



数据工程实验指导书

Experiment Manual of Data Engineering

□ 陈刚 郝建东 赵成 刘斌 编著



国防工业出版社

National Defense Industry Press

数据工程实验指导书

陈刚 郝建东 赵成 刘斌 编著

国防工业出版社

· 北京 ·

内 容 简 介

本书是与张宏军教授等编著的《作战仿真数据工程》教材配套的实验指导书。本书根据《作战仿真数据工程》教材的组织特点编排了实验内容。紧密配合理论教学,合理安排数据工程的实验教学,使学生能够在有限的实验课时中,加深对所学知识的理解与掌握,熟练使用常用的数据工程相关处理软件,培养和提高学生的实际操作水平。

本书涵盖了作战仿真数据规划、数据建模、数据集成、元数据集拟制、数据分析与数据挖掘、数据可视化、数据管理等方面的实验项目,实验内容饱满、步骤详细,有助于提高学生的实验效果。

本书可作为相关专业研究生和高年级本科生的实验教材,也可作为工程技术人员的参考书。

图书在版编目(CIP)数据

数据工程实验指导书/陈刚等编著. —北京:国防工业出版社,2015.9
ISBN 978-7-118-10512-4

I. ①数... II. ①陈... III. ①数据管理—实验
IV. ①TP274—33

中国版本图书馆 CIP 数据核字(2015)第 221794 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 100048)

北京嘉恒彩色印刷有限责任公司

新华书店经售

*

开本 787×1092 1/16 印张 13 $\frac{3}{4}$ 字数 317 千字

2015 年 9 月第 1 版第 1 次印刷 印数 1—3500 册 定价 42.00 元

(本书如有印装错误,我社负责调换)

国防书店:(010)88540777

发行邮购:(010)88540776

发行传真:(010)88540755

发行业务:(010)88540717

前 言

随着经济发展、科技创新以及军事技术的不断变革,数据呈现出爆炸式的增长,对数据的管理难度日益加大;同时,人们也逐渐意识到这庞大的数据中蕴藏着重要的科学价值、经济价值、社会价值和军事价值。数据资源的建设发展越来越受到重视,如何有效地管理数据、分析数据、挖掘数据成为人们研究的新方向。由此,产生了一门新兴学科——数据工程。

编者长期从事作战仿真方面的技术研究,并重点围绕该领域的数据建设完成了大量的工作,积累了一些宝贵的工作经验。在院校多年数据工程的课程讲授中,经过反复的实验教学环节的锤炼,形成了许多实验课程的合适素材。为规范实验内容,严格实验训练,达到实验目的,编者一直在对实验教学进行研究,探索在有限的课时条件下,合理组织作战仿真数据工程的实验内容,使学生通过实验教学能够巩固对理论知识的理解,提高动手能力。

本书内容涵盖了作战仿真数据规划、数据建模、数据集成、元数据集拟制、数据分析与数据挖掘、数据可视化、数据管理等方面的实验项目。在一些实验项目中,包含了若干个子实验,在具体实验教学过程中,可以根据课程设计、课时要求和实验条件的不同,选取适合的子实验组合实施。通过本书的引导,希望学生在反复练习中能深刻理解实验内涵,提高独立操作解决问题的能力,并有所发挥创新。

编写过程中,编者将多年积累的实验素材进行了整理和组织安排,并参考引用了《数据分析与数据挖掘》、《数据可视化》、《数据管理新技术》等课程的实验指导书。在此,对所有支持本书编写工作的老师表示衷心感谢。

由于编者水平有限、编写时间紧迫,加之数据工程理论与技术不断发展,书中难免存有错误和不妥之处,敬请专家和广大读者不吝批评指正。编者将进一步完善和充实本书的内容。

编著者

2015年6月于南京

目 录

实验一	数据资源规划	1
1.1	实验目的	1
1.2	实验内容和要求	1
1.3	实验环境	1
1.4	实验报告	1
1.5	实验讲义	1
1.5.1	IRP2000 工具简介	1
1.5.2	数据资源规划案例练习	8
实验二	数据建模	19
2.1	实验目的	19
2.2	实验内容和要求	19
2.3	实验环境	19
2.4	实验报告	19
2.5	实验讲义	19
2.5.1	PowerDesigner 工具简介	19
2.5.2	数据建模案例练习	20
实验三	数据集成	35
3.1	实验目的	35
3.2	实验内容和要求	35
3.3	实验环境	35
3.4	实验报告	35
3.5	实验讲义	35
3.5.1	军事演习数据综合处理平台简介	35
3.5.2	数据集成案例练习	37
实验四	元数据集设计	79
4.1	实验目的	79
4.2	实验内容和要求	79
4.3	实验环境	79
4.4	实验报告	79

