

知识技术

曾民族 主编

及其应用

科学技术文献出版社

知识技术

及其应用

清华大学出版社

知识技术及其应用

主 编 曾民族

副主编 梁战平 张振海

乔晓东 陈 豫

科学技术文献出版社

Scientific and Technical Documents Publishing House

北 京

图书在版编目(CIP)数据

知识技术及其应用/曾民族主编. -北京:科学技术文献出版社,2005.11

ISBN 7-5023-5134-5

I. 知… II. 曾… III. 知识学 IV. G302

中国版本图书馆 CIP 数据核字(2005)第 124868 号

出 版 者 科学技术文献出版社

地 址 北京市复兴路 15 号(中央电视台西侧)/100038

图书编务部电话 (010)58882959,(010)58882958(传真)

图书发行部电话 (010)68514035(传真),(010)68514009

邮 购 部 电 话 (010)58882952

网 址 <http://www.stdph.com>

E-mail: stdph@istic.ac.cn

策 划 编 辑 科 文

责 任 编 辑 周 玲 唐 玲

责 任 校 对 唐 炜

责 任 出 版 王芳妮

发 行 者 科学技术文献出版社发行 全国各地新华书店经销

印 刷 者 利森达印务有限公司

版 (印) 次 2005 年 11 月第 1 版第 1 次印刷

开 本 787×960 16 开

字 数 657 千

印 张 36

印 数 1~5000 册

定 价 65.00 元

© 版权所有 违法必究

购买本社图书,凡字迹不清、缺页、倒页、脱页者,本社发行部负责调换。

知识技术及其应用

编 委 会

主 任 邹大挺(中国科技信息研究所,原所长,研究员)

副主任 闫 巍(中国国防科技信息中心,主任,高级工程师)

王明亮(清华同方 CNKI,总经理)

贺德方(万方数据,总裁,研究员)

委 员

中国科技信息研究所:

邹大挺(中国科技信息研究所,原所长,研究员)

梁战平(中国科技信息研究所,原所长,博导,研究员)

武夷山(中国科技信息研究所,总工,研究员)

张满年(中国科技信息研究所,研究员)

中国国防科技信息中心:

闫 巍(中国国防科技信息中心,主任,高级工程师)

胡均平(中国国防科技信息中心,副主任,研究员)

李业惠(中国国防科技信息中心,研究员,博导)

清华同方光盘股份有限公司:

王明亮(清华同方 CNKI,总经理)

刘学东(清华同方 CNKI,副总经理)

徐如镜(交通部科学研究院,研究员)

万方数据股份有限公司:

贺德方(万方数据,总裁,中国科技信息研究所所长,研究员)

蒋勇清(万方数据,常务副总经理,研究员)

乔晓东(万方数据,副总经理,研究员)

张 旭(万方数据,副总经理,研究员)

知识技术及其应用

编辑部

主 编 曾氏族(中国国防科技信息中心,研究员)

副主编

梁战平(中国科技信息研究所,原所长,博导,研究员)

张振海(清华同方 CNKI,副总经理)

乔晓东(万方数据,副总经理,研究员)

陈 豫(中国国防科技信息中心,研究员)

编 委

武夷山(中国科技信息研究所,总工,研究员)

王惠临(中国科技信息研究所,研究员,博导)

张 旭(万方数据,副总经理,研究员)

薛德军(清华同方 CNKI,部门经理,博士)

曾 铮(中国国防科技信息中心)

知识技术及其应用

作 者

序 言 邹大挺(中国科技信息研究所,原所长,研究员)

第一章 曾民族(中国国防科技信息中心,研究员)

第二章 曾民族(中国国防科技信息中心,研究员)

第三章 薛德军(清华同方 CNKI,部门经理,博士)

第四章 秦 健(美国纽约雪城大学信息研究学院,副教授)

第五章 曾 蕾(美国肯特州立大学图书情报学院,教授)

第六章 于 波(清华同方 CNKI,产品经理)

第七章 罗 霄(清华同方 CNKI,项目经理)

第八章 曾 铮(中国国防科技信息中心)

第九章 周晓英(中国人民大学,副教授)

第十章 曾建勳(万方数据,资源总监,研究员)

曾民族(中国国防科技信息中心,研究员)

第十一章 陈 豫(中国国防科技信息中心,研究员)

第十二章 赵俊杰(中国科技信息研究所)

佟贺丰(中国科技信息研究所)

第十三章 王惠临(中国科技信息研究所,研究员,博导)

吴 丹(中国科技信息研究所)

第十四章 张翼燕(中国科技信息研究所)

李佳宁(中国科技信息研究所)

武助宇(中国科技信息研究所)

王惠临(中国科技信息研究所,研究员,博导)

第十五章 张 进(美国威斯康辛大学信息研究系,教授)

第十六章 张丽华(清华同方 CNKI)

第十七章 张 旭(万方数据,副总经理,研究员)

第十八章 乔晓东(万方数据,副总经理,研究员)

张 玲(万方数据)

郭家义(万方数据)

