
1.2.1 信息检索的必要性	4
1.2.2 信息检索的基本过程	8
1.2.3 信息检索效果的衡量指标	10
1.2.4 信息检索系统的发展历史	10
第 2 章 互联网及其信息资源服务	14
2.1 互联网简介	14
2.1.1 国际互联网的发展历史	14
2.1.2 中国互联网的发展历史	17
2.1.3 互联网的基本工作原理	18
2.2 互联网信息资源服务	20
2.2.1 远程登录服务	20
2.2.2 文件传输服务	23
2.2.3 电子邮件服务	25
2.2.4 网络新闻服务	26
2.2.5 名址服务	28
2.2.6 文件索引服务	28
2.2.7 信息浏览服务	29
2.2.8 其他信息服务	35
第 3 章 搜索引擎介绍	38
3.1 搜索引擎的发展	38
3.1.1 国外搜索引擎的发展历史	39
3.1.2 中国搜索引擎的发展历史	65
3.2 搜索引擎的原理与工作机制	73

3.2.1 搜索引擎工作机制	73
3.2.2 查询结果的显示模式和排序依据	76
3.3 特种搜索引擎	80
3.3.1 元搜索引擎	80
3.3.2 FTP 搜索引擎	86
3.3.3 多媒体搜索引擎	88
3.3.4 地图搜索引擎	95
3.3.5 特殊搜索引擎	99
第 4 章 搜索引擎的检索方法	104
4.1 基于关键词的基本检索方法	104
4.1.1 布尔检索	104
4.1.2 词组检索	111
4.1.3 模糊检索	115
4.1.4 字段检索	117
4.2 基于 Web 目录的分类检索方法	130
4.3 检索策略与典型案例分析	134
4.3.1 合理选择检索关键词	134
4.3.2 综合使用各种检索方法	140
4.3.3 间接检索方法	144
4.3.4 其他方法	153
第 5 章 利用搜索引擎进行信息分析和决策	155
5.1 百度的信息分析决策功能	155
5.2 Google 的信息分析决策功能	163
5.3 其他搜索引擎的信息分析决策功能	166
第 6 章 经济管理类网络图书资源的检索	169
6.1 经济管理类网络书目检索	169
6.1.1 图书馆书目系统	170
6.1.2 图书网站的网络书目	178
6.1.3 网络书目数据库	184
6.2 经济管理类电子图书的全文检索与下载	187
6.2.1 全文电子图书数据库	188
6.2.2 利用搜索引擎获取全文电子图书	195
第 7 章 经济管理类网络学术论文资源的检索	198
7.1 经济管理类网络电子学术论文的检索方法	198

7.1.1	电子学术论文的数据库检索	199
7.1.2	搜索引擎的电子论文检索	215
7.1.3	经济管理类主题数据库论文检索	227
7.2	利用引文信息获取相关学术研究资源的基本方法	231
7.2.1	参考文献查询与管理	231
7.2.2	引文索引查询	237
7.2.3	结合参考文献和引文索引获取相关学术研究资源的基本方法	250
第 8 章	经济管理类其他网络文献资源的检索	257
8.1	网络专利文献的检索	257
8.1.1	从专利主管机构检索	258
8.1.2	专利数据库检索	258
8.1.3	搜索引擎中的专利信息检索	264
8.2	网络科技报告检索	272
8.2.1	国内网络科技报告的检索	272
8.2.2	国外网络科技报告的检索	273
8.3	网络标准文献检索	277
8.3.1	国内网络标准文献的检索	278
8.3.2	国际网络标准文献的检索	279
第 9 章	经济管理类网络信息资源的检索	281
9.1	新闻信息检索	281
9.2	名录信息检索	286
9.2.1	企业名录信息检索	286
9.2.2	产品名录信息检索	295
9.2.3	院校和研究机构的检索	304
9.3	商贸信息检索	307
9.3.1	商品供求信息检索	307
9.3.2	交易市场信息检索	314
9.3.3	会展信息检索	316
9.4	价格信息检索	317
9.4.1	中国国家信息中心的价格数据库	317
9.4.2	产品价格信息的服务站点检索	319
9.4.3	专业的产品价格数据库检索	323
9.5	统计信息检索	325
9.5.1	统计数据检索	325
9.5.2	年鉴信息检索	328
9.6	金融信息检索	338