4.5	实验讲义	79
4.5.1	XMLSpy 工具简介	79
4.5.2	元数据集设计练习	83
实验五	数据分析与数据挖掘	135
5.1	实验目的	135
5.2	实验内容和要求	135
5.3	实验环境	135
5.4	实验报告	135
5.5	实验讲义	135
5.5.1	SPSS 软件简介	135
5.5.2	数据分析与数据挖掘的设计练习	136
实验六	数据可视化	179
6.1	实验目的	179
6.2	实验内容和要求	179
6.3	实验环境	179
6.4	实验报告	179
6.5	实验讲义	179
6.5.1	Xcelsius 工具简介	179
6.5.2	数据可视化练习	182
实验七	数据管理	198
7.1	实验目的	198
7.2	实验内容和要求	198
7.3	实验环境	198
7.4	实验报告	198
7.5	实验讲义	198
7.5.1	实验 1 关系型数据库 Mysql 的基本操作	198
7.5.2	实验 2 面向对象数据库	202
7.5.3	实验 3 Mongo 常用操作	205
7.5.4	实验 4 图数据库实践	210
参考文献		214

实验一 数据资源规划

实验计划学时:4 学时。

1.1 实验目的

1. 理解基于稳定信息过程的数据规划方法的关键步骤。
2. 掌握数据规划常用工具 IRP2000 的使用方法。
3. 强化学生建立数据工程化建设的思想,培养承担数据工程建设的基本能力。

1.2 实验内容和要求

利用数据资源规划的理论方法,结合学生在大学期间学习生活情况的调查与分析,通过 IRP2000 工具,完成学生信息管理系统的数据资源规划草案。学生学习生活的相关情况分析,参见实验讲义。

1.3 实验环境

1. 硬件:计算机一台,推荐使用 Windows XP 操作系统。
2. 软件:数据资源规划工具 IRP2000,截图软件。

1.4 实验报告

完成本次实验后,需要提交的实验报告主要包括:

1. 利用 IRP2000 规划后的各个阶段的截图,以及相应的文字说明。
2. 整理形成初步的数据资源规划草案。

1.5 实验讲义

1.5.1 IRP2000 工具简介

目前,国内数据规划自动化工具较少,最具有代表性的工具是由大连圣达计算机有限公司开发的 IRP2000(如图 1-1 所示),该工具能够全面支持企业信息资源规划的需求分析与系统建模两个阶段的工作。



图 1-1 IRP2000 主界面

通过 IRP2000 工具,可以进行以下工作:

