

基于领域知识的需求信息获取*

陆汝铃 金芝 万荣林

夏幼明

(中国科学院数学研究所 北京 100080) (云南师范大学数学系 昆明 650092)

摘要 本文提出一种基于领域知识的需求信息获取方法,相应的 MIS 开发环境 PROMIS 已成功地应用于多个领域.

关键词 基于知识的系统,需求说明,管理信息系统.

在管理信息系统 MIS(management information systems)的开发中,屡见有失败的例子.失败的根本原因在于需求分析没有做好.而需求分析做不好的原因又在于用户和软件工程师之间缺乏良好的合作.用户往往不能确切和清楚地说明他的需求,甚至还要保留在 MIS 研制过程中随时修改需求的自由.^[1]这一来,软件工程师就很难作出恰当的需求分析,从而也就很难完成研制能有效运行的 MIS 的任务.

目前软件工作者在需求分析中研究和使用的主要方法是形式化或半形式化方法,这些方法使用了不同的需求说明语言.形式化方法的优点是它能引导软件工程师或程序员按照严格的步骤写出需求说明^[2~5],但使用任何形式化方法的前提是已经确切和完整地收集到了用户的需求信息.如果做不到这一点,则不管形式化方法在理论上如何严密,它是不能帮助我们构造有效的 MIS.也就是说,形式化方法不能解决我们在上面提到的用户不能很好合作的问题.针对这个问题而提出的一个解决方案是:尽可能多地把用户吸引到 MIS(或一般地,应用软件)的开发中来,使用户能尽早发现正在研制中的软件与他的需求之间的偏差,从而及时修改需求说明和设计.但是,形式化方法很难被直接用作软件工程师和用户之间合作的基础,形式化的需求说明语言也很难在这两部分人中间起沟通作用,因为他们的背景往往很不相同,更何况在研制和运行 MIS 的过程中,需求信息可能改变,修改设计或重新设计都将给这两部分人带来沉重的负担.

我们认为,不仅应该一般地把用户吸引到 MIS 的开发过程中来,而且在可能的情况下,应该把开发 MIS 的钥匙交到用户手里,让用户自己来定义、开发、维护和修改他的 MIS.在这里,关键是要免去用户学习和掌握软件开发知识的负担,还要免去用户按形式化方法做需

* 本研究得到国家“八五”攻关项目、国家 863 高科技项目及国家自然科学基金资助.陆汝铃,1935 年生,研究员,博士生导师,主要研究领域为人工智能,计算机动画,知识工程,软件工程.金芝,女,1962 年生,副研究员,主要研究领域为人工智能,软件工程,知识工程.万荣林,1971 年生,助教,主要研究领域为人工智能,软件工程.夏幼明,1956 年生,讲师,主要研究领域为人工智能,计算机科学理论.

本文通讯联系人:陆汝铃,北京 100080,中国科学院数学研究所

本文 1995-08-31 收到

求分析的负担. 为了实现这个目标, 本文提出了一种基于知识的 MIS 开发方法 KISSME (knowledge intensive software system manufacture engineering), 这个方法要点是以一个大容量知识库来支持 MIS 的开发, 该知识库包含 2 个部分: 软件工程知识库和领域知识库. 前者使用户不必自己掌握软件开发知识, 后者使用户不必自己掌握有关应用领域的需求分析知识. 在这个知识库的支持下, 用户只需把本单位的业务情况描述清楚, 系统即可自动生成一个有效的 MIS. 我们还设计和实现了相应的 MIS 开发环境 PROMIS (PROtotyping MIS). 不难看出, KISSME 方法和 PROMIS 系统是面向最终用户的 (此处, “最终用户” 特指 MIS 的用户, 而非 PROMIS 一类工具系统的用户). 它们有助于填平软件工程师和最终用户之间的鸿沟, 并有利于在某些领域中把开发 MIS 的钥匙交到最终用户手里.

