



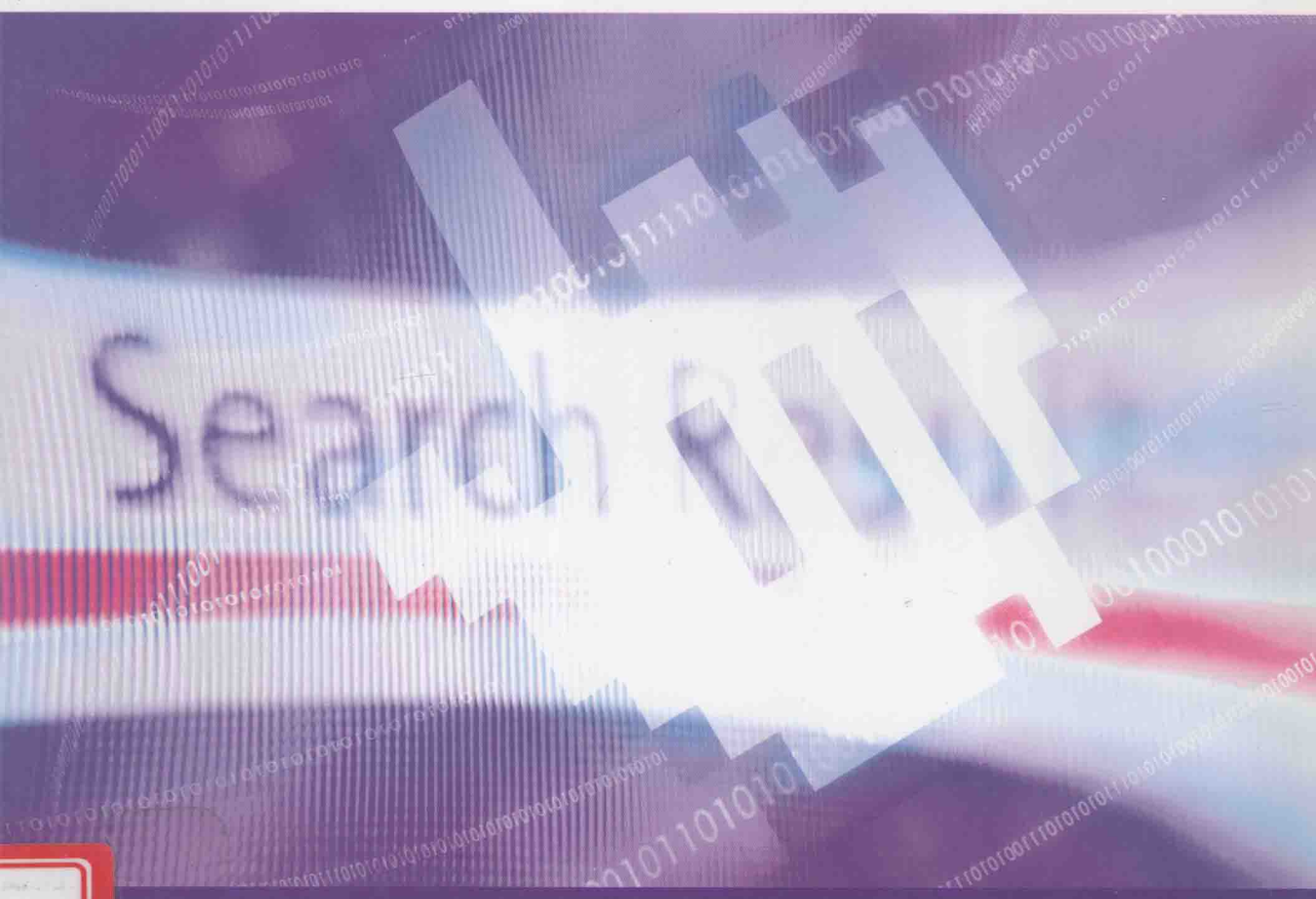
GAODENG XUEXIAO ZHUANYE JIAOCAI

• 高等学校专业教材 •

信息检索与利用教程

xinxi jiansuo yu liyong jiaocheng

彭奇志 主编



中国轻工业出版社

ZHONGGUO QINGGONGYE CHUBANSHE

高等学校专业教材

信息检索与利用教程

彭奇志	喻 萍	彭奇志	主编
张 群	江惜春	沈艳红	编著
		严而清	
		张逸新	主审

 中国轻工业出版社

图书在版编目(CIP)数据

信息检索与利用教程 / 彭奇志主编. —北京 : 中国轻工业出版社, 2009.1

高等学校专业教材

ISBN 978 - 7 - 5019 - 5191 - 8

I. 信 … II. 彭 … III. 情报检索 - 高等学校 - 教材 IV .G252.7

中国版本图书馆 CIP 数据核字(2005)第 145198 号

责任编辑: 林 媛

策划编辑: 林 媛 责任终审: 孟寿萱 封面设计: 刘 鹏

版式设计: 马金路 责任校对: 李 靖 责任监印: 胡 兵

出版发行: 中国轻工业出版社(北京东长安街6号, 邮编: 100740)

印 刷: 三河市世纪兴源印刷有限公司

经 销: 各地新华书店

版 次: 2009年1月第1版第5次印刷

开 本: 787×1092 1/16 印张: 17

字 数: 428千字

书 号: ISBN 978-7-5019-5191-8

定 价: 30.00元

读者服务部邮购热线电话: 010—65241695 85111729 传真: 85111730

发行电话: 010—85119845 65128898 传真: 85113293

网 址: <http://www.chlip.com.cn>

Email: club@chlip.com.cn

如发现图书残缺请直接与我社读者服务部联系调换

81351J1C105ZBW

序

信息获取能力是现代社会人才素质的基本要素,也是人才竞争优势的重要体现。信息检索作为人类获取信息的主要手段与技术,在人类的知识传播和科学研究中具有承上启下的作用,是人类知识组织的“超链接”。近年来,随着社会政治经济的飞速发展,尤其是互联网技术的应用与发展,信息的增长与传播速度达到了前所未有的高度,也正因为如此,信息检索的作用也更加凸显。