1. 业务功能分析。支持业务模型的建立,用“职能域——业务过程——业务活动”3层列表描述的业务功能结构,如图 1-2 至图 1-4 所示。

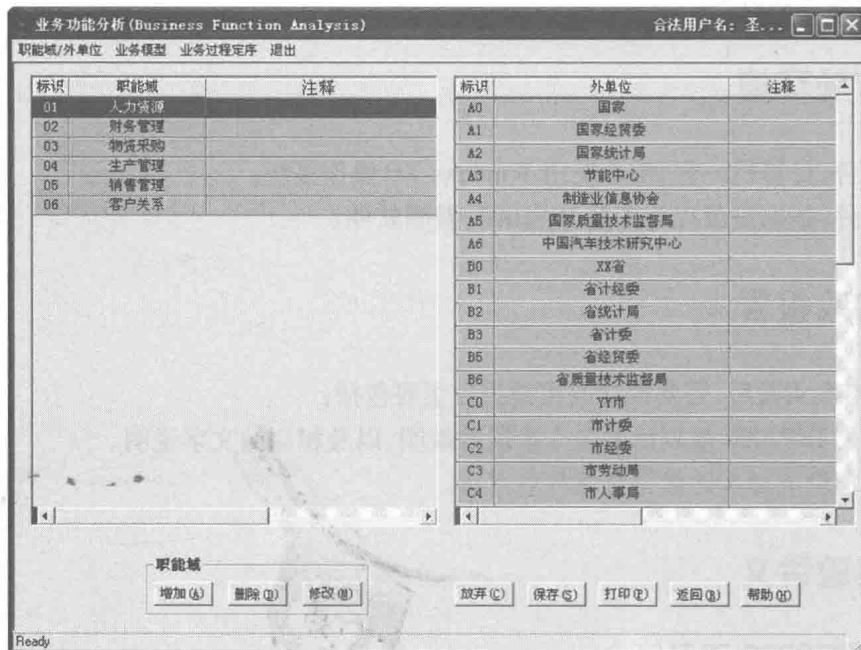


图 1-2 职能域/外单位

2. 业务数据分析。支持用户视图分析、数据元素/数据项的聚类分析和各职能域输入/输出数据流的量化分析,如图 1-5 至图 1-8 所示。

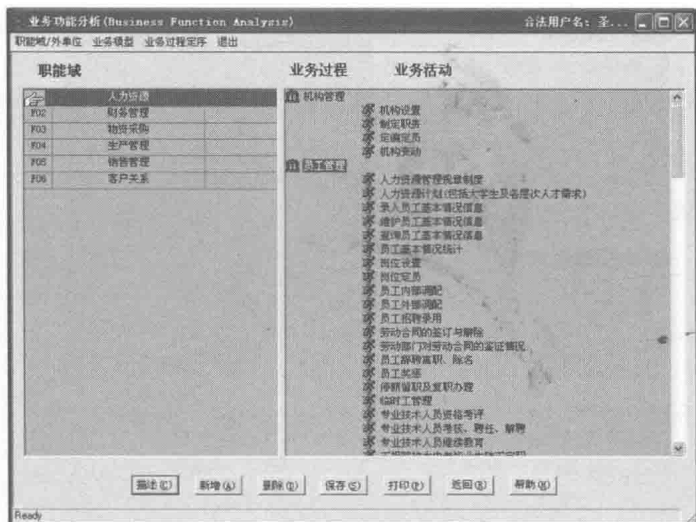


图 1-3 业务模型

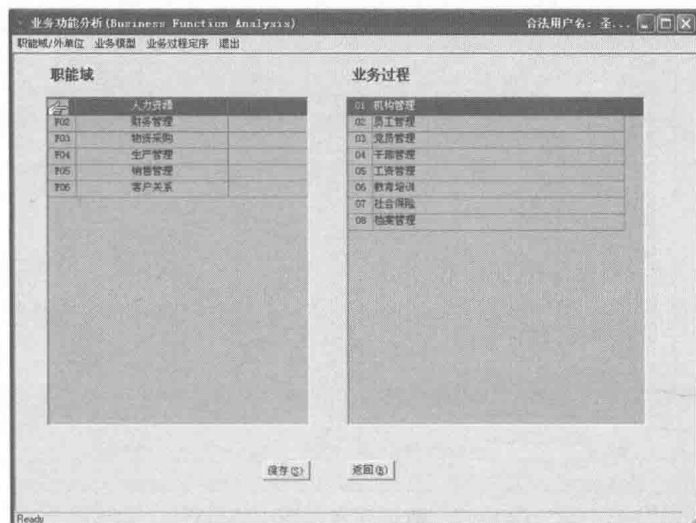


图 1-4 业务过程定序



图 1-5 用户视图

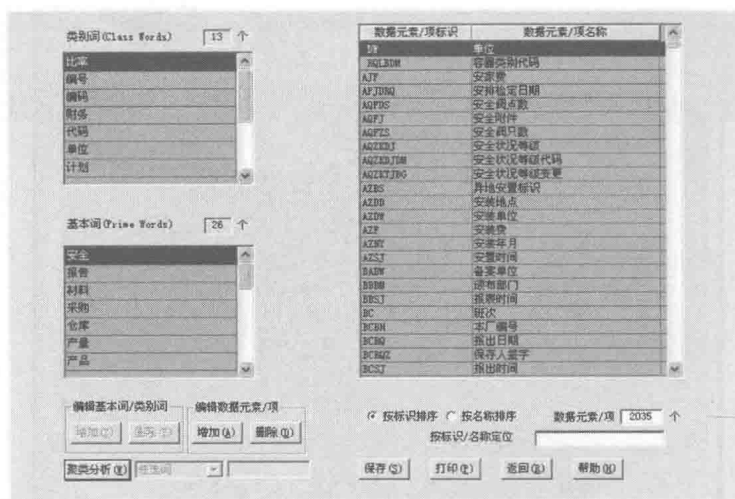


图 1-6 数据元素/数据项



图 1-7 数据项在视图中的分析



图 1-8 数据流

3. 系统功能建模。支持功能模型的建立,用“子系统——功能模块——程序模块”的3层结构来表示系统的逻辑功能模型,如图1-9所示。

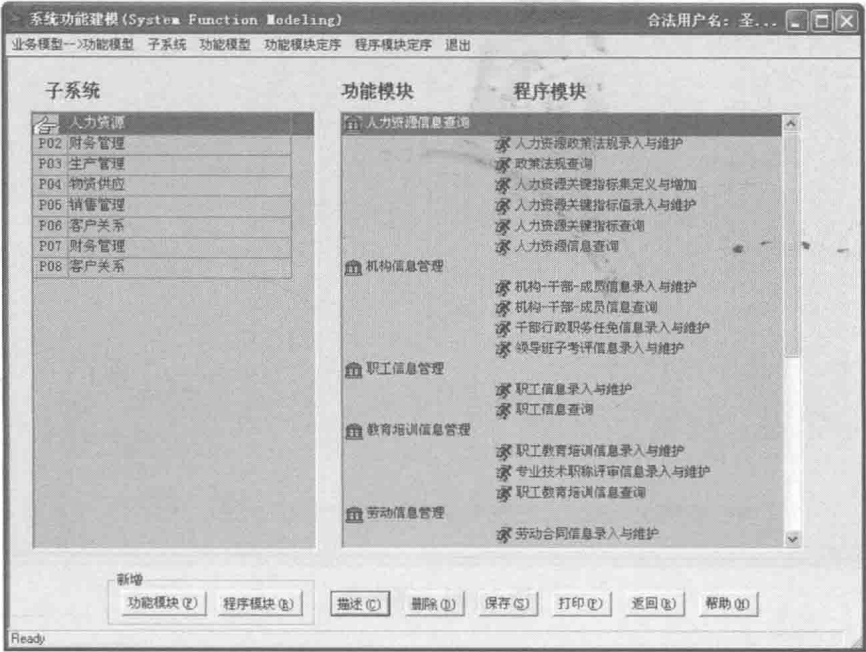


图 1-9 功能模型

4. 系统数据建模。从概念主题数据库的定义开始,支持用户视图分组与基本表定义,落实逻辑主题数据库的所有基本表结构,建立全域和各子系统数据模型,如图1-10至图1-12所示。