9.6.1 股票信息检索	338
9.6.2 专业金融信息数据库检索	342
9.7 经济分析信息检索	348
9.7.1 国内经济分析信息数据库	348
9.7.2 国外经济分析信息数据库	350
9.8 专业术语检索	352
9.8.1 利用词典站点来检索	352
9.8.2 网络专业百科知识检索	360
9.9 人物信息检索	373
9.9.1 利用人物数据库来检索	373
9.9.2 互联网人物信息的检索方式	377

参考文献	383
------	-----

第 1 章 导 论

本书主要介绍在互联网上检索各种信息资源的方法,其中对经济类信息检索有专门的说明。不过,在正式介绍之前,需要首先了解一些比较基本的概念和原理,以方便后续书中内容的学习,同时也有助于那些对信息检索原理不是很了解的读者,使其掌握一些信息检索的基本理论内容,从而加深对各种实践操作的理解。本章主要介绍信息资源和信息资源检索的一些基本概念。

1.1 信 息 资 源

随着信息技术的发展和信息技术对人类生活生产行为影响力的不断加深,信息资源本身逐渐成为和能源、材料并列的当代世界三大资源之一,而且它的重要性在很多领域甚至逐渐超过了后两者。正如美国哈佛大学研究小组著名的资源三角形观点:“没有物质,什么都不存在;没有能量,什么都不会发生;没有信息,任何事物都没有意义。”

显然,这个现象与人类对信息资源价值的重视密不可分。虽然直到 20 世纪 80 年代信息资源才作为一种独立的资源形式,然而近三十年来,特别是随着互联网的高速发展,信息资源逐渐成为一种对国家发展,对人类生活生产至关重要的战略资源。有意思的是,信息资源的可获取性也逐渐得到提高,因此不论是企业、个人还是国家政府,都在广泛充分地利用信息资源来提高驾驭物质资源和能源资源的能力。

所谓信息资源,是指人类社会活动所产生和涉及的一切文件、资料、图表和数据等信息的总称,它的存在形式包括文字、音像、印刷品、电子信息、数据库等。从广义来看,它是指信息活动中各种要素的总称,既包含信息本身,也包含与信息相关的人员、设备、技术、资金等因素;但是从狭义来看,人们通常所说的信息资源只限于信息本身,是指各种载体和形式的信息的集合。

从信息来源的角度,人们通常把信息资源分为以下四种形式。

1. 体裁信息资源

它指以人体为载体,通过口头语言和身体语言(体态)这些信息交流符号创造和传播并能为他人识别的信息。参与社会信息交流的每个人都是一个独立的信息源。它其实也是最为古老的一种形式,在人类的早期,大多数体现知识的信息都是以语言口授的方式得以保存,如各种远古神话传说等,例如《论语》就是记载孔子言语的一本书。

但是,口语信息资源并没有随着时代的发展而逐渐变得不再重要,相反,在现代社会,

口语信息依然是一种极为重要的信息资源,甚至有学者把这种口语信息资源称之为“零次信息资源”。很多不经正式渠道流通的信息,各种存储在人类大脑中的知识,往往都只是通过口语的形式传播,也通常只保存在人们的脑海中。显然,人们有必要将其收集整理出来,以便利用。就像前段时间,有人曾经组织过对著名老科学院和老电影艺术家的人物采访,据此这些珍贵的口语信息才能保存下来。一些诸如百度知道之类的网络百科全书也正是利用这种口语信息资源来提供信息查询服务的。

2. 实物信息资源

从严格意义上讲,实物并非信息资源,但是一切物质实体蕴含着的丰富信息均可视为实物信息,它给人们提供了充分认识事物的物质条件。依据实物的人工与天然特性又可将实物信息资源分为以自然物质为载体的天然实物信息资源和人工实物为载体的人工实物信息资源。

在信息资源获取活动中,人们往往通过获取实物来间接得到信息,如产品展览会上展出的各类产品,通过了解这些产品,人们可以得到很多关于市场和竞争企业的相关信息,同时也能够了解该产品的一些具体细节信息。对于经济类信息资源而言,实物信息资源及其相应的数据信息都是一些重要的信息资源。