网络的发展与普及在为人类获取信息带来了方便的同时,也为人类信息的获取带来了新的困难。信息资源的不断丰富,给人们在信息的海洋中选择获取信息带来了比以前更大的难度,不仅要检索,还要搜索与挖掘;信息发布的媒介类型的多样化使信息的冗余度越来越高;手工检索技术、数据库检索技术和网络搜索技术等不断涌现与发展为检索者提出了计算机知识、专业知识以及检索知识的更高要求;网络搜索技术的不完善也使得检索结果往往难以满足用户的要求,检索结果集往往非常庞大,以至于用户在浏览结果信息时要花费大量的时间,违背了检索的基本目的;同时,网络信息资源的生命周期越来越短,更新越来越快……这些都使得人们在现代社会中获取能满足自己需求的有用信息越来越困难。网络的发展对人类信息的获取来说是一把双刃剑,其实更多的体现在其负面性方面。

彭奇志先生主编的《信息检索与利用教程》一书,正是为了满足现代社会信息活动不断发展变化的需要,为了提高人们信息获取的知识能力而推出的。现在社会上出版的有关检索(包括手工检索、计算机检索和网络检索)的著作、教材颇不少,其中不乏具有影响力和生命力的著作,对信息检索学科产生过重要作用。彭奇志先生及其团队在本书的构思与写作过程中,集思广益,传统与现代相结合,理论与实践相结合,重点与一般相结合,形成了自身的特色。从内容方面来看,该书对信息资源进行了较好的裁剪与组织,对传统的手工检索工具进行了大量压缩,对数据库与网络检索工具重点进行了突出,有些内容甚至进行了全面的剖析,充实了该书的内容与实际应用价值。同时,该书对检索方法与技术也进行了很好的安排,按照教学的需要逐步地展开这些内容,循序渐进,便于读者学习和阅读。

期盼彭奇志先生出版更多更好的著作。

谢新洲

2006年1月于北京大学

前 言

21 世纪是人类全面进入知识经济和实现社会、经济信息化的时代,信息素质将成为衡量人才素质的重要标准。新世纪所需人才的各种素质与能力都有赖于全面信息素质的培养,包括:信息基本知识理论素养的培养,信息能力(含对信息的搜集获取、加工处理、组织管理、分析评价、思维创新与交流)的提高和信息意识的增强。其中特别是信息意识对人的信息能力具有重要的制约作用。信息检索与利用课是目前高校对学生实行信息教育的惟一途径,传统文献检索课在教学内容、手段、教学管理、教师素质等方面与学生信息能力、信息素质的培养存在着较大的差距,因此必须改革传统的教学,对学生进行全面的信息素质教育,以适应新形势的要求,信息检索与利用教材的创新势在必行。本书是在充分调研现有的教材和高校实际教学的基础上,归纳与总结,创新性地对教材的内容做了适当的编排与构建,具体体现在:

1. 精简手工检索工具教学内容

结合专业特点,对传统手工检索工具进行精选,把那些最常用的、必需的检索工具做了介绍。并注意它们之间内容、特点的对比分析,让学生掌握它们之间的内在联系,真正掌握检索的基本原理,为机检教学打下必要的基础。

2. 以计算机检索内容为主体

大幅增加了计算机检索内容,使本门课程的教学重点从手工检索转移到计算机信息检索。针对现在大多数图书馆引进了大量的电子资源,因此本书重点介绍了电子资源的检索与利用,同时编写了相应的教学实例,有利于学生上机实习,增强学生的学习兴趣。

3. 增加网上信息获取的内容

在网络信息骤增的今天,迅速提高学生获取和利用网上信息的能力已成为当前一个紧迫的任务。本书介绍网络基本知识、基本功能的同时,还以较大的篇幅突出介绍了网络信息的获取、网络信息资源利用与评鉴等。

4. 融手工检索与计算机检索于一体

把手工检索和计算机信息检索的基本理论与方法有机地合二为一,融会贯通,这是本书的特色之一。

5. 增加了学术研究与信息分析利用的内容

过去的文献检索教材,虽然提到了文献的利用,但绝大多数缺乏完整的信息利用方面的内容。信息的分析研究与综合利用是信息检索的最终目的之一,也是信息能力培养的核心。因此,本书在编写信息检索技能与技巧的基础上,增加了信息分析与利用、课题选题知识、学术论文撰写与投稿、期刊评价体系等内容,旨在培养学生能够用科学研究的方法,从现象到本质地进行创造性思维,并经过分析结合,得到创新的、智能化的增值性文献信息产品,这也是信息能力培养的根本所在。

本书的编写力求内容新颖、资料准确、实用性强、结构清晰、文字简练。在介绍综合性信息检索的基础上,重点阐述了人文社会科学信息检索。主要内容包括:信息检索基础理论、检索基本方法、工具书及其网络资源、Internet 搜索工具、综合性检索系统、中文专业性检索系统、外文专业性检索系统、特种文献信息检索、专题信息资源检索、学术研究与信息利用等。

各章节的编写工作具体分工为：第1章、第2章由彭奇志编写，第3章由严而清编写，第4章由喻萍编写，第5章由江惜春、张群、喻萍编写(其中江惜春编写5.1~5.3、5.6~5.7，张群编写5.4，喻萍编写5.5)，第6章由张群、沈艳红编写(其中张群编写6.1~6.8，沈艳红编写6.9)，第7章由沈艳红编写，第8章由彭奇志、江惜春、严而清编写(其中彭奇志编写8.1~8.2，严而清编写8.3，江惜春编写8.4~8.5)，第9章由喻萍编写，第10章由彭奇志编写。本书的大纲、统稿和修改由彭奇志负责，喻萍、沈艳红参与了部分章节的讨论与修改。全书由江南大学图书馆馆长张逸新教授审定。

本书得到了滕颖、谢东、王红等老师的关心和帮助，北京大学谢新洲教授为本书作序，在此表示衷心的感谢。囿于作者的学识和水平，书中难免有缺漏及不当之处，恳请读者不吝指正。