知识技术及其应用

评 审 人

第 一 章 梁战平(中国科技信息研究所,原所长,研究员,博导)

武夷山(中国科技信息研究所,总工,研究员)

第 二 章 曾 蕾(美国肯特州立大学图书情报学院,教授)

王惠临(中国科技信息研究所,研究员,博导)

第 三 章 陈 豫(中国国防科技信息中心,研究员)

第 四 章 王惠临(中国科技信息研究所,研究员,博导)

薛德军(清华同方 CNKI,部门经理,博士)

第 五 章 陈文广(北京大学信息管理系,教授)

沈 英(中国科学院文献情报中心,研究院,博导)

第 六 章 陈 豫(中国国防科技信息中心,研究员)

第 七 章 陈光祚(武汉大学信息管理学院,教授,博导)

第 八 章 沈 英(中国科学院文献情报中心,研究员,博导)

苏新宁(南京大学信息管理系,教授,博导)

第 九 章 曾民族(中国国防科技信息中心,研究员)

荣毅虹(首都师范大学信息管理系,副教授)

第 十 章 李广建(北京师范大学信息管理系,教授)

第十一章 赖茂生(北京大学信息管理系,教授,博导)

- 第十二章 缪其浩(上海图书馆副馆长,研究员)
冷伏海(中国科学院文献情报中心,研究员,博导)
- 第十三章 曾民族(中国国防科技信息中心,研究员)
- 第十四章 陈 豫(中国国防科技信息中心,研究员)
李明镜(微软中国研究院)
- 第十五章 朱东华(北京理工大学,教授)
余红梅(中国科技信息研究所,博士)
- 第十六章 曾建勳(万方数据,资源总监,研究员)
陈光祚(武汉大学信息管理学院,教授,博导)
- 第十七章 梁战平(中国科技信息研究所,原所长,博导,研究员)
陈 峰(中国科学院科技政策与管理科学研究所,教授)
- 第十八章 张智雄
陈 豫(中国国防科技信息中心,研究员)

序 言

随着信息和通信革命的深入发展,信息交流环境、信息系统和信息资源等发生了迅速而深刻的变化。我们又一次站在历史的转折点,重新认识信息情报服务的基本理念、基本功能、基本结构、基本机制和基本环境,重新构筑新的服务模式,把握新的发展方向。凭借科技情报工作者的敏锐性,中国科技信息研究所、总装科技信息研究中心、清华同方光盘有限公司和北京万方数据股份有限公司的专家、学者们,共同组织编写了《知识技术及其应用》一书,以引起业界同行的关注,共同迎接知识时代的新挑战。

知识技术及其应用是一个崭新的前沿领域,在国内外都尚未形成体系,只是在不同领域和不同侧面,从理论上和实践上进行了探索,但是这些探索进展很快,越来越成为信息科学界和情报界的关注焦点。我们的专家、学者们,通过文献调研,抓住一些前沿性、前瞻性、开拓性的研究成果加以梳理、综合和集成,最后确定重点放在知识技术、语义网、知识组织、知识构建、知识网站设计、知识发现、知识服务、协同技术、信息检索、文本内容技术、个性化服务、多媒体内容检索、情报分析、竞争情报系统设计、数字图书馆等主题上。每一个主题都对新知识、新技术进行准确、通俗解读,客观描述最近1~2年内的发展和应

用,指出了关键技术和主攻方向,分析案例,展望未来等。经过近一年时间的努力奋斗,终于完成了《知识技术及其应用》一书的编写任务。

《知识技术及其应用》一书的问世,反映和记录了我国科技情报工作者与时俱进的精神风貌。它的出版为建设和发展 21 世纪我国情报学理论和实践提供了新思路,它的出版将引起业界对前沿技术的关注,进而结合我国实际进行创新活动,实现跨越式发展,它的出版对于广大科技管理干部和高校师生来说,是进行知识更新的一部实用教材,它的出版也是科技情报界迎接我国科技情报事业 50 周年纪念的最好礼物。

知识技术及其应用方兴未艾,本书介绍其进展、技术、观点、理论都不是最终的结果,目的是对业界同行起抛砖引玉的作用,引发大家的思考和讨论,期待大家批评指正。

邹大挺

2005 年 6 月 1 日



科学技术文献出版社方位示意图

(京)新登字 130 号

内 容 简 介

知识技术及其应用是一个崭新的前沿领域,国内外都尚未形成体系,但越来越成为信息科学界和情报界的关注焦点。本书的问世,反映和记录了我国科技情报工作者与时俱进的精神风貌,它的出版为建设和发展 21 世纪我国情报学理论和实践提供了新思路。本书内容涵盖知识技术概况、语义网、知识组织、知识构建、知识网站设计、知识发现、知识服务、信息检索、多媒体内容检索、文本内容技术、个性化服务、情报分析、竞争情报系统设计、数字图书馆等共十八章。

