

基于实体-关系-问题建模体系的信息资源整合建模研究

王 宁¹, 韩胜菊², 李怀明¹, 王延章^{*1}

(1.大连理工大学 管理学院, 辽宁 大连 116024;
2.大连水产学院 信息工程学院, 辽宁 大连 116023)

摘要: 整合政府、企业的信息资源,需要对信息资源整合的建模方法进行研究.为此提出了一种基于实体-关系-问题($E-R-P$)的信息资源整合建模方法,介绍了实体、关系和问题三者之间的关系,以及 $E-R-P$ 建模过程,并比较了 $E-R-P$ 建模与传统 ER 建模的差别.结合信息资源整合实际项目,阐述了这种建模方法.该建模方法可为信息资源整合的系统实现提供方法性指导.

关键词: 实体-关系-问题;信息资源整合;建模过程;建模方法
中图分类号: N945.12 **文献标识码:** A

0 引言

随着计算机技术的发展,越来越多的政府和企业已经建立起了自己的信息资源库和信息系统,但“信息孤岛”现象也日趋严重^[1].为摆脱各部门、各系统信息无沟通、各自为政的局面,实现局域、广域、异构操作系统和数据库环境下的信息集成,将分散的信息资源更好地统一、整合、管理,进行信息资源整合显得尤为关键.而对信息资源整合建模直接影响到整合项目的成败.

以往对企业的建模,按目的和用途可分为功能模型(如 CIM-OSA GRAI网络、IDEF0等)、组织模型(如 GRAI CIM-OSA组织建模等)、过程模型(如 工作流、Petri 网、CPM、PERT、IDEF3 等)、资源模型(如 IEM、CIM-OSA等)^[2].但这些建模方法主要是针对信息系统的构建或某种业务的处理,而信息资源整合是在原有的系统、资源和网络基础上的建设过程,与以往的信息系统建设有一定的差别.

虽然目前信息资源整合已经成为信息化建设的热点与重点,但针对这一领域的建模理论研究并不多.同时,信息资源整合又具有跨组织、跨网络、跨系统等特点.信息资源具有数据类型、结构、语义关系复杂,数据来源多样,信息标准变化快,对数据的处理要求各不相同等特点^[3].因此

需要一种通用性强、扩展性强、灵活、简便的建模方法对信息资源整合进行建模,进而推动信息资源整合的实现.

针对上述情况,本文通过对基于实体-关系-问题($E-R-P$)建模体系和建模过程的研究,结合实际项目,提出一种基于 $E-R-P$ 建模体系的信息资源整合建模方法.

1 $E-R-P$ 模型体系

1.1 概述

按哲学观点,一个客观事物总是可分的,可以分成单元、元素或下一个层次事物,这些单元、元素、下一层次事物(或简称为子事物)是一种客观实在,并构成客观事物的内在因素活动.为了方便,称这些客观实在为实体(entities).

基于哲学中“事物是普遍联系的”观点,构成一个客观事物的所有实体都是相互联系着的.这种联系体现在实体之间的相互依赖、相互作用的关系(relations)之中.不难理解,任一个客观事物的运动或任一领域的知识都可以分解成有关实体和相应关系加以描述.

问题(problems)是人的主观愿望或要求与客观事物系统状态的差异,因此,问题是关于人们主观目标(目标的概念来源于客观事物系统)要

收稿日期: 2005-04-20; 修回日期: 2006-06-03.
基金项目: 国家自然科学基金资助项目(70271045); 辽宁省自然科学基金资助项目(2051066).
作者简介: 王 宁(1973-),男,博士;王延章*(1952-),男,教授,博士生导师.
©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

求和有关的客观事物系统的集合^[4].

论域 (domain) 是人们研究一个客观事物系统 S 时,所涉及到的研究范围. 如果研究的是某个系统的内部问题,那么,论域可以是这个系统;如果研究的是几个系统间的问题,那么,论域就是这几个系统的集合.

这样,对客观事物系统 S 的表示可分别从实体、关系、问题和论域的角度进行,从而形成了客观事物系统的 $E-R-P$ 体系. 系统可以用如下模型来表示:

$$S = \int_D (E, R, P)$$

式中: D 表示系统的论域 (或作用空间); E 表示实体; R 表示关系; P 表示问题; $\int$ 表示系统的建模.