3. 文献信息资源

文献信息资源是用一定的记录手段将系统化的信息内容存储在各类载体上而形成的一类信息资源,这些载体包括印刷型载体(Printed Form)、电子型载体(Electronic Form)、缩微型载体(Micro Form)和声像型载体(Audio-Visual Form)等。文献信息资源是信息资源中的主体部分,也是信息搜集、存储、检索和利用的主要对象,也是本书主要的讨论对象。

它可以按照加工的深度分为如下四种:

一是零次文献信息资源。它是指最原始的文献信息资源形式,虽未公开交流,但它是生成一次文献信息的主要素材。具体形式包括未经记录或者未形成文字材料的口头交谈信息,还有未公开于社会即未经正式发表的原始文献,或没正式出版的各种文献资料。通常获取难度很大,但是在商业和军事领域,它具有特殊的利用价值,也被称为“灰色信息”。

二是一次文献信息资源。它主要是指一些具有原创性的文献信息,如各种论文、专著和新闻等。此类信息价值较大,通常也是人们最终所希望获取的信息内容。常用的一次文献主要包括图书、期刊、会议文献、学位论文、专利文献、标准文献、科技报告、政府出版物、产品样本和产品目录、档案,统称为十大文献信息源。其中,图书、期刊(报纸)被称为普通文献(白色文献);会议文献、学位论文、专利文献、标准文献、科技报告、政府出版物、产品样本和产品目录以及档案八种类型文献被称为特种文献(灰色文献),它是一种介于图书与期刊之间的文献类型,通常在出版发行方面或获取途径方面比较特殊,因而被称为特种文献。

三是二次文献信息资源。它主要是指对大量一次文献进行收集整理后形成的信息资源,如摘要和目录索引等,这些文献信息资源的主要目的是提供人们一种查询一次文献的途径和方法。

四是三次文献信息资源。它主要是指在对二次文献进行整理加工的基础上,按照某一个领域和学科方向编撰的带有综合性的文献信息,如百科全书和词典等。

文献信息资源也可以按照信息内容及其检索方式的不同分为如下三种类型:

一是全文信息资源(Full-text Information Resource)。它就是用户希望获取的最终的信息内容,通常都是一次文献信息内容。随着计算机技术的发展,今天的大多数文献信息资源都可以通过互联网来获取相应的电子全文版本。

二是书目信息资源(Bibliographic Information Resource)。书目是相对于全文而言,也就是说,此类文献信息资源通常都不是人们希望获取的最终信息内容,只是进一步检索的依据和途径,即二次文献信息资源。借助这些书目型信息资源,人们可以更为方便地找到所需的全文信息资源。当然,即便是没有全文信息,此类信息源依然很重要,人们可以据此来了解某一学科的发展趋势和某一机构或者个人的科研能力,它也是很多科研评价的重要参考依据之一。

三是数值信息资源(Numeric Information Resource)。它主要提供各种原始数据资料,以便于学者进行科研工作时使用,如经济统计数据、人口地理数据和产品参数数据等。通常也被称为事实信息资源。

4. 网络信息资源

今天,网络信息资源往往特指以互联网为纽带连接起来的和以互联网为主要交流、传递、存储手段与形式的信息资源。具体是指所有以电子数据形式把文字、图像、声音、动画等多种形式的信息存储在光、磁等非纸介质的载体中,并通过网络通信、计算机或终端等方式再现出来的资源。网络信息资源通过网络将原本相互独立、分布于世界各地的数据库、信息中心、文献中心等连接在一起,形成一个内容与结构全新的信息载体。

相对于传统文献信息源而言,它的信息来源复杂,质量参差不齐。它既包括以网络资源形式存在的文献信息资源内容,也包括网络媒体发布的纯网络信息资源形式,同时还包括诸如大量由互联网用户发布的网页信息内容,特别随着移动设备的广泛使用,实时通信信息资源日益成为一种重要的网络信息资源。它们数量巨大、增长迅速,传播方式也极为快捷,对现代社会的影响力逐渐增大,也成为了人们日常获取信息的主要来源之一。