本书适用于情报学专业广大师生及专业人士参考。

科学技术文献出版社是国家科学技术部系统唯一一家中央级综合性科技出版机构,我们所有的努力都是为了使您增长知识和才干。

目 录

第一章	在知识基础设施框架内重构信息服务业	(1)
第二章	把信息转变为知识——知识技术及其研究领域	(22)
第三章	网格技术在知识服务中的应用	(61)
第四章	以知识体系为基础的语义网	(99)
第五章	知识组织系统	(121)
第六章	知识网站的设计	(138)
第七章	知识管理和服务	(170)
第八章	知识发现和挖掘	(228)
第九章	从信息构建到知识构建	(260)
第十章	“泛在知识环境”——“后数字图书馆”的发展目标和研究方向	(297)
第十一章	信息检索技术	(321)
第十二章	情报分析方法和技术	(345)
第十三章	文本内容处理技术	(379)
第十四章	基于内容的多媒体信息检索	(405)
第十五章	信息检索可视化	(431)
第十六章	网络出版与电子期刊	(461)
第十七章	情报技术与竞争情报系统	(504)
第十八章	个性化信息服务	(529)

第一章 在知识基础设施框架 内重构信息服务业

曾民族

前言

从 20 世纪 80 年代开始,现代信息服务以信息资源数字化、信息传递网络化、信息加工计算机化、信息服务系统化为主要特征,全面变革传统信息服务模式,让人们享受信息技术带来的信息文明,成为现今社会信息化的重要标志之一。然而,人们通常对信息技术以打破摩尔定律的超速度发展和更新换代习以为常,却很少注意到现代信息服务的模式也在悄悄地演变。今天现代信息服务又像当年传统图书馆信息服务那样落后于时代的步伐。信息资源数字化的结果,出现了如下局势:人类产生、收集信息的能力超过了人类组织、管理和有效利用信息的能力,人们急切需要的信息通常被实际不需要的信息垃圾淹没掉,也就是奈斯比特所说的“信息淹溺,知识饥荒”^[1]。信息传递网络化并没有彻底解决信息共享的技术障碍,网上仍然存在无数的信息孤岛;信息加工计算机化仍停留在仅能取代人类简单体力劳动的低能水平,绝大部分智能信息处理仍需依靠人工完成;信息服务系统化以系统、设备为中心,颠倒了信息服务业人与系统的主从关系。现有信息服务模式已经落伍,知识服务正在成为信息服务业的主旋律。把信息服务提升为知识服务不但是信息服务自身发展的必由之路,而且具备现实条件。越来越多的信息研究中心、信息服务机构和 IT 有关企业已在着手向知识服务方向重新定位和设计未来。本文通过对科研环境、信息技术、信息资源、信息需求等方面已经发生或正在发生的演变的考察,探讨 21 世纪科技先进国家构建知识服务基础结构的现状,研究如何面向知识服务的特性构建知识基础设施。

1 环境分析

1.1 科研环境新的革命

同 20 世纪 80 年代相比,科技人员从事研究、发展、教学的科技环境在发生革命性变化。首先,科技资源信息化。科技资源主要包括四个部分:人力资源、计算资源、科研仪器设备、科学基础数据和科技信息资源。计算资源网络化,科学家个人可利用由超级计算机组成的计算网格。高级计算不再限于少数研究领域和研究机构(如气象预报、高能物理等),而是普及到各个科学工程领域,如生物、化学、社会科学、环境科学、医学、纳米科学等。目前几乎所有科研仪器设备都具有数字化控制、远程共享、结果可视化等特点,科学家个人可远程操纵和利用大型、昂贵科学仪器设备。依靠各种监测仪器可实时实地采集、网络传输和存储、计算机处理、远程访问海量科学基础数据。科技信息出现数字化图书馆、期刊数字化、网络杂志等出版和流通模式,传统的科技信息资源需要重新定义。其次,科研方式和手段巨大变化。理论/分析和实验/观测两种传统科研方法同计算机仿真/建模相结合,开拓新的可能性和实现新的精密度。科研信息、数据的实时获取与处理和大规模计算成为分析、发现和预测的主要手段之一。科学研究对象不再是简单孤立系统,而是涵盖更大范围、跨学科、整合众多来源的原始数据、使用最先进的数据工具来分析和模拟它们之间错综复杂联系并以更形象的方式使之可视化的复杂系统。科学家现在可跨越传统学科边界开展科研工作,例如环境学家利用气候模型进行研究,物理学家直接利用天文学的观测结果,社会学家就像研究大众行为一样分析科学家之间的交流行为。先进的网络使人、工具、信息联系在一起,多系统之间可消除地理、时间、机构和学科的屏障,实现异地、跨学科、跨系统的协同研究。这些变化导致了已有的研究方法和科学门类的变革。

1.2 信息技术指数增长

信息技术以指数增长速度日新月异。计算机速度每 18~24 个月增加一倍,存储密度每 12~24 个月增加一倍,网络带宽每 9~12 个月增加一倍,信息技术正在整体更新换代。1992—2002 十年间,网络速度从 100 Kpbs 提高为 Gpbs,计算速度从 100 Mflop 提高为 teraflop,存储容量从 100GB 提高为 PetaBytes。1999 年,世界仅两台计算机达到每秒 1