彭奇志

2006年1月

目 录

第1章 信息检索基础理论	(1)	2.3.2 检索标识的确定	(34)
1.1 新经济与信息化	(1)	2.4 检索方法	(34)
1.1.1 基于信息与知识的经济	(1)	2.4.1 工具法	(34)
1.1.2 基于网络的信息化发展历程	(2)	2.4.2 引文法	(35)
1.2 信息与知识	(4)	2.4.3 循环法	(35)
1.2.1 信息(Information)	(4)	2.5 计算机检索技术	(36)
1.2.2 知识(Knowledge)	(5)	2.5.1 布尔逻辑算符	(36)
1.2.3 情报(Intelligence)	(6)	2.5.2 位置算符	(37)
1.3 信息资源	(6)	2.5.3 截词检索	(38)
1.3.1 信息资源的概念	(6)	2.5.4 短语或词组检索	(38)
1.3.2 信息资源的类型	(6)	2.5.5 优先算符	(38)
1.4 信息检索	(12)	2.5.6 限制检索	(38)
1.4.1 信息检索的概念	(12)	2.5.7 精确与模糊检索	(39)
1.4.2 信息检索的类型	(12)	2.5.8 加权检索	(39)
1.5 信息检索的变革与意义	(13)	2.6 检索策略	(40)
1.5.1 信息检索的变革	(13)	2.6.1 手工检索策略	(40)
1.5.2 信息检索的意义	(14)	2.6.2 计算机检索策略	(42)
1.6 信息检索原理	(15)	2.7 检索效果的评价	(44)
1.7 检索语言	(16)	2.7.1 检索效果的评价指标	(44)
1.7.1 检索语言的概念	(16)	2.7.2 查全率和查准率的互逆关系	(46)
1.7.2 检索语言的类型	(17)	2.7.3 影响查全率和查准率的因素	(46)
1.7.3 主题语言	(18)	2.7.4 提高检索效果的措施	(47)
1.7.4 分类语言	(21)	第3章 工具书及其网络资源	(49)
第2章 信息检索基本方法	(26)	3.1 工具书概述	(49)
2.1 检索工具	(26)	3.1.1 工具书的特点	(49)
2.1.1 检索工具的定义与类型	(26)	3.1.2 工具书的评价和挑选	(50)
2.1.2 手工检索工具	(26)	3.1.3 工具书的排检方法	(50)
2.1.3 机械检索工具	(27)	3.2 工具书的手工检索	(51)
2.1.4 数据库	(28)	3.2.1 目录	(51)
2.1.5 计算机检索系统	(30)	3.2.2 索引	(53)
2.2 检索途径	(32)	3.2.3 文摘	(54)
2.2.1 内容特征途径	(32)	3.2.4 百科全书	(56)
2.2.2 外表特征途径	(33)	3.2.5 字典、词典	(58)
2.3 检索标识	(33)	3.2.6 手册	(59)
2.3.1 检索标识的定义与类型	(33)	3.2.7 年鉴	(61)

3.2.8 名录、指南	(62)	5.6 OCLC 联机检索系统	(112)
3.3 网上工具书	(64)	5.6.1 FirstSearch 概述	(113)
第4章 Internet 搜索工具	(69)	5.6.2 FirstSearch 检索方法	(115)
4.1 Internet 概述	(69)	5.7 剑桥科学文摘(CSA)	
4.2 WWW 信息搜索的基本方式	(70)	数据库	(118)
4.3 网站目录及搜索引擎		5.7.1 概况	(118)
工作原理	(70)	5.7.2 CSA 检索方法	(118)
4.3.1 网站目录工作原理	(70)	第6章 中文专业性检索系统	(122)
4.3.2 搜索引擎的工作原理	(71)	6.1 中国资讯行	(122)
4.4 国内外典型搜索工具	(72)	6.1.1 概述	(122)
4.4.1 Google	(72)	6.1.2 检索方法	(123)
4.4.2 百度(Baidu)	(74)	6.1.3 检索结果的处理	(125)
4.4.3 搜狐(SOHU)	(76)	6.2 人大复印报刊资料	(125)
4.4.4 雅虎(YAHOO!)	(77)	6.2.1 概述	(125)
4.4.5 新浪(Sina)	(78)	6.2.2 全文数据库的检索	(126)
4.4.6 Excite	(80)	6.3 法律数据库	(128)
4.4.7 元搜索引擎 Metacrawler	(81)	6.3.1 中国法律信息网	(128)
4.4.8 学术专题搜索引擎 Scirus	(82)	6.3.2 北大法律信息网	(129)
第5章 综合性检索系统	(85)	6.4 全国报刊索引	(132)
5.1 万方数据资源系统	(85)	6.4.1 印刷型《全国报刊索引》	(132)
5.1.1 概述	(85)	6.4.2 全国报刊索引数据库	(133)
5.1.2 子系统及检索方法	(85)	6.5 国研网	(136)
5.2 CNKI 中国知网	(88)	6.5.1 概述	(136)
5.2.1 概述	(88)	6.5.2 国研网的主要栏目	(136)
5.2.2 数据库类型	(88)	6.5.3 国研网的检索方法	(138)
5.2.3 检索方法	(90)	6.6 中国经济信息网	(139)
5.3 维普资讯网	(91)	6.6.1 概述	(139)
5.3.1 数据库类型	(91)	6.6.2 中经网的检索	(140)
5.3.2 检索方法	(92)	6.7 CALIS 数据库	(142)
5.4 Dialog 国际联机检索系统	(95)	6.7.1 概述	(142)
5.4.1 概述	(95)	6.7.2 CALIS 的数据库	(142)
5.4.2 Dialog 系统的特点	(96)	6.8 NSTL 数据库	(144)
5.4.3 Dialog 提供的几个常用功能	(96)	6.8.1 概述	(144)
5.4.4 Dialog 系统的数据库	(97)	6.8.2 数据库及其检索方法	(145)
5.4.5 Dialog 系统的检索	(98)	6.8.3 NSTL 提供的其他服务	(147)
5.5 ISI Web of Knowledge	(104)	6.9 中文电子图书	(147)
5.5.1 概述	(104)	6.9.1 概述	(147)
5.5.2 印刷版引文索引检索方法	(106)	6.9.2 超星数字图书馆	(148)
5.5.3 引文索引检索方法举例	(108)	6.9.3 书生之家	(149)
5.5.4 Web of Science 检索	(110)	6.9.4 方正 Apabi 数字图书	(149)

