



科技信息工作自动化概论

徐学文 主编



国防工业出版社
National Defense Industry Press

总装部队军事训练“十五”统编教材

科研试验系列

科技信息工作自动化概论

徐学文 主编

国防工业出版社

·北京·

图书在版编目(CIP)数据

科技信息工作自动化概论 / 徐学文主编. —北京:国防工业出版社, 2008. 3

总装部队军事训练“十五”统编教材. 科研试验系列
ISBN 978 - 7 - 118 - 05480 - 4

I. 科... II. 徐... III. 科技情报 - 情报工作 - 自动化 - 教材 IV. G356

中国版本图书馆 CIP 数据核字 (2007) 第 182571 号

※

国防工业出版社出版发行

(北京市海淀区紫竹院南路 23 号 邮政编码 100044)

国防工业出版社印刷厂印刷

新华书店经售

*

开本 880 × 1230 1/32 印张 13½ 字数 382 千字

2008 年 3 月第 1 版第 1 次印刷 印数 1—5000 册 定价 35.00 元

(本书如有印装错误, 我社负责调换)

国防书店: (010)68428422

发行邮购: (010)68414474

发行传真: (010)68411535

发行业务: (010)68472764

第三届总装备部军事训练教材 编辑工作委员会

主任委员 张建启

副主任委员 曹保榆 夏长法 侯贺华 郭文敏

委员 (按姓氏笔画排序)

于俊民 王宜标 王泽民 尤广志

冯 章 朱双华 朱忠刚 刘树军

刘瑞成 安敏建 李方洲 李治三

肖力田 迟宝山 张忠华 张海东

陈永光 胡利民 侯 鹰 姜世忠

聂 峰 倪红星 徐 航 郭 勇

黄伟强 彭华良 裴承新 潘贤伦

秘书长 聂 峰

办公室主任 田 禾

办公室副主任 石根柱 郝 刚

办公室成员 李国华 李立法 郑晓娜

第三届总装备部军事训练教材 科技情报系统编委会

主任委员 余德泉

副主任委员 臧明

委员 (按姓氏笔画排序)

刘林山 苏洪余 李国华

李浩志 陈豫 黄伟强

秘书 石根柱

序

军事训练教材是部队开展军事训练和培养高素质科研试验与管理人才的重要基础。“十五”期间是我军加速武器装备现代化建设的关键时期,随着科学技术不断发展,新武器、新装备大量投入部队使用,急需编写相应的配套教材,来满足部队军事训练和人才培养的需求。为此,总装司令部印发了《总装部队军事训练教材建设“十五”计划》,并组织部分专家、学者编著了这套总装部队军事训练“十五”统编教材。

编著这套总装部队军事训练“十五”统编教材是国防科研试验事业继往开来的大事,也是体现国防科研试验技术水平的一个重要标志。它以新时期军事战略方针为统揽,以军委和总装首长关于加强军事训练工作的一系列重要指示为指导,以《军事训练与考核大纲》、《继续教育科目指南》为依据,坚持科学性、前瞻性和实用性相结合,不断满足军事训练和人才培养对教材的需求,为圆满完成武器装备科研试验和管理保障任务提供了有力的技术支持。

“十五”统编教材共计69部,内容涉及科研试验、陆军装备科研订购、通用装备保障和试验后勤等4个系列的28个系统。这套教材既总结升华了武器装备科研试验和管理保障经验,又反映了国内外最新动态和发展方向,是对国防科研试验工程技术系列教材建设的进一步延续和扩展,是一批高质量的精品教材。其使用对象主要是部队具有大专以上学历的科技人员和管理干部,也可供院校有关专业师生使用或参考。