对实体、关系和问题的数学描述,可参见文献 [3-4].

$E-R-P$ 模型的关键是引入了 P . P 涵盖了需求、问题,以及问题的求解. 对信息资源整合来说,系统当前状态和目标状态之间存在差异,这个差异就是问题. 问题求解的过程就是对这种差异的解决,可以对应着相关的结构搭建.

模型是对一个系统的抽象, $E-R-P$ 模型是对系统的结构和功能的抽象,把结构和功能结合起来,能更全面地描述一个系统,更好地指导系统的实现. $E-R-P$ 通过对实体、关系、问题的描述,一方面可以实现对系统结构和功能的抽象描述,指导系统的构建;另一方面,可以使人们更好地理解系统的结构和功能,确保客户、最终用户以及系统开发人员对系统具备统一的认识.

1.2 实体、关系、问题间的关系

实体、关系与问题三者相互依存,共同描述系统的功能和结构. 实体是点,关系是边,问题涵盖在点与关系的描述中,也涵盖在系统整体中 (如图 1). 单是点和边不能完全地描述系统本身,问题贯穿于整个系统的构造过程当中,指导系统结构

的确定. 系统的构建应当以问题 (任务) 为起点,通过对实体、关系和问题的描述来实现. 实体的划分、关系的确定,是对问题的求解.

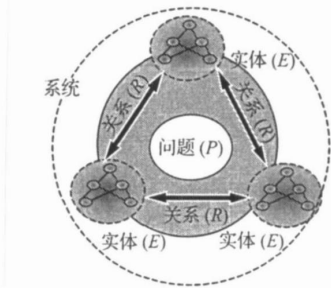


图 1 实体、关系、问题三者的关系
Fig. 1 The relation of entity, relation and problem

1.3 $E-R-P$ 建模体系与 ER 模型的差异

传统的 ER 模型是一种用来描述系统结构的 概念模型,主要用于信息系统中的数据建模. 信息系统主要是用来记录信息、展示信息的,所以传统的 ER 模型在一定程度上还能满足其数据建模的需要. 可对于政府信息资源整合来说,传统的 ER 模型就显得不够了. 一方面,传统的 ER 模型没能体现出系统要做什么,没有体现出当前结构和理想结构的差异,没有体现出目标和驱动. 它只是一种对状态和结构的描述,没有体现出“实现 (承约)”. 另一方面,任何信息系统都不是万能的,只能是一个系统对应一个业务,不存在普遍适用的数据库. 以往的 ER 模型不能描述通用的东西,只能是对具体问题的分析. 而 $E-R-P$ 模型从问题 P 出发,研究实体与关系,使模型更具有普遍性,这也是 $E-R-P$ 建模体系的优势所在. 表 1 将 ER 模型和 $E-R-P$ 模型进行了对比.

表 1 ER 模型与 $E-R-P$ 模型比较
Tab. 1 Comparison of ER model with $E-R-P$ model

模型	建模的类型	模型的性质	应用的主要领域	主要表述方式
ER	结构模型	概念模型	数据建模	ER 图
$E-R-P$	结构与功能模型	包含了概念与逻辑	系统建模、知识表示	形式化语言描述

1.4 $E-R-P$ 建模过程

系统建模首先要明确系统所涉及的范围与作

用空间,即确定论域;然后根据系统的需求,明确系统的目标和建模的目标;再根据目标问题,研究

问题的解决方案,划分实体,描述关系,并制订出相应的 $E-R-P$ 模型(如图 2).

系统的建模,是对问题先进行概括,再对问题进行分解的过程. 问题最终可分解成原子问题,直接对应操作.

建模过程根据系统的层次划分,逐层建立 $E-R-P$ 模型,是自顶向下的过程. 论域由大变小,系统逐步细化,问题、目标、求解方案逐层分解. 下一层的模型在一定程度上是上一层问题的实现. 建模的层次数目可以根据项目的实际需要而定(如图 3).