第7章 外文专业性检索系统	(151)	8.1.4 中国专利信息检索	(177)
7.1 Elsevier Science 全文		8.1.5 欧洲专利局 esp@cenet 网站	(180)
电子期刊	(151)	8.1.6 美国专利信息检索	(183)
7.1.1 概述	(151)	8.1.7 IPDL 日本专利信息检索	(185)
7.1.2 SDOS 检索方法	(151)	8.1.8 其他专利信息检索系统	(187)
7.2 Kluwer 电子期刊	(153)	8.2 研究报告	(187)
7.2.1 概述	(153)	8.2.1 概述	(187)
7.2.2 检索方法	(154)	8.2.2 我国研究报告及检索	(188)
7.2.3 检索技术	(155)	8.2.3 国外研究报告及检索	(189)
7.3 Springer Link 电子期刊	(155)	8.3 学位论文的检索	(192)
7.3.1 概述	(155)	8.3.1 概述	(192)
7.3.2 功能说明	(156)	8.3.2 国内学位论文的检索	(192)
7.3.3 检索方式	(157)	8.3.3 国外学位论文的检索	(193)
7.4 Ebscohost	(159)	8.4 会议文献的检索	(197)
7.4.1 概述	(159)	8.4.1 概述	(197)
7.4.2 数据库检索	(159)	8.4.2 国内会议文献的检索	(198)
7.5 Wiley InterScience	(160)	8.4.3 国外会议文献的检索	(199)
7.5.1 概述	(160)	8.5 标准文献	(204)
7.5.2 检索方法	(161)	8.5.1 概述	(204)
7.6 Blackwell Synergy	(161)	8.5.2 标准文献的检索	(205)
7.6.1 概述	(161)	第9章 专题信息资源检索	(210)
7.6.2 检索功能	(162)	9.1 经贸类专题信息检索	(210)
7.7 Gale 数据库	(162)	9.1.1 国内经贸类信息检索	(210)
7.7.1 概述	(162)	9.1.2 国外经贸类信息检索	(212)
7.7.2 Gale 信息资源检索	(163)	9.2 法学类专题信息检索	(215)
7.8 Ingenta	(165)	9.2.1 国内法学信息检索	(215)
7.8.1 概述	(165)	9.2.2 国外法学信息检索	(218)
7.8.2 检索方法	(165)	9.3 语言文学专题信息检索	(221)
7.9 法律资源全文数据库	(166)	9.3.1 国内语言文学类信息检索	(222)
7.9.1 UMI 法律全文数据库	(166)	9.3.2 国外语言文学类信息检索	(224)
7.9.2 Westlaw	(166)	9.4 艺术类专题信息检索	(227)
7.9.3 LexisNexis	(167)	9.4.1 艺术类图书检索	(228)
7.10 国外电子图书	(167)	9.4.2 艺术、建筑类论文与	
7.10.1 NetLibrary	(167)	资料检索	(228)
7.10.2 ebrary	(168)	9.4.3 艺术类网上信息资源检索	(229)
第8章 特种文献信息检索	(170)	9.5 其他专题信息检索	(231)
8.1 专利信息检索	(170)	9.5.1 教育信息资源	(231)
8.1.1 专利知识概述	(170)	9.5.2 政府信息资源	(232)
8.1.2 国际专利分类法	(174)	9.5.3 求职信息资源	(233)
8.1.3 专利文献及专利信息检索	(175)	第10章 信息分析与学术研究	(235)

10.1 信息分析	(235)	10.3.4 学位论文的写作要求	(244)
10.1.1 信息分析的概念与功能	(235)	10.3.5 学位论文写作的基本步骤	(247)
10.1.2 信息分析的类型	(235)	10.4 科学论文投稿	(249)
10.1.3 信息分析方法的来源	(236)	10.4.1 论文写作与投稿存在	
10.1.4 文献信息分析方法	(237)	的问题	(249)
10.2 课题选题与学术		10.4.2 论文投稿之忌	(251)
文献利用	(238)	10.4.3 投稿实用技巧	(251)
10.2.1 选题策略	(238)	10.5 期刊评价与引文	
10.2.2 研究课题信息的获取	(239)	索引系统	(253)
10.2.3 选题程序	(240)	10.5.1 期刊评价	(253)
10.2.4 提高文献阅读质量是正确选题的		10.5.2 核心期刊的界定	(253)
基本保证	(241)	10.5.3 国内期刊评价与引文	
10.3 学位论文的撰写	(242)	索引系统	(254)
10.3.1 学位论文概述	(242)	10.5.4 国外期刊评价与引文	
10.3.2 学位论文的特点	(243)	索引系统	(255)
10.3.3 学位论文的基本格式	(244)	参考文献	(258)

第 1 章 信息检索基础理论

1.1 新经济与信息化

1.1.1 基于信息与知识的经济

社会发展到今天,人们在能够熟练控制物质和能源资源之后,开始进入到控制只有人类才有能力处理的信息知识的时代。只有人类具有思维、语言和智力,而这都是以信息知识为内容的,是信息技术的革命推进了这一进程。信息与知识在我们的生活和生产中起着不可或缺甚至主导的作用。

1962 年美国经济学家弗里茨·克卢普首次提出知识经济这一概念;20 世纪 70 年代美国丹尼尔·贝尔等人对“后工业社会”中的新经济形态有了进一步的认识;80 年代墨西哥记者阿尔温·托夫勒所著的《第三次浪潮》明确提出与知识经济相近的“后工业经济”;1984 年美国奈斯比特在《大趋势》中深化了信息经济的概念;1986 年美国福莱斯特在《高技术社会》中提出了“高技术经济”;1988 年中国邓小平提出“科学技术是第一生产力”;1990 年联合国研究机构提出“知识经济”的概念,并第一次明确了这种新型经济的性质。