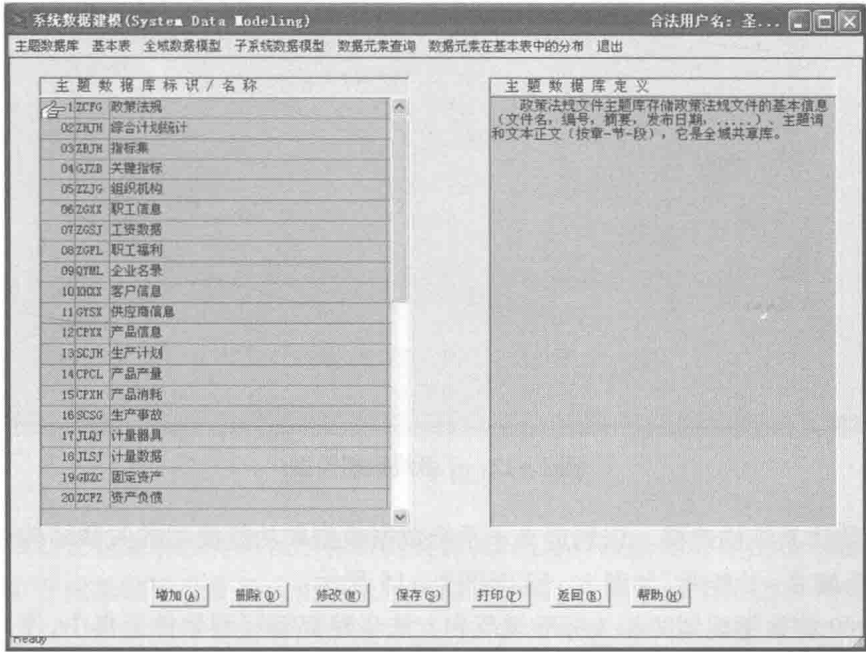


图 1-10 主题数据库

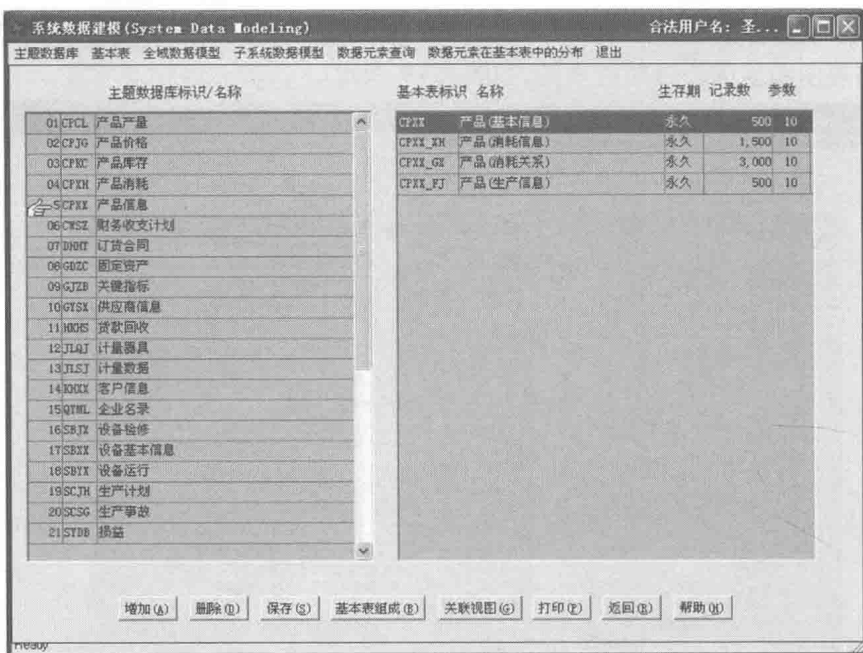


图 1-11 基本表

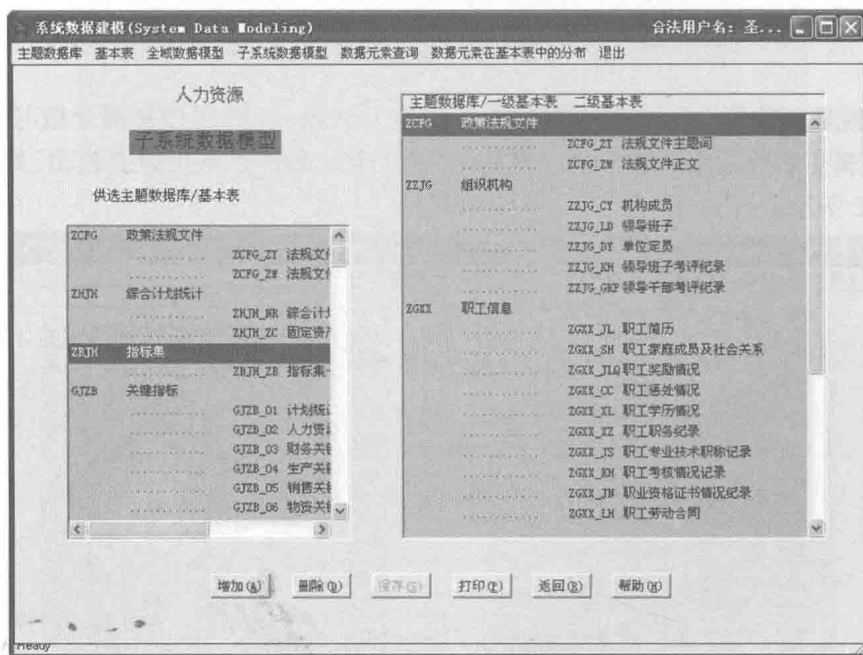


图 1-12 子系统数据模型

5. 系统体系结构建模。识别定义子系统数据模型和功能模型的关联结构,自动生成子系统和全域 C-U 矩阵,如图 1-13 至图 1-15 所示。