值得说明的是,本书介绍的信息资源检索方式主要为网络信息资源,其中包括 Web 网站资源、搜索引擎和各个传统文献数据库提供的网站检索站点等。

1.2 信息资源检索

所谓信息资源检索,有时也称为信息资源获取、信息资源查询和信息资源搜索等。它们的意思相差不大,都表示用户利用现代信息检索系统来获取所需信息资源内容的过程。虽然随着信息资源重要性的不断提升,信息资源数量的不断增多,信息资源的可获取性不断增强,人们获取所需信息资源的能力也在增强,但是在很多领域、很多时间,人们对信息资源的获取效果依然不满意,造成这种现象的原因是多方面的,除了技术和资源本身需要改进外,增强信息资源获取意识、加深信息资源形式和种类的了解、提高个人的信息资源

检索能力,这些都能极大地改善这种情况。

1.2.1 信息检索的必要性

很多人都经常说“做功课”。比如旅游出行前,需要对旅游目的地做必要的了解,通过现有的网络搜索引擎和地图服务功能,现代人几乎可以在网上提前完成所有的行程安排及票务预订,了解注意事项。特别随着移动设备的广泛使用,即使是在出行中,遇到问题时,移动端的信息资源检索服务依然可以实时地提供大量有价值的资讯服务。表 1.1 给出了常见旅行安排所需的信息检索网站及其功能。

表 1.1 常见旅行安排所需的信息检索网站及其功能

旅行安排	国内常用网络信息资源站点	国外常用网络信息资源站点
行程了解、旅游攻略	百度	Google
机票预订	携程、同程	Priceline、Expedia、Orbitz
旅店预订	携程、艺龙	Booking、Airbnb
交通服务	滴滴打车	Uber
餐饮	大众点评网	Yelp

再如在生活和工作中遇到问题怎么办?过去似乎手边总是需要一本诸如“百事通”之类的手册,但今天只需打开互联网,查询一下即可。如有同事使用惠普打印机出了问题,老是在液晶屏上显示一个 Printer Mispick 的提示,找了半天说明手册。其实直接在百度上输入 Printer Mispick,就可以解决问题了,如图 1.1 所示。



图 1.1 在百度中查询 Printer Mispick 的结果页面(截取于 2015-3)

在现代社会,对于个人而言,提高信息检索意识是一种重要的基本技能。有人称之为“搜商”(Search Quotient, SQ)^①,可以把它看成是一种与智商、情商并列的人类智力因素,也就是人类通过某种手段获取新知识的能力,其本质就是查询信息和搜索信息的能力。

如果把信息检索的必要性再说大一些,人生的几件大事可能都与信息检索有关!

比如高中毕业选择高校和专业,仅仅查阅那个小小的高校专业介绍显然不够。看看大家对这些高校和专业的关注程度和相关网络信息,才可以更好地帮助我们做出选择。如百度提供的高校搜索风云榜就按照关注度对这些高校进行了排序,如图 1.2 所示。



图 1.2 百度提供的高校搜索风云榜(截取自 2015-3)

当然如果能知道专门提供此类高校专业信息的站点,则可以获取更为准确的参考信息,如教育部学位与研究生教育发展中心主办的“中国学位与研究生教育信息网”就公布有年度中国大学的学科排名信息,如图 1.3 所示。

再如就业找工作,其实就是就业信息检索。甚至连找对象这种事情也都成为现代互联网信息检索服务产业中一个很大的市场,如百合等各种婚恋介绍站点,更不要说买房买车之类的事情了。

对于企业而言,在与经济有关的各个领域,信息检索服务及其利用形式更是无处不在。比如利用搜索引擎进行广告推广已经成为一种常见的市场营销策略,用户只需在搜索引擎中输入一些查询词,搜索引擎就会把相应的广告有效地推送给用户浏览。甚至争夺自己网站在著名搜索引擎检索结果的排名位置,也成为一个专门的行当,被称为“搜索引擎优化”(Search Engine Optimization, SEO)。当然很多搜索引擎自己也开始直接在检

① “搜商”概念最早由中搜的总裁陈沛提出。

2012年全国高校学科评估结果

0201 理论经济学

本一级学科中，全国具有“博士一级”授权的高校共37所，本次有27所参加；还有部分具有“博士二级”授权的高校也参加了评估；参加评估的高校共52所，以下按得分高低的高校依次排列（得分相同者并列）。