美国经济自 1991 年 3 月回升以来,到 2000 年 2 月,连续 9 年保持 2%~4% 的增长率,创下了历史上最长的繁荣纪录。1999 年 10 月,美国的失业率也降至 30 年来的低水平,只有 4.1%,与此同时消费品通货膨胀率只有 19%,是 34 年来增幅最小的一年。这就是前一段时间讨论和研究颇多的新经济,高增长率、低失业率、低通货膨胀率被认为是它的典型特征。有人认为新经济就是网络经济,强调网络化以后的特点;有人认为新经济就是知识经济,主要强调知识在新经济中的地位和作用。还有人认为新经济就是信息经济,说新经济的主导产业是信息产业;还有数字经济、全球经济的说法。1996 年亚太经济合作组织把美国这种新型经济明确定义为“以知识为基础的经济”(knowledge-based economy,简称为“知识经济”),即建立在知识的占有、配置、生产、使用、消费之上的经济,美国政府也宣布,美国的“新经济”就是“知识经济”。可以这样认为,新经济就是以信息和知识经济为主的新型的经济发展模式,它现在已经处于经济的中心地位,人类活动的每个层面可以由以网络信息技术、高技术为代表的知识信息产业的发展加以改变。

新经济是一种知识经济,知识经济与信息经济有相似之处,这是因为信息与知识概念本身就是一种相近的概念,知识是和人脑的信息处理功能相联系的概念,信息与知识都是人脑处理的对象,而知识是人脑信息处理的结果,是所得到的对客观世界的认识结果。我国对“信息经济”的界定始于 20 世纪 80 年代初,现在一般认为,信息经济是关于信息价值、信息的经济地位,经济信息的收集、处理、储存、控制、组织、综合以及追求信息在生产、科学技术及其他一切社会活动中的经济效益的新型经济形式。知识经济是由信息技术的革命催生的,说新经济是一种信息经济一点也不过分,至于数字经济、网络经济都可以归于信息经济的范畴。

1.1.2 基于网络的信息化发展历程

自 1963 年 1 月日本学者梅棹忠夫在《信息产业论》一文中首次涉及“信息化”这个概念,描绘了信息革命和信息化社会的前景,人类社会正经历着从物质和能量主导的社会向信息社会推进的过程。信息化的含义,一般是指在全社会包括经济、政治、文化、生活等各个方面广泛应用现代信息技术,有效地开发和利用信息与信息资源,促进经济发展和社会进步的过程。

从 1946 年世界第一台通用电子数字计算机 ENIAC 在美国诞生,到 1957 年前苏联成功发射了人类历史上第一颗人造地球卫星,计算机与通信技术的结合,使社会信息化迅猛发展起来。20 世纪 60 年代以来,全球掀起了两次大的信息化浪潮。第一次是以计算机为中心的信息技术应用以及信息产业化、产业信息化的浪潮,其标志是计算机(Computer)、通信(Communication)和内容(Content)三者的结合,可称之为计算机革命和产业革命;第二次是以网络为中心的社会信息化浪潮,其标志是“DIN”即数字化(Digitalization)、信息化(Information)和网络化(Networking),可称之为数字革命和网络革命。美国未来学家保罗·萨福将 Internet 的变革称为计算机领域的第三次浪潮,他认为,第一次浪潮是信息处理;第二次浪潮是信息获取;第三次浪潮是所有计算机联网。

1.1.2.1 全球信息化

1984 年美国国家科学基金会(NSF)决定将 ARPANET、NSFNET、MILNET 等几个网络合并成一个网络——Internet,这个网不仅覆盖美国全国,而且还延伸到世界各大洲。20 世纪 90 年代中期,Internet 以每年翻一番的速度持续增长,成为世界上规模最大、信息资源最丰富的计算机网络。从运行在 Internet 上的主机数看,由 1987 年的 1 万台,经过 1997 年的爆炸性增长,至 1998 年 1 月达到 6 800 万台。至 1996 年 7 月,Internet 已覆盖了全球 173 个国家和地区,连接了 13.4365 万个网络,入网主机总数达 1 288.1 万台。从 Internet 的用户数看,据美国国际数据公司(IDC)估计,1995 年为 830 万,1996 年增加到 3 100 万。据美国努阿 Internet 调查公司的调查结果,截至 1998 年 2 月,全世界的 Internet 用户数达 1.13 亿,2000 年全球使用 Internet 的人数已经突破了 4 亿人。资料显示,2002 年全球因特网用户已经超过 5 亿,站点 2 000 万个,电子商务营业额达到 1 万多亿美元;至 2003 年年底全球因特网用户已经达到 6 亿。

在 Internet 发展的前期,信息传播速率慢、带宽窄,难以支持声音、图像、视频等多媒体信息的传输,也无法满足发达国家信息化增长的需要。在美国副总统戈尔(Gore)的建议下,1993 年 2 月,美国总统克林顿在国会以《国情咨文》的形式发表了“促进美国经济增长的技术——经济发展的新方向”的报告,提出了一项划时代意义的计划“国家信息基础结构:行动纲领”(National Information Infrastructure: Agenda for Action, NII),称为“信息高速公路”计划。报告指出,建设信息高速公路可以带动新学科和交叉学科的发展,形成一大批高新技术群体,增强美国的国际竞争力。戈尔说,信息高速公路的建设将永远改变人们生活、工作和相互沟通的方式,产生比“工业革命”更为深刻的影响。美国政府 1993 年 9 月正式确立这一行动计划,为实施 NII,美国政府签署了总统令,成立了“美国国家信息基础设施顾问委员会”和“信息基础设施特别工作小组”,共同推进 NII 的发展。NII 对全球产生了很大的影响。各国纷纷仿效,制定本国的信息高速公路计划,从而掀起了全球信息高速公路的热潮。