IRP2000 将数据规划的相关标准规范和方法步骤都编写到软件工具中,使用可视化、易操作的程序,引导规划人员执行标准规范,使信息资源规划工作的资料录入、人机交互和自动化处理的工作量比例保持在 1:2:7,因而能高质量、高效率地支持数据规划工作。

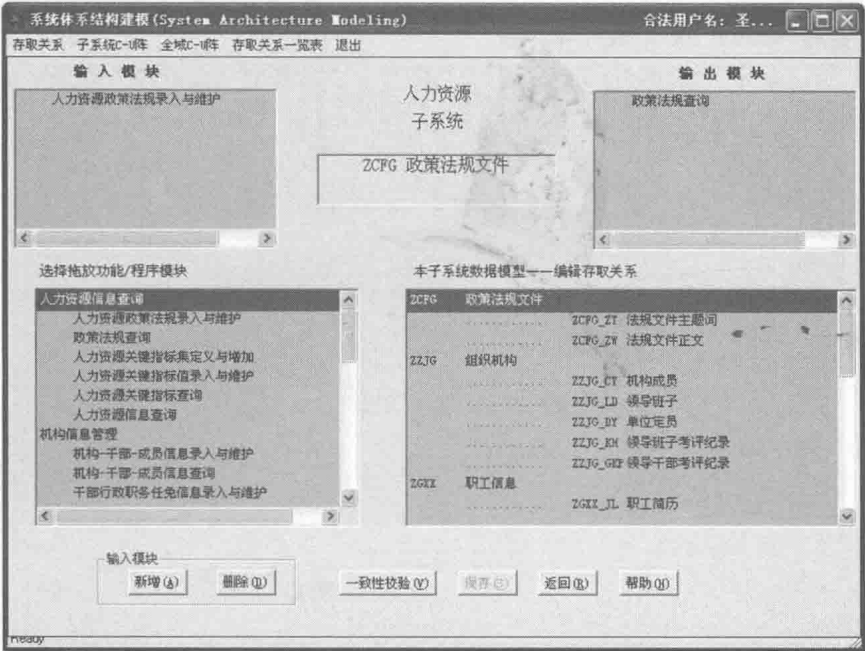


图 1-13 存取关系



图 1-14 子系统 C-U 矩阵

该工具能帮助企业继承已有的程序和数据资源,诊断原有数据环境存在的问题,建立统一的信息资源管理基础标准和集成化信息系统总体模型,在此基础上可以优化提升已有的应用系统,引进、定制或开发新应用系统,高起点、高效率地建立新一代的信息网络。