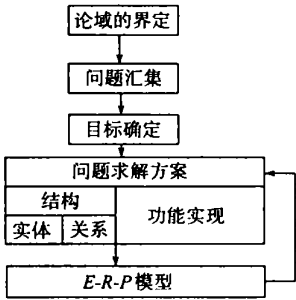


图 2 一个层面的 $E-R-P$ 建模过程
Fig. 2 The $E-R-P$ modeling process on one layer

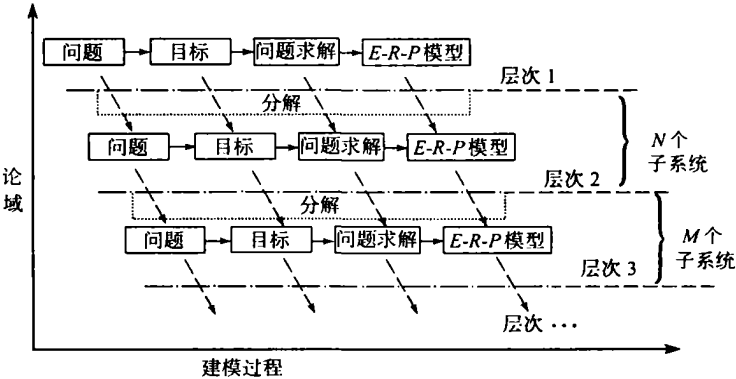


图 3 $E-R-P$ 建模过程
Fig. 3 The $E-R-P$ modeling process

2 信息资源整合建模

信息资源整合的 $E-R-P$ 建模是建模方法和建模过程的实现. 本文以 $\times\times$ 市的工商、国税、地税 3 个局的信息资源共享为例, 介绍基于 $E-R-P$ 的信息资源整合的建模. 遵循上面介绍的 $E-R-P$ 建模过程, 进行分层次、分阶段的信息资源整合系统建模, 并将其分为 5 个阶段.

整合前的状态是: 3 个局的数据各自输入, 有重叠、相互独立; 存在着进行信息共享的问题; 在管理上, 要求信息产业局负责这个信息资源共享项目, 并进一步利用这些信息资源为市民和政府服务.

2.1 第一阶段: 建模准备

首先界定论域, 确定信息整合所涉及的部门、设施、资源以及研究的范围, 在这一阶段, 仅是对问题、实体以及一些简单关系的罗列和整理, 还没

构成目标系统; 将参与信息资源整合的实体划分出来, 明确其相关属性; 根据实际需要提出问题, 理解信息资源整合中各部门的需求, 整理出需要解决的各个问题(如图 4(a)).

(1) 系统描述

论域: {工商局, 地税局, 国税局, 信息产业局, 政府办公网络, 各局需要共享的信息资源, 管理制度, 计算机技术, 信息使用者}

实体集: {工商局, 地税局, 国税局}

问题集: {工商局与地税局的信息共享需求, 国税局与地税局的信息共享需求, 国税局与工商局的信息共享需求}

问题细分为

{地税局对工商局的信息共享需求, 地税局对国税局的信息共享需求, 工商局对地税局的信息共享需求, 工商局对国税局的信息共享需求, 国税局对地税局的信息共享需求; 国税局对工商局的

信息共享需求}

子问题细分为 (以第一个子问题为例,其他略)

地税局对工商局的信息共享需求: {纳税人税务登记情况, 纳税人年终申报纳税额及经营情况, 非正常户信息, 税务机关对纳税人的监管信息}

子子问题 (以第一个子子问题为例,其他略):

纳税人税务登记情况: {纳税人名称, 企业登记注册号, 税务登记号, 企业地址, 纳税登记机关, 注销税务登记的纳税人名称, 企业登记注册号, 税务登记号, 地址, 注销日期, 注销原因}

(2) 数学描述

$D = \{E_1, E_2, E_3, E_0, Net, Resource, Management, Technology, User\}$

$E = \{E_1, E_2, E_3\}$
 $P = \{P_1, P_2, P_3\}$, 其中, $P_1 \triangleq \{E_1 \cap E_2\}; P_2 \triangleq \{E_1 \cap E_3\}; P_3 \triangleq \{E_2 \cap E_3\}$.