1994 年 3 月,美国副总统戈尔在布宜诺斯艾利斯举行的信息技术联盟(ITU)世界电子通信发展大会上发言,首次提出“全球信息基础设施”(Global Information Infrastructure)概念。GII 由地方、国家和地区的网络组成。作为“网络的网络”,GII 将推动全球信息共享,推动全球

的相互联系和通信,从而创造一个全球的信息大市场。由于要求在众多国家之间进行合作的努力,GII 为所有参与者创造经济与社会利益,包括创造就业机会,推动经济增长,完善基础设施,向顾客提供廉价的服务。

由于用户的急剧增加和各种应用的不断增多,使 Internet 不堪重负,1996 年 10 月 6 日,美国总统克林顿宣布了 NGI(Next Generation Internet),即下一代 Internet 计划。该计划的主要内容是:将美国国家科学基金会(NSF)、国防部(DOO)、能源部(DOE)和国家航空航天局(NASA)支持的现有网络进行扩展和扩容,建设新的宽带、高速、多媒体网,开发新的应用。

1.1.2.2 中国信息化发展历程

中国的信息化建设起步可追溯到 20 世纪 80 年代初期,从国家大力推动电子信息技术应用开始,经历了准备阶段、启动阶段、展开阶段、发展阶段。

1. 准备阶段

1984 年,国务院发出通知指出,为了迎接世界新的技术革命,加速我国四个现代化的建设,必须重点地发展信息产业,应当把电子工业摆到国民经济发展的非常重要的位置上。从 1988 年至 1992 年,国家经济委员会、机电部、国家科委和电子信息技术推广应用办公室,在推动传统产业技术改造、EDI 技术、CAD/CAM 等领域,做了大量工作,不断推动电子信息技术应用向纵深发展。

2. 启动阶段

我国信息化正式起步于 1993 年,国家提出信息化建设的任务,启动了金卡、金桥、金关等重大信息化工程,拉开了国民经济信息化的序幕。同年 12 月,成立了国家经济信息化联席会议,加强统一领导,确立了推进信息化工程实施、以信息化带动产业发展的指导思想。1994 年 5 月成立了国家信息化专家组,作为国家信息化建设的决策参谋机构。1996 年以后,中央和地方都确立了信息化在国民经济和社会发展中的重要地位,信息化在各领域、各地区形成了强劲的发展潮流。国务院于 1996 年 1 月成立了信息化工作领导小组,统一领导和组织协调全国的信息化工作。

3. 展开阶段

国务院信息化工作领导小组确立了国家信息化的定义和国家信息化体系六要素,进一步充实和丰富了我国信息化建设的内涵;提出了信息化建设“统筹规划,国家主导;统一标准,联合建设;互联互通,资源共享”的二十四字指导方针。初步形成了中国电信、中国移动、中国联通、中国网通、中国铁通等多家电信运营公司开展市场竞争的格局。与此同时,会同有关部门,积极推动政府上网工程、企业上网工程和电子商务,在国民经济信息化方面,做了大量工作。十五届五中全会在 2000 年提出要走一条新型的工业化道路,就是以信息化带动工业化,以工业化来支持信息化。

4. 发展阶段

“十五”规划明确提出要把推进国民经济和社会信息化放在优先位置。十六大提出“以信息化带动工业化,以工业化促进信息化,走新型工业化道路”以来,我国积极促进信息技术在国民经济和社会发展中各个领域的应用,在推进电子政务、信息资源开发利用、信息安全及信息化法律法规的建设和促进信息产业发展等诸多方面,取得了重要进展。随着《电子签章法》、《电信条例》的实施,《政府信息公开条例》、《个人数据保护法》等法律法规的研究和制定,信息化发展的环境得到了不断改善。我国信息化建设也有力地促进了信息产业的发展。信息产品的制造业规模已跃居全国各个工业部门首位,信息产业已成为我国的基础产业、支柱产业和先导产业。

国家信息主干网的建设大大加快了我国信息化的步伐。在 Internet 网络方面,我国的发展速度越来越快。根据《第三次中国互联网络信息资源数量调查报告》,截至 2003 年 12 月,全国网民

逼近 8 000 万。全国域名 1 187 380 个,全国网站数约 595 550 个,其中 com 和 com.cn 比例最高。关于网站平均页面访问次数,商业网站最多,其次分别是教育、科研机构及企业网站;网站信息最主要来源是采集,比例高达 88.8%,95.4% 网站提供简体中文阅读,31.3% 提供全站搜索;全国网页总数已经达到 311 864 590 个,相比 2002 年,2003 年网页数量增长了近 98.5%;截至 2003 年 12 月,全国在线数据库总量为 169 867 个,其中企业网站拥有数据库最多,商业网站其次。到了 2005 年 1 月,全国域名数已达 185.23 万个,全国网站数约为 66.89 万个,全国网页总数约为 6.5 亿个,全国网页总字节数约为 20537GB,全国在线数据库约为 30.6 万个。

1.2 信息与知识

1.2.1 信息(Information)

信息是普遍存在的,一切信息来源于自然界,来源于人类社会,人们的生产、生活、科研以及社会活动等都是信息产生的源泉。信息的定义具有多样性,不同的学科,从不同的角度对信息的定义有所不同。

关于信息的定义,有文可考的信息概念已不下百种。据考证,汉语中的“信息”一词,最早见于《三国志》中“正数欲来,信息甚大”的记载。唐朝杜牧在其诗《寄远》中有云:“塞外音书无信息,道旁车马起尘埃”。其中的“信息”指音讯、消息。中国《辞海》(1989 年版)将信息定义为音讯,消息;通讯系统传输的对象。《辞海》中的第一个解释是传统的信息含义,第二个则明显带有现代技术的特征。我国《情报与文献工作词汇 基本术语》(GB/T 4894—1985)中关于信息的定义是:物质存在的一种方式,一般指数据、消息中所包含的意义,可以使消息中所描述的事件的不定性减少。