期望这套教材能够有益于部队高素质人才的培养,有益于武器装备科研试验和管理保障任务的完成,有益于国防科技事业的进步。

总装备部军事训练教材

编辑工作委员会

二〇〇五年十二月

前 言

《科技信息工作自动化概论》是为从事信息处理和服务工作的情报专业技术人员编写的教材。内容包括信息收集、加工和服务中的基础技术及其应用开发,重点介绍信息处理的原理、算法、系统实现的基本概貌以及主要技术进展。本书的选材侧重于当前信息处理的关键和热点技术,以较为成熟和可应用的技术为主,兼顾有发展前景的新兴技术。在章节设计上,分上、中、下三篇,上篇为科技信息工作自动化变革与进展,共两章,分别介绍科技信息体系和科技信息处理技术进展;中篇为技术篇,介绍信息处理自动化的基础技术和算法,共五章,分别介绍信息检索技术、分类与聚类技术、自然语言处理技术、科技信息访问控制技术和科技信息的可信度与适用性;下篇为应用篇,介绍系统实现技术和应用系统,共七章,分别介绍 XML 相关技术、Web Service 技术、多媒体信息处理、信息系统平台建设、知识管理技术、Agent 技术和智能信息服务、个性化信息服务相关技术。

本书是在中国国防科技信息中心多年来的研究和实践的基础上完成的。参加编写工作的同志有该中心的专业技术干部和第一线的技术骨干,也有硕士、博士研究生和特约的外单位专家、教授。各章的具体执笔者为:第1、2章:徐学文,第3章:孙登峰,第4章:罗威,第5章:马红妹,第6章:田丰,第7章:陶国强,第8章:田丰、戴春生,第9章:戴春生,第10章:张洪斌、许志敏,第11章:袁建州,第12章:曾铮,第13章:陶国强,第14章:马红妹。全书由游宏梁、马红妹统稿成书,中国国防科技信息中心的许多同志为本书提供了有益的思路 and 重要参考资料。在成书过程中参考或引用了许多署名和未署名的研究人员的研究成果,是集体智慧和共同劳动的结果。谨此向为本书的写作和出版工作作出贡献的所有

参与者表示深切的感谢。

科技信息处理自动化技术是一个正在迅速发展的技术领域,新方法和新技术层出不穷,要很好地概括其全貌和跟上技术的发展是很困难的,而我们由于水平所限,难免挂一漏万,存在疏漏和不当之处,恳请广大同行和读者不吝赐教,批评指正,以便今后补充完善。

编者

2007年5月

内 容 简 介

本书围绕科技信息工作自动化发展的过程,介绍科技信息工作主要环节上发生的技术变革和出现的新技术。内容包括信息收集、加工和服务中的基础技术及其应用开发,重点介绍信息处理的原理、算法、系统实现的基本概貌以及主要技术进展。选材侧重于当前信息处理的关键和热点技术,以较为成熟和可应用的技术为主,兼顾有发展前景的新兴技术。

本书是一本为从事信息处理和服务工作的情报专业技术人员编写的教材,面向高等院校科技情报和相关领域的本科生、研究生及从事科技信息工作的科技人员。

目 录

上篇 变革与进展

第1章 绪论	3
1.1 概述	3
1.2 文献型和非文献型信息	5
1.3 信息和信息体系	9
1.4 科技信息体系开发	12
1.4.1 建立信息处理的体系结构和技术框架	13
1.4.2 筹划多站点分布式的联合信息处理系统	14
1.4.3 建设集成化、专业化、个性化的信息服务系统	14
1.4.4 加强对网上信息资源描述和管理的研究	15
1.4.5 重视非文献信息处理技术	15
参考文献	15
第2章 科技信息处理技术的进展	17
2.1 概述	17
2.2 信息采集与获取	22
2.2.1 特定领域(Specific Topic)的高效率信息采集	22
2.2.2 Web 语料库(Corpus)	23
2.2.3 Web 预取(Pre - Fetching)	23
2.2.4 网页智能搜索	23
2.3 信息分类与挖掘	24
2.3.1 文本自动分类	24
2.3.2 文本挖掘	26