学校名称	学科评估得分
中国人民大学	91
北京大学	87
武汉大学	86
南开大学	84
复旦大学	84
南京大学	81
厦门大学	79
吉林大学	78
上海交通大学	78
清华大学	78
北京师范大学	76
中山大学	76
四川大学	73

2012年科研院所学科评估结果

本次共有28个科研院所（含宽口径）的44个学科也参加了评估，科研院所的得分按与高校相同的方法计算得出，其学科整体水平得分如下。

一级学科名称	高校和科研院所学科评估得分	科研院所名称	学科整体水平得分
0101 哲学	10002	中国科学院	91
0201 理论经济学	10001	中国社会科学院	87
0202 应用经济学	10486	中国社会科学院	86
0301 法学	10005	中国社会科学院	84
0302 政治学	10240	中国社会科学院	84
0303 社会学	10353	中国社会科学院	81
0401 教育学	10384	中国社会科学院	81
0402 心理学	10536	中国社会科学院	79
0403 体育学	10183	中国社会科学院	78
0501 中国语言文学	10272	中国社会科学院	78
0502 外国语言文学	10697	中国社会科学院	76
0503 新闻传播学	10027	中国社会科学院	76
0601 数学	10140	中国社会科学院	73
0602 物理学	10651	中国社会科学院	73
0603 化学	10003	中国社会科学院	73
0604 生物学	10422	中国社会科学院	73
0605 医学	10487	中国社会科学院	73
0606 农学	10520	中国社会科学院	73
0607 工学	10419	中国社会科学院	73

图 1.3 中国学位与研究生教育信息网公布的高校专业排名信息(截取于 2015-3)

索结果中呈现商品的具体信息,无须用户进一步单击链接去了解价格等商品信息,如图 1.4 所示。



图 1.4 在百度中查询“自行车”的结果页面(截取于 2015-3)

互联网电子商务在这几年得到了大力发展,诸如淘宝之类的网络购物站点也如雨后春笋般陆续出现并取得巨大成功。其实,网络购物的关键就在信息检索,这也是影响用户使用感受的一个最为明显的因素。如果这种系统不能很好地帮助用户找到自己所需的商品,恐怕用户就不愿意使用它们了。为此,淘宝在自己主页的显著位置上放置了一个检索框,同时也在各个商品的浏览页面中集成了各种方便用户的检索功能,甚至还提供了高级检索功能,并要求用户对此提出意见,如图 1.5 所示。



图 1.5 淘宝商品的“高级检索”界面(截取于 2015-3)

从总体来看,今天互联网上的信息量已经呈现出一种爆炸性增长的态势。据报道,由中国互联网信息中心(CNNIC)发布的《第 35 次中国互联网络发展状况统计报告》显示,截至 2014 年 12 月,我国域名总数为 2060 万个,中国网站总数为 335 万,年增长 4.6%,而国际出口带宽为 4 118 663 Mb/s^①,年增长 20.9%^②。面对着这个海量的信息资源,用户使用网络信息检索的能力现状如何呢?事实上,人们难以有效地获取所需知识。主要原因在于这种信息资源的增长速度远远超出了人们能够处理它们的能力。约翰·奈斯比特(John Naisbitt)在《大趋势》一书中是这样形容人们目前所处的困境:“信息是丰富的,而我们正在渴求知识(Rich Data But Poor Information)。”^③

搜狗实验室在 2007 年曾经发表过一篇研究论文^④。文中指出,在对搜狗搜索引擎一

① b/s 是指每秒传输位数(Bits Per Second),它是衡量网络带宽的重要指标,今天人们使用的网络带宽普遍可以达到 100 Mb/s 左右。

② 第 35 次中国互联网络发展状况统计报告。http://www.cnnic.net.cn/hlwzfzyj/hlwxyzbg/201502/P020150203551802054676.pdf. 2015-3.

③ [美]J. 奈斯比特. 大趋势[M]. 北京: 新华出版社, 1998.

④ 余慧佳, 刘奕群, 张敏, 茹立云, 马少平. 基于大规模日志分析的搜索引擎用户行为分析[J]. 中文信息学报, 2007, 21(1).