英文“Information”一词,有情报、资料、消息、报道之意,我国曾译为情报或信息,后经国家有关部门核准,建议该词一律译成信息。在中国台湾、香港地区,该词被译为“资讯”。

美国数学家、信息论创始人 C. E. 香农在其发表的论著《通信的数学理论》中把信息定义为“用来清除随机事件的形式的不定性的东西”,信息就是不定性减少的量,是两次不定性之差。信息的量值与信息的随机性有关,如信息无法预估消息或信号中所蕴含的内容或意义,即预估的可能性越小,信息量就越大。

控制论创始人美国数学家 N. 维纳在《信息控制论》一书中认为信息“不是物质,也不是能量”,是人们在适应外部世界并使这种适应反作用于外部世界过程中,同外部世界进行交换的内容的总称。

心理学家认为,信息不是知识。信息是存在于人们意识之外的东西,它存在于自然界、印刷品、硬盘以及空气之中;知识则存在于人们的大脑之中,它是与不确定性(Uncertainty)相伴而生的,人们一般用知识而不是信息来减少不确定性。

信息和计算机专家认为,信息是数据处理的最终产品,是经过收集、记录、处理,以能检索的形式存储的事实或数据。原始数据中产生信息,信息中产生知识。信息管理专家 F. W. 霍顿给信息下的定义是:信息是按照最终用户决策的需要,经过处理和格式化的数据,处理可以是自动化的或手工的,由数据转化为信息是由信息处理者自己完成的。

英国的《牛津词典》认为:信息就是谈论的事情、新闻和知识。美国《韦氏词典》解释为:信息是通讯的事实,是在观察中得到的数据、新闻和知识。这里“信息”又与“知识”和“数据”联系在一起。日本《广辞苑》把信息释义为:信息是所观察的事实和知识。

综上所述,信息尽管是一个应用十分广泛的概念,但对信息的认识还在探索,还没有一个公认统一的定义。总之,信息是与物质和能量密切相关的,同时又是与物质和能量不同的。参考以上专家学者的研究成果,可以理解为,信息是事物运动的状态和存在方式的反映,信息具有以下主要属性:普遍性、客观性、传递性、依附性、多态性、共享性、时效性、无限性、抽象性。从不同的角度划分,信息可以分为以下几种类型。按信息产生的客体性质来分类,可分为自然信息、生物信息、机器信息和社会信息。从信息的内容划分,可有经济信息、文化信息、科技信息、决策信息、工业信息、军事信息等。从信息与应用过程的时间划分,可分为预测信息、实时信息、反馈信息。按可开发程度分为潜在信息和现实信息。然而,信息不等于知识,“信息”和“知识”是两个不同的概念。

1.2.2 知识(Knowledge)

《辞海》(1989年版)对知识的解释为:人类认识的成果或结晶。依反映对象的深刻性,可分为生活知识和科学知识,依反映层次的系统性,可分为经验知识和理论知识。经验知识是知识的初级形态,系统的科学理论是知识的高级形态。现代认知心理学把知识定义为:个体通过与其环境相互作用后获得的信息及其组织。也就是把知识看成是个体头脑中的一种内部状态。同时认知心理学又根据知识在头脑中表征特点分为两种类型,一类是陈述性知识,另一类是程序性知识。前者主要以命题网络或图式表征,后者以产生式系统表征,这构成了知识本质上的差异。

一般认为,知识是人类在改造客观世界的实践中所获得的认识和经验的总和。人脑通过对客观事物发出的信息的接受、选择和处理,得到对事物一般特征的认识,形成了概念。在反复的实践和认识过程中,人脑通过对相关概念的判断、推理和综合,加深了对事物本质的认识,构成了人们头脑中的知识。

根据经济合作与发展组织(OCED)出版的《以知识为基础的经济》报告,知识又可分为以下几种类型:第一类是“知事”(Know-what),指关于事实方面的知识,这类知识通常被近似地称为信息;第二类是“知因”(Know-why),指自然原理和规律方面的科学理论,这类知识往往在专门研究机构如实验室和大学形成;第三类被称为技术诀窍或专有技术(Know-how),指做某些事情的技艺和能力,许多企业的技术情报和商业秘密被归入这一类知识。获取上述知识的途径有两种:一是直接来源于产生信息的客观事物;二是通过信息载体或媒介。

知识具有以下主要属性:

(1) 实践性 社会实践是一切知识产生的基础和检验知识的标准,科学知识对实践有重大指导作用。

(2) 规律性 人们对实践的认识,是一个无限的过程,人们获得的知识在一定层面上揭示了事物及其运动过程的规律性。

(3) 渗透性 随着知识门类增多,各种知识可以相互渗透,形成许多新的知识门类,形成科学知识的网状结构体系。

(4) 继承性 每一次新知识的产生,既是原有知识的深化与发展,又是更新的知识产生的基础和前提。知识被记录或被物化为劳动产品后,可以世代相传和利用。

知识是信息的一部分。人们不仅能通过信息感知世界、认识世界和改造世界,而且能将获得的信息转变成知识作为认识和改造世界的武器。把信息转化为知识,再把知识转化为智慧,是一种动态过程,是一种开拓过程。反过来,智慧又会转化为新知识,新知识又会转化为新信息,人们通过一定的手段和社会传递过程,借助媒体传给使用者。

1.2.3 情报(Intelligence)

情报是与知识和信息密切相关的概念,人们最早把情报用于军事,作为战时有关敌情的报告。在关于情报概念的讨论中,所提出的情报定义大都与信息 and 知识有关。如情报是“被人们所利用的信息”、“被人们感受并可交流的信息”;情报是“为了解决一个特定问题所需要的知识”、“借助语言、符号在载体上传递的知识”等。

我们认为,情报是特定对象为特定目的而搜集的有使用价值的知识或信息。不同部门、单位或个人所需要的知识或信息,就成为不同部门单位或个人的情报。也就是说,不同的情报对应于不同的用户。例如,科技部门所需要的知识或信息称为科技信息或科技情报;决策部门所需的知识或信息称为决策情报,还有军事情报、经济情报等。信息、知识和情报三者的关系,简单地说,信息包含知识和情报,而知识和情报是交叉的关系,信息、知识和情报在一定条件下可以相互转换。