2.3.3 数据挖掘	27
2.3.4 网络信息挖掘	29
2.4 信息存储与检索	31
2.4.1 独立搜索引擎	31
2.4.2 元搜索引擎	34
2.4.3 检索 Agent	35
2.4.4 搜索引擎的发展	36
2.5 量化信息处理技术	37
2.5.1 概率和数理统计方法	38
2.5.2 神经网络(ANN)和机器学习	38
2.5.3 模糊技术和粗糙集理论	40
2.5.4 进化计算	42
2.5.5 离散傅里叶变换和离散小波变换	43
2.5.6 Agent 和 Multi-Agent 技术	45
2.6 信息服务技术	47
2.6.1 个性化信息服务	47
2.6.2 信息集成服务	50
2.6.3 Web 集成服务模式 and XML	53
2.6.4 应用服务	57
2.6.5 客户关系管理	58
2.6.6 信息服务的智能化技术	58
2.7 科技信息处理应用	64
2.7.1 网络信息挖掘技术应用	64
2.7.2 非文本信息检索	67
2.7.3 信息综合集成	71
参考文献	74

中篇 技术篇

第3章 信息检索技术	79
------------------	----

3.1	概述	79
3.2	基本概念	80
3.2.1	用户任务	80
3.2.2	文档的逻辑视图	81
3.2.3	检索原理	81
3.2.4	检索过程	82
3.3	信息检索模型	83
3.3.1	分类	84
3.3.2	形式化表达	85
3.3.3	经典信息检索模型	85
3.4	索引与搜索	90
3.4.1	倒排文档	90
3.4.2	顺序搜索	93
3.5	查询	95
3.5.1	基于关键词的查询	95
3.5.2	模式匹配	97
3.5.3	查询操作	98
3.6	示例系统	99
3.6.1	中文分词	100
3.6.2	词条权值的计算	101
3.6.3	文档的向量表示	103
3.6.4	用户查询的向量表示	104
3.6.5	相似度计算	104
3.6.6	检索算法	107
	参考文献	108
第4章	分类与聚类技术	109
4.1	概述	109
4.2	分类技术	112
4.2.1	kNN 分类方法	112
4.2.2	决策树方法	113

4.2.3 Bayesian 分类法	114
4.2.4 SVM 分类方法	116
4.2.5 投票分类 (Voted Classification)	118
4.2.6 文本分类的特征选择方法	119
4.2.7 评测分类效果的方法	121
4.3 聚类技术	125
4.3.1 等级聚类技术	126
4.3.2 分割聚类技术	130
4.3.3 聚类结果的修正方法	134
4.3.4 评测聚类效果的方法	135
参考文献	138
第5章 自然语言处理技术	141
5.1 概述	141
5.2 自然语言分析技术	142
5.2.1 词法分析	142
5.2.2 句法分析	147
5.2.3 语义分析	150
5.2.4 世界知识的表示与篇章分析	157
5.3 统计自然语言处理技术	162
5.3.1 语料库语言学简介	162
5.3.2 Shannon 的噪声信道模型	163
5.3.3 统计语言模型的构造	164
5.3.4 基于优先的分析技术	164
5.4 自然语言处理技术的应用	166
5.4.1 自动文摘	166
5.4.2 信息抽取	170
参考文献	174
第6章 科技信息访问控制技术	175
6.1 概述	175
6.2 访问控制	176

6.2.1 访问控制定义	176
6.2.2 访问控制策略	177
6.3 访问控制实现机制	181
6.3.1 访问控制矩阵	182
6.3.2 访问能力表 (CL) 机制	184
6.3.3 访问控制表 (ACL) 机制	184
6.3.4 CL 与 ACL 机制的比较	186
6.4 基于角色的访问控制 (RBAC)	188
6.4.1 RBAC 的概念	188
6.4.2 角色	189
6.4.3 RBAC 的形式化描述	190
6.4.4 RBAC 的发展	191
6.4.5 ANSI RBAC 标准 2004	192
参考文献	196
第7章 科技信息的可信度与适用性	197
7.1 概述	197
7.2 NASA 的信息质量保证机制	199
7.2.1 信息质量的基本标准	199
7.2.2 信息的质量分级管理	201
7.2.3 信息质量评审方法	202
7.2.4 管理机制	203
7.3 校核、验证与确认管理技术体系	203
7.3.1 VV&A 的含义	204
7.3.2 VV&A 的发展现状	204
7.3.3 VV&A 管理技术体系的构成	206
7.4 科学试验信息的 VV&A 过程示例	208
7.4.1 科学试验数据获取通用过程	209
7.4.2 科学试验数据获取过程中的 VV&A 活动	210
参考文献	214