数据字典是软件工程用来记存应用系统中数据定义、结构和相互关联的概念。随着系统的复杂化和从建设到运行的全程管理的需要,数据字典发展成元库。IRP2000 创建

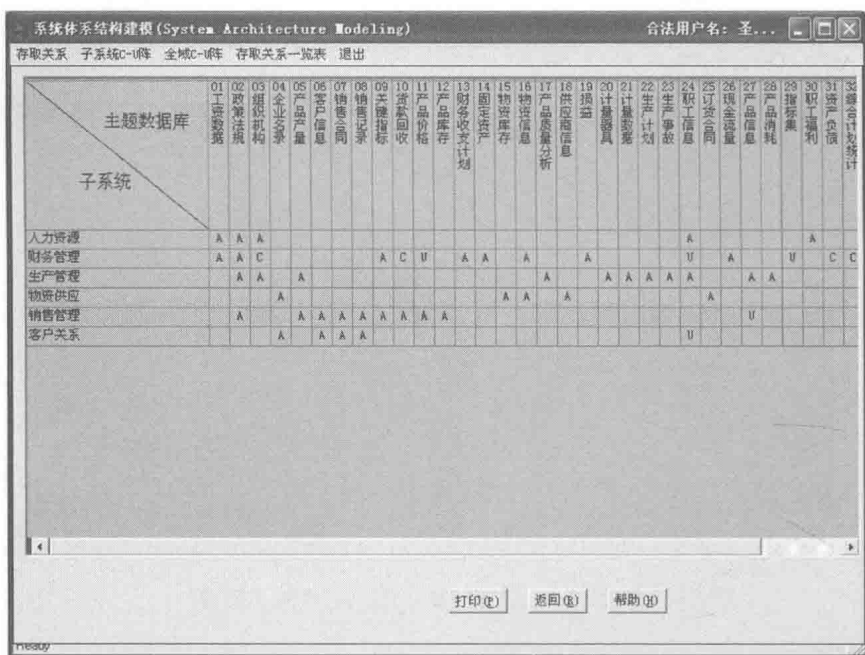


图 1-15 全域 C-U 矩阵

的、贯穿数据规划到应用系统开发全过程的元库,称作信息资源元库。

1.5.2 数据资源规划案例练习

1.5.2.1 学生学习生活情况的调查与分析

经调查发现,学生从入学到毕业在学校所要经历的活动主要包括报到、学习、借还书、购物和毕业等环节。学生持入学通知书到学校报到,学校给学生发放学生证、校园卡和住宿通知书。在平时的学习过程中,学生首先要选课,课程结束后还要参加考试,取得考试成绩。在校期间,学生还会到图书馆借书还书、去食堂吃饭、去学校超市购物等。在校学习结束后,学生申请毕业,经审核后,学校给学生发放毕业证书。由此可以得到顶层数据流程图如图 1-16 所示。

1.5.2.2 学生信息管理系统数据规划步骤

依照学生信息管理系统的数据流程图,以学生的“报到”和“学习”两大活动为例,讲述如何使用 IRP2000 做系统的数据规划,“借还书”“购物充值”和“毕业”三大活动请根据步骤自主完成数据规划。

步骤1 完成业务功能分析

点击主界面“业务功能分析”选项,进入“业务功能分析”界面。点击菜单“职能域/外单位”,点击“增加”按钮,添加职能域“报到”和“学习”,点击“保存”按钮,点击“返回”按钮。