1.3 信息资源

1.3.1 信息资源的概念

关于信息资源(Information Resources)的定义,目前尚无统一认可的定义。高波、吴慰慈将信息资源定义为:“信息资源是经过人类采集、开发并组织起来的各种媒介信息的有机集合,也就是说信息资源既包括制品型的文献资源,也包括非制品的电子资源”。代根兴、周晓燕则认为信息资源是“人类选取、组织、序化的有用信息的集合”。我国 2004 年 10 月 27 日审议通过的《关于加强信息资源开发利用工作的若干意见》,将“信息资源”定义为有使用价值或者潜在使用价值的各种信息的总称。信息资源与自然资源、人力资源共同构成支撑现代社会发展的资源体系。信息资源的开发利用不是几个部门、几个行业、几个地区的事情,而是涉及政治、经济、文化、军事和全社会各个领域,覆盖现代化建设全局的重大战略。

信息与信息资源两个概念并不能完全等同。信息是普遍存在的,但并非所有的信息都是信息资源。我们认为,只有经过人类加工、有使用价值或者潜在使用价值的信息才可称为信息资源。信息资源具有客观性、传递性、依附性、动态性、共享性、时效性、增长性等特点。

1.3.2 信息资源的类型

信息资源种类繁多,数量巨大。常见的信息资源涵盖书籍、期刊、报纸、只读光盘、软盘、磁带、计算机、因特网、卡通、电视、广播、电影、盒式录音带、录像带、激光唱片等。信息资源的类型划分应该遵循文献信息分类的同一律:在同一个区分等级上,类目划分时必须采用同一个标准,否则,就会造成同位类目外延的交叉现象。信息资源按照不同的划分标准,可以分为不同的类型。

1.3.2.1 按加工程度划分

按加工程度,信息资源可以划分为零次信息资源、一次信息资源、二次信息资源和三次信息资源。

(1) 零次信息资源 指未向社会公开和正式报道,只供一定范围内使用的信息资源。如手稿、私人笔记、会议(发布会、座谈)记录、信件、设计草稿、网络私人聊天等。它具有直接性、及时性、新颖性、随机性、非存储检索等典型特征。零次信息资源在当代信息社会具有愈来愈重要的作用:①获取零次信息资源可以补充记录信息和正规渠道的不足,因为科学技术日新月异

异,新知识、新概念、新术语、新数据层出不穷,而且常常通过非正式渠道以零次信息的形式传播;②在信息系统不健全、信息服务水平不高或信息渠道不畅的情况下,可以通过捕捉零次信息弥补信息的获得;③在市场环境中,零次信息资源占有较大比例,它们反映着市场供求、价格、竞争状态的变化,是市场调查和分析的重要依据;④随着因特网信息的急剧增长,零次信息的传递超越了时空限制,传播量、传播速度和影响面越来越大。

(2) 一次信息资源 指以科学研究、研制成果、生产或产品信息等为基础撰写的,未经他人加工、重组、压缩,通过某种方式(含网络)向社会公开发布或报道的信息资源。如专著、教材、学术期刊论文、科技报告、专利说明书等。一次信息资源一般内容详尽、分散、无序。

(3) 二次信息资源 指将各种分散、无序的一次信息资源,按照一定的规则整理、归纳、提炼和简化,按一定的规律组织而形成的供查找一次信息资源的线索型信息资源。它能有效地提高检索和使用一次信息资源的效率。主要包括目录、题录、文摘、索引类具有标引和检索性质的信息资源。如图书馆馆藏书目、联机目录、专题文摘、引文索引、网络文摘题录型数据库、搜索引擎检索结果等。

(4) 三次信息资源 是在对一次信息资源和二次信息资源进行评价和筛选后,经过分析、提炼和压缩,按知识门类或专题综合加工而成的信息资源,也称为参考工具书。其内容上具有概括性、浓缩性、可检索性和参考性。如百科全书、年鉴、字词典、标准信息、综述和述评等。

1.3.2.2 按信息载体的类型划分

从信息载体的角度,信息资源可分为纸质信息资源和非纸质信息资源两大类。

1. 纸质信息资源

这是一种以纸质材料为载体的信息资源。按其记录方式又可分为:

(1) 印刷型信息资源,包括图书、期刊、报纸等。

(2) 书写型信息资源,包括手稿、日记、信件和原始档案等。

2. 非纸质信息资源

指存贮在非纸质介质载体上的信息资源。按其存贮信息方式、载体材料及使用方法等又可分为:

(1) 缩微型信息资源,包括缩微胶片和缩微胶卷等。

(2) 声像型信息资源,包括唱片、磁带、录音带、录像带、胶卷、胶片和幻灯片等。

(3) 电子信息资源,指以二进制的格式把文字、声音、图像和动画等存贮在磁盘或磁带等载体上,需要通过计算机、网络通信或终端等设备进行阅读的信息资源。包括光盘数据库、网络数据库、联机数据库、电子图书、电子期刊、电子报纸和因特网信息资源等等。

1.3.2.3 按存在状态和开发程度划分

按照存在状态和开发程度,信息资源可分为潜在的信息资源和现实的信息资源。潜在的信息资源是指个人在学习、认知和实践过程中储存在大脑中的信息资源,其特点是只能供个人所用。现实的信息资源则是指潜在信息资源经过个人表述之后能够为他人所利用的信息资源,它们最主要的特征是具有社会性,通过特定的符号表述和传递,可以在特定的社会条件下广泛地连续往复地为人类所利用,是一种无限再生的信息资源。而现实的信息资源又可分为口语信息资源、体语信息资源、实物信息资源、网络信息资源和文献信息资源。

1. 口语信息资源

口语信息资源是人类以口头语言如谈话、演讲、授课、讨论、歌唱等方式表述出来的信息资源。它没有形成文字记录,仅靠口头流传,其特点是传递迅速、互动性强,但容易传失、久传易