问题细分:

$P_{12} = \{P_{12}, P_{21}\}; P_2 = \{P_{13}, P_{31}\}; P_3 = \{P_{23}, P_{32}\}$
其中 $P_{12} \triangleq \{C_1, C_2, E_1 \mapsto E_2\}$, C_1 为 E_1 的属性集, C_2 为 E_2 的属性集, $E_1 \mapsto E_2$ 表示从实体 E_1 到 E_2 需求. 其他类似.

子问题细分:

$P_{12} = \{P_{121}, P_{122}, P_{123}, P_{124}, P_{125}\}$
 $P_{21} = \dots$
 $\dots$

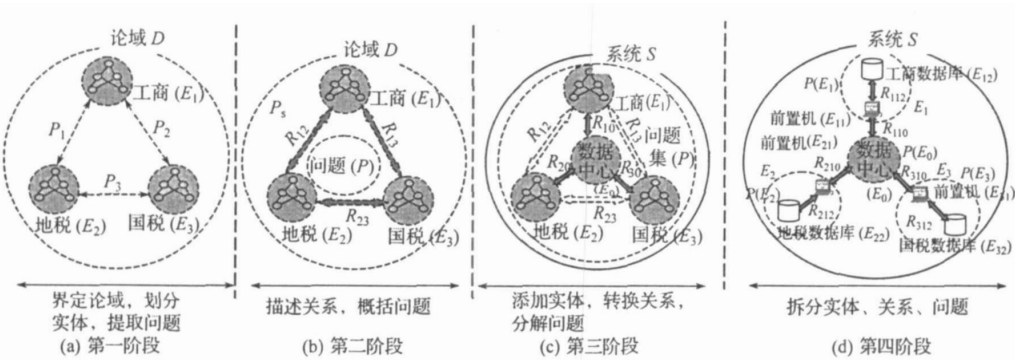


图 4 信息资源整合建模

Fig. 4 Information integrating method

2.2 第二阶段: 粗糙模型

从上一阶段的问题出发,描述各实体间的关系,明确部门间信息整合的数据交换关系. 组织结构关系; 汇总各细小问题,对问题进行概括,总结出信息资源整合的整体性问题; 从系统的角度出发,研究问题和系统的结构 (如图 4(b)).

(1) 数学描述

$E = \{E_1, E_2, E_3\}$
 $R = \{R_{12}, R_{13}, R_{23}\}$
 $P(E_1) = \{P_1, P_2\}; P(E_2) = \{P_2, P_3\}; P(E_3) = \{P_1, P_3\}$
 $P_s = \prod (P_1(R_{12}(E_1, E_2)), P_2(R_{13}(E_1, E_3)), P_3(R_{23}(E_2, E_3)))$

其中 $\prod$ 表示问题的概括, P_s 表示系统问题.

(2) 关系矩阵

$$\begin{matrix} & E_1 & E_2 & E_3 \\ \begin{matrix} E_1 \\ E_2 \\ E_3 \end{matrix} & \begin{bmatrix} 0 & R_{12} & R_{13} \\ R_{12} & 0 & R_{23} \\ R_{13} & R_{23} & 0 \end{bmatrix} \end{matrix}$$

由于研究的是不同实体间的整合问题,忽略实体自身的关系,记为 0.

(3) 相应的 $E-R-P$ 阵

$$\begin{bmatrix} 0 & R_{12} & R_{13} \\ R_{12} & 0 & R_{23} \\ R_{13} & R_{23} & 0 \end{bmatrix} \times \begin{bmatrix} P(E_1) \\ P(E_2) \\ P(E_3) \end{bmatrix} = \begin{bmatrix} P(E_2)^* R_{12} + P(E_3)^* R_{13} \\ P(E_1)^* R_{12} + P(E_3)^* R_{23} \\ P(E_1)^* R_{13} + P(E_2)^* R_{23} \end{bmatrix}$$