点击菜单“业务模型”,选择职能域“报到”,点击“增加”按钮,添加业务过程“注册”和“安排住宿”。

选择业务过程“注册”,点击“增加”按钮,添加相应的业务活动“学生基本信息录入”“学生基本信息维护”“校园卡办理”“校园卡信息维护”“学生证办理”“学生证信息维

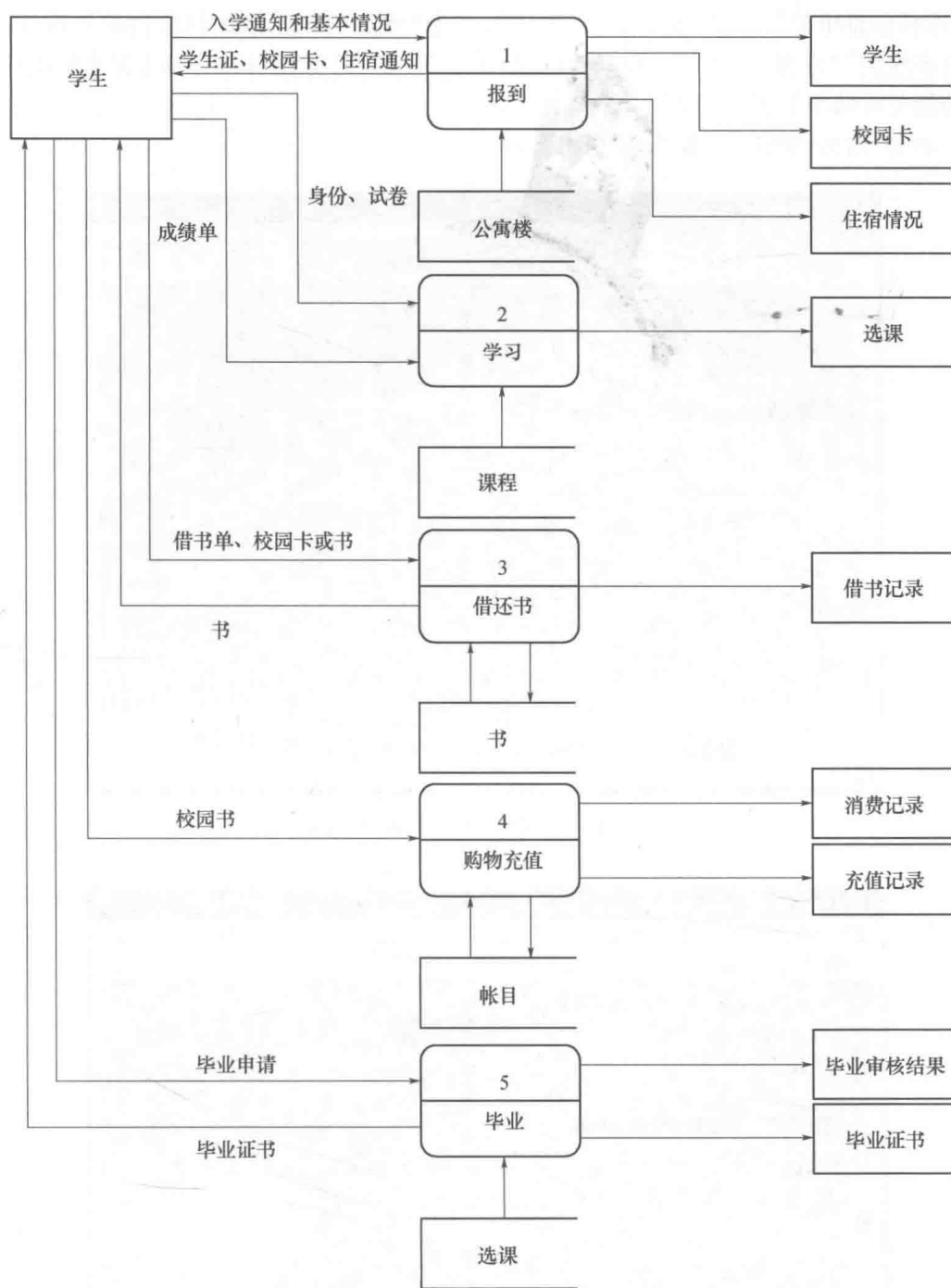


图 1-16 学生信息管理系统顶层数据流程图

护”。选择业务过程“安排住宿”，点击“增加”按钮，添加相应的业务活动“学生住宿信息录入”和“公寓楼信息维护”。点击“保存”按钮，以上要素形成的业务功能分析效果如图 1-17 所示。

选择职能域“学习”，点击“增加”按钮，添加业务过程“选课”和“考试”。选择业务过程“选课”，点击“增加”按钮，添加相应的业务活动“学生选课信息录入”“选课信息维护”

和“课程信息维护”。选择业务过程“考试”，点击“增加”按钮，添加相应的业务活动“考试内容维护”“设置考试信息”和“考试成绩录入”。点击“保存”按钮，以上要素形成的业务功能分析效果如图 1-18 所示。