下篇 应用篇

第8章 XML	217
8.1 概述	217
8.2 XML 语言	217
8.2.1 XML 文档概要	217
8.2.2 XML 逻辑结构和物理结构	219
8.2.3 XML 组成元素	221
8.2.4 XML 文档结构	222
8.2.5 XML 名称空间	225
8.2.6 XMLSchema	227
8.3 XML 信息定位	231
8.3.1 XML 信息集合	232
8.3.2 XPath	232
8.3.3 XPointer	233
8.3.4 XLink	234
8.4 XML 的处理	239
8.4.1 XSL 转换	239
8.4.2 DOM	240
8.4.3 SAX	240
参考文献	241
第9章 Web Service	242
9.1 概述	242
9.2 Web Service 架构	243
9.3 Web Service 应用示例	245
9.4 相关的协议和数据类型	248
9.4.1 可使用的3种协议	248
9.4.2 Web Service 支持的数据类型	254
9.5 Web Service 的高级话题	255

9.5.1 关于 SOAP 协议	255
9.5.2 关于 SOAP 头文件	256
9.5.3 控制整理 XML	259
9.5.4 Web Service 的安全性	262
参考文献	263
第 10 章 多媒体信息处理	264
10.1 概述	264
10.1.1 多媒体及多媒体技术概念	264
10.1.2 多媒体数据类型及其主要特征	265
10.1.3 多媒体计算机系统组成	266
10.1.4 多媒体技术的发展和应	267
10.2 多媒体信息处理技术	267
10.2.1 音频信息的处理技术	267
10.2.2 图像信息的处理技术	270
10.2.3 视频信息的获取技术	275
10.2.4 多媒体信息检索技术	276
10.3 流媒体技术及应用	281
10.3.1 流媒体基本概念	282
10.3.2 流媒体技术原理	282
10.3.3 流媒体播放软件系统	288
10.3.4 流媒体技术应用	290
10.3.5 流媒体技术应用方案实例	292
参考文献	296
第 11 章 信息系统平台建设	297
11.1 概述	297
11.1.1 网络平台体系	298
11.1.2 应用服务体系	298
11.1.3 运行管理体系	299
11.2 网络平台体系的建设	299
11.2.1 组建网络平台体系的基本步骤	299

11.2.2	机房环境建设	302
11.2.3	综合布线	305
11.2.4	网络平台设备的选择	308
11.2.5	网络安全子网的建设	318
11.2.6	网络平台设备的连接	320
11.2.7	网络配置	321
11.2.8	网络平台竣工文档	322
11.3	应用服务体系建设	323
11.3.1	数据支持平台的建设	323
11.3.2	应用基础平台的建设	324
11.3.3	应用系统的建设	326
11.3.4	服务体系的建立	330
11.4	运行管理体系建设	331
11.4.1	运行管理的组织与人员	331
11.4.2	运行管理的规章制度	331
11.4.3	网络管理	331
11.4.4	信息管理	332
11.4.5	安全管理	333
11.4.6	应用系统的管理	334
	参考文献	334
第 12 章	知识管理技术	335
12.1	概述	335
12.2	知识管理技术的基本定义	336
12.3	知识管理的核心流程	339
12.3.1	知识获取	339
12.3.2	知识分类与存储	340
12.3.3	知识分享和扩展	341
12.3.4	知识更新	342
12.3.5	知识利用	343
12.4	知识管理技术的分类	344