这里的“+”和“*”是广义的“加”和“乘”,如

$P(E_1) * R_{12}$ 表示与实体 E_1 相关的问题集 $P(E_1)$ 中,与关系 R_{12} 相对应的问题。

2.3 第三阶段: 模型改进

基于对上一阶段系统问题的解决,需要在信息产业局设立数据中心。在原来 3 个实体的基础上,添加数据中心实体,通过数据中心来解决共享以及资源再利用问题。从而,原来的工商局和地税局的信息共享关系,转换为工商局从数据中心提取和提交信息关系,以及数据中心向地税局提供和提取信息的关系。原来的一个关系,用后来的两个关系代替和表达(如图 4(c))。同时,将问题分解为网络连通问题、数据交换问题、信息安全问题、数据库建设问题、数据中心设置问题、软件实现问题等。

(1) 数学描述

$$E = \{E_1, E_2, E_3, E_0\}, \text{其中 } E_0 \text{ 为数据中心}$$
$$R_{12} = R_{10} \cup R_{20}; R_{13} = R_{10} \cup R_{30}; R_{23} = R_{20} \cup R_{30}$$
$$P = \coprod (P(Network), P(DataChange), P(Security), P(DB), P(DataCenter), P(Software)), \text{其中 } \coprod \text{ 表示问题的分解。}$$
$$\{P\} = \{P(Network), P(DataChange), P(Security), P(DB), P(DataCenter), P(Software)\}$$

(2) 关系矩阵

$$\begin{matrix} & E_0 & E_1 & E_2 & E_3 \\ \begin{matrix} E_0 \\ E_1 \\ E_2 \\ E_3 \end{matrix} & \begin{pmatrix} 0 & R_{10} & R_{20} & R_{30} \\ R_{10} & 0 & r & r \\ R_{20} & r & 0 & r \\ R_{30} & r & r & 0 \end{pmatrix} \end{matrix}$$

其中 r 表示间接关系,在关系矩阵中也称为虚关系。

(3) $E-R-P$ 阵

$$\begin{pmatrix} 0 & R_{10} & R_{20} & R_{30} \\ R_{10} & 0 & r & r \\ R_{20} & r & 0 & r \\ R_{30} & r & r & 0 \end{pmatrix} \times \{P\}$$

其中 $R_{ij} * \{P\}$ 表示 R_{ij} 所涉及到的相关问题。

2.4 第四阶段: 模型细化

基于上一阶段的问题,进行实体的拆分、关系

的分解,并将问题进一步分解。这一阶段建立的 $E-R-P$ 模型也是上层问题求解的过程。

将原来的实体按一定拆分规则进行划分,并描述其属性,原来的各局被拆分成前置机和原有系统数据库两个实体。实体拆分后,必然有新的实体间关系需要描述,原有的关系被转化,需要建立更底层的实体间关系。同时,也细化出一系列新的问题(如图 4(d))。

数学描述如下:

$$E_1 = \{E_{11}, E_{12}\}; E_2 = \{E_{21}, E_{22}\}; E_3 = \{E_{31}, E_{32}\}$$
$$R_{10} = R_{110} \cup R_{112}; R_{20} = R_{210} \cup R_{212}; R_{30} = R_{310} \cup R_{312};$$
$$P(E_1) = \coprod (P(GetDB), P(CollectDB), P(Control), P(Show), P(DataContrast), P(DataMirror))$$
$$P(E_0) = \coprod (P(GetDB), P(CollectDB), P(Control), P(ConCtrlDB), P(Show), P(DataExchange))$$

2.5 第五阶段: 最终模型

在原来的实体和关系的基础上,解决上一阶段所提出的问题,进一步拆分实体、细化关系、定义问题。将前置机、数据中心等实体细化为共享库、采集库、管理库等数据库实体以及数据交换中心、数据交换适配器、数据镜像、数据比对等数据交换子系统,以及数据展示和管理等数据应用子系统,最终描述出系统的目标结构模型——信息资源整合模型(如图 5)。而该模型的实现,涉及到数据仓库技术与元数据管理等诸多方面,将是下一层次的问题。对信息资源整合模型及其实现的详细介绍可参见文献 [5]。

$E-R-P$ 建模体系遵循系统工程的“系统是可分的;系统的部分和部分之间,整体和部分之间存在着有机的联系;系统具有整体功能”^[6] 的基本思想。这种针对信息资源整合的建模体现出了目前状态与理想状态的差异。建模过程本身也是对这种差异的描述和解决过程。基于 $E-R-P$ 建模体系设计出的信息资源整合模型,已经在多个政府部门的信息资源整合项目中得到应用,并取得了很好的实际效果。