点击“返回”按钮，点击菜单“退出”。

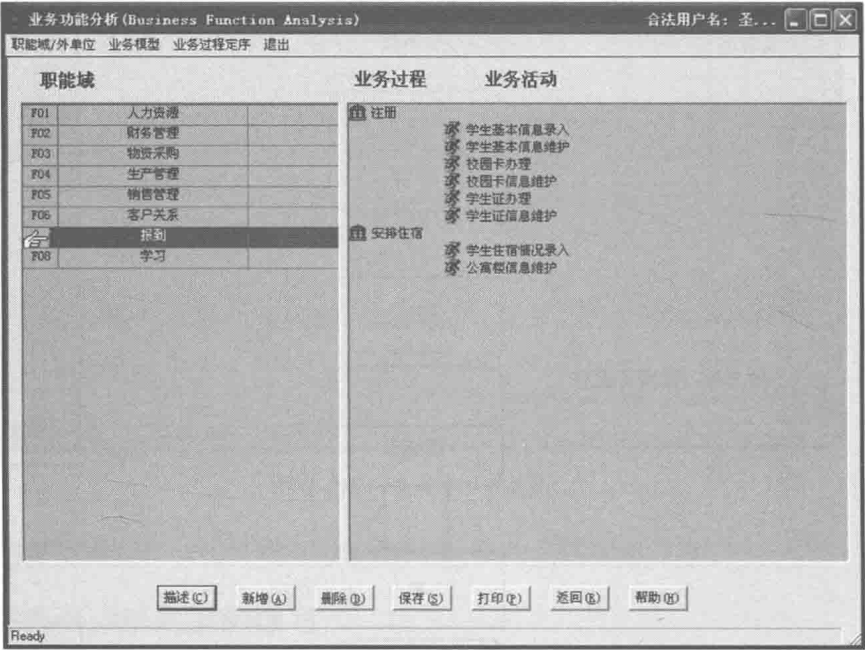


图 1-17 职能域“报到”的分析结果

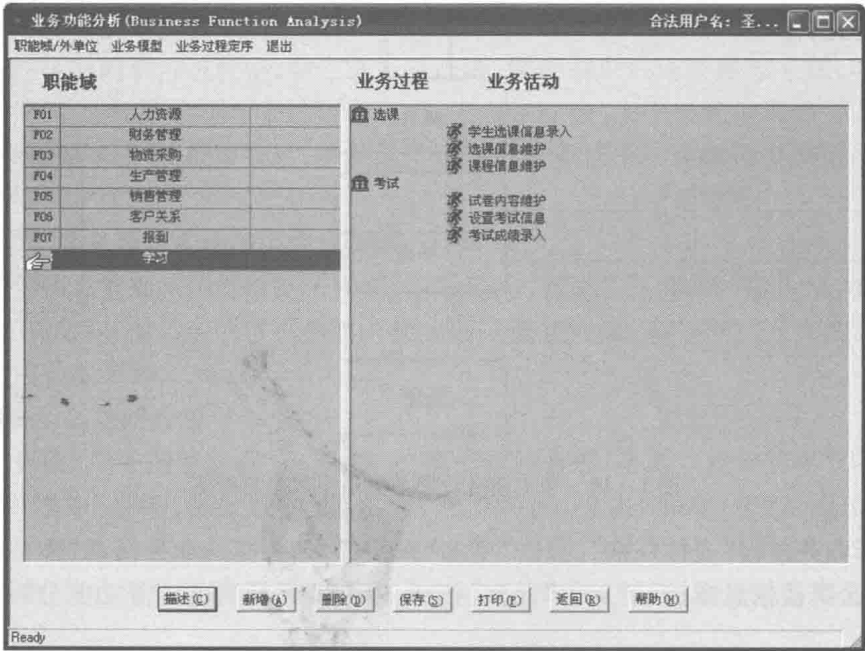


图 1-18 职能域“学习”的分析结果

步骤2 分析系统的用户视图、数据元素

点击主界面“业务数据分析”选项,进入“业务数据分析”界面。点击菜单“用户视图”,选择职能域“报到”,点击“增加”按钮,输入用户视图“学生报到表”“校园卡申请表”“学生证申请表”“学生信息查询”“校园卡信息查询”“学生证信息查询”“住宿申请表”“公寓楼信息查询”,点击“保存”按钮,如图 1-19 所示。

标识	职能域	视图标识	用户视图名称	流向	类型	生存期	记录数	统计参数
01	人力资源	D072101	学生报到表	存储	单证	永久	20,000	1
02	财务管理	D072102	校园卡申请表	存储	单证	永久	20,000	1
03	物资采购	D072103	学生证申请表	存储	单证	永久	20,000	1
04	生产管理	D072104	学生信息查询	存储	单证	动态	20,000	1
05	销售管理	D072105	校园卡信息查询	存储	单证	动态	20,000	1
06	客户关系	D072106	学生证信息查询	存储	单证	动态	20,000	1
07	报到	D072107	住宿申请表	存储	单证	年	20,000	1
08	学习	D072108	公寓楼信息查询	存储	单证	动态	500	1

共有视图 8 个

用户视图登记

增加(A) 删除(D)

用户视图组成(C) 打印(P) 保存(S) 返回(B) 帮助(Q)

图 1-19 “报到”的用户视图

选择职能域“学习”,点击“增加”按钮,输入用户视图主要包括“学生选课单”“选课信息查询”“课程信息查询”“考试成绩录入”“试卷内容录入”“考试信息设置”“考试信息查询”,点击“保存”按钮,如图 1-20 所示。

标识	职能域	视图标识	用户视图名称	流向	类型	生存期	记录数	统计参数
01	人力资源	D082101	学生选课单	存储	单证	永久	40,000	1
02	财务管理	D082102	选课信息查询	存储	单证	动态	40,000	1
03	物资采购	D082103	课程信息查询	存储	单证	动态	1,000	1
04	生产管理	D082104	考试成绩录入	存储	单证	永久	40,000	1
05	销售管理	D082105	试卷内容录入	存储	单证	年	40,000	1
06	客户关系	D082106	考试信息设置	存储	单证	年	1,000	1
07	报到	D082107	考试信息查询	存储	单证	动态	1,000	1
08	学习							

共有视图 7 个

用户视图登记

增加(A) 删除(D)

用户视图组成(C) 打印(P) 保存(S) 返回(B) 帮助(Q)

图 1-20 “学习”的用户视图