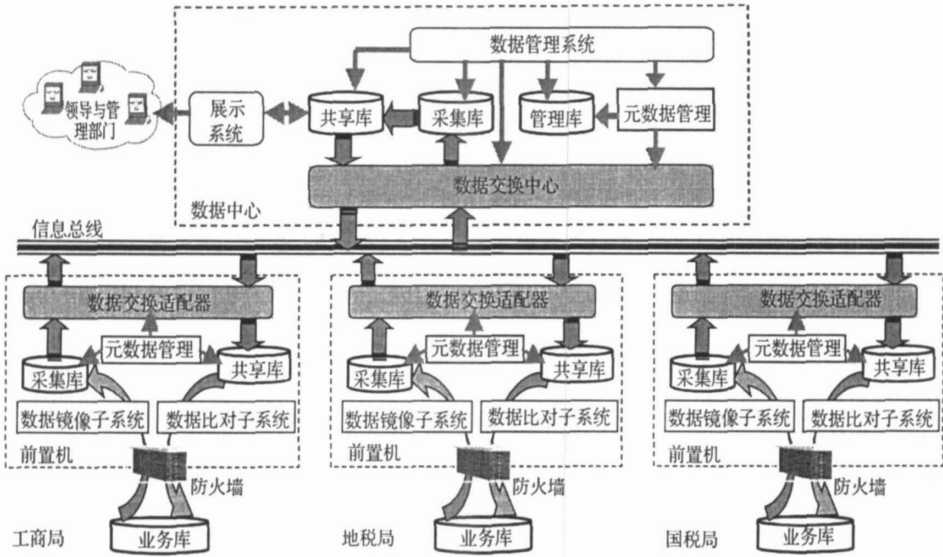


图 5 信息资源整合模型

Fig. 5 The information integration model

3 结 语

基于本文提出的建模方法,逐层对整合项目进行分析和设计,能够较大幅度地体现整合的需求和功能,指导结构的规划,更好地实现对信息资源整合的建模,达到信息资源共享的目的.实践证明,这种建模方法可以指导信息资源整合项目的实施,同时对一般的信息系统建模也具有一定的参考价值.

参考文献:

[1] 魏铁进. 网络环境下我国信息资源整合问题的思考 [J]. 郑州航空工业管理学院学报, 2003, 21(4): 99-100

[2] 范玉顺,王 刚,高 展,等. 企业建模理论与方法导论 [M]. 北京: 清华大学出版社, 2001

[3] 于 淼,王延章. 基于实体-关系-问题建模体系的政务资源整合研究 [J]. 系统工程与电子技术, 2004, 26(5): 647-651

[4] WANG Yan-zhang, YANG Bai, WANG Zhong-tuo. ERP System of Knowledge Representation for DSS [C] // Proceedings of International Conference on Information & Systems. Hangzhou International Academic Publishers, 1991: 852-856

[5] 王 宁,王延章,叶 鑫,等. 一种基于数据中心的政府信息资源整合系统架构设计 [J]. 计算机应用研究, 2005, 22(9): 67-68

[6] 王众托. 系统工程引论 [M]. 北京: 电子工业出版社, 1991: 1-3

Research on information integrating method based on entity-relation-problem modeling system

WANG Ning¹, HAN Sheng-ju², LI Huai-ming¹, WANG Yan-zhang^{* 1}

(1.School of Manage., Dalian Univ. of Technol., Dalian 116024,China;
2.School of Electr. and Inf. Eng., Dalian Fisheries Univ., Dalian 116023, China)

Abstract In order to integrate the information resources in governments and enterprises, a research on modeling method of information resources integration is needed. Thus, a modeling method of information resources integration based on the study of entities-relations-problems (*E-R-P*) is proposed. The relationships among entity, relation and problem and modeling process are introduced. The differences between *E-R-P* model and traditional *ER* model are compared. The modeling method is expounded combining with the practical program of information resources integration. The modeling method can provide the methodological guidance for the systematic realization of information integration.

Key words entity-relation-problems; information integration; modeling process; modeling method