1 PROMIS 系统

作为一个 MIS 开发环境, PROMIS 的核心任务是构造 MIS 使用单位业务情况的一个概念模型, 然后把此概念模型转换成能运行的 MIS. 为此目的, 在具体构造每个 MIS 时, PROMIS 的领域知识库中存放着适用于该应用领域的抽象概念模型. 最终用户输入该用户单位的具体业务信息. PROMIS 运用此具体业务信息于抽象概念模型之上, 对抽象概念模型作匹配、代真、剪裁、扩充、改造等工作, 即得到一个确切地描述该单位业务情况的具体概念模型. 因此, PROMIS 的主要技术关键之一是设计和实现适合于不懂软件技术的最终用户输入业务信息的界面, 并正确地分析和理解这些业务信息. 应该强调指出: 对用户输入的分析 and 理解也不能离开领域知识的支持.

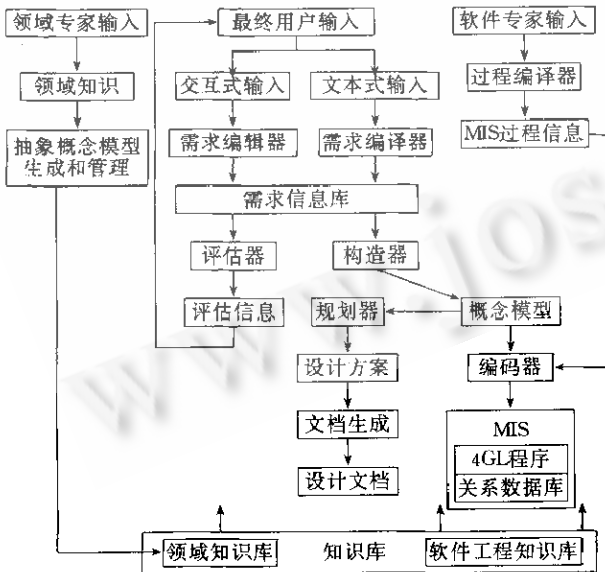


图1 PROMIS的体系结构

PROMIS 的主要组成模块如下:
 (1)交互式需求信息输入; (2)类自然语言描述需求信息输入; (3)需求编辑器; (4)需求编译器; (5)需求评估器; (6)概念模型生成; (7)4GL 代码生成; (8)关系数据库生成; (9)软件设计生成; (10)设计文档生成; (11)专家需求信息输入; (12)过程型需求信息编译; (13)领域知识的输入和管理. PROMIS 的体系结构如图 1 所示.

这里有几个问题需要说明. 首先, PROMIS 不是只面向最终用户. 它面向 3 类用户: 最终用户可以利用 PROMIS 中已有的知识库构造 MIS; 软件专家可以利用 PROMIS 的过程式

语言编写复杂和精细的 MIS 功能程序 (主要是数据库操作); 领域专家可以输入领域知识以增添新的抽象概念模型. 当然最主要的用户还是最终用户.

其次, 最终用户有 2 种输入方式. 一是回答 PROMIS 的提问, 用交互方式输入.

PROMIS 的提问是基于领域知识的,它以用户能理解的方式显示当前应用领域的抽象概念模型,然后由用户在模型上实施剪裁操作,以得到具体的概念模型.二是用类自然语言文本输入.我们设计并实现了类自然语言 BIDL (business information description language),用 BIDL 语句可以描述企事业单位的具体业务信息,而不必使用软件术语. PROMIS 利用 BIDL 包含的信息对抽象概念模型进行剪裁,也可以得到具体概念模型.关于这 2 种输入方式我们在下面的 2 节中还要作深入的讨论.

第 3, PROMIS 提供 2 种类型的输出. 它可以直接输出一个能运行的 MIS, 如同我们在前面说过的那样, 这个 MIS 由 2 部分组成: 一组 4GL (某个第 4 代语言) 的程序和一个与该 4GL 相应的关系数据库 (一组关系及其数据集). 通过 4GL 编译程序加工, 即成为一个可执行的 MIS. 除此之外, PROMIS 还可提供另一种输出: 所需 MIS 的设计文档. 它可以完成通常是软件工程师的需求分析和系统设计工作, 并把余下的编码任务交给程序员去做. 完成设计文档分 2 步走. 第 1 步拟出设计方案 (例如对象和类的划分、属性和变量的确定等); 第 2 步利用自然语言生成技术生成文本形式的设计文档, 相当于一本已经编辑好的设计文件, 可以直接打印成一本小册子.

第 4, PROMIS 还有一个需求评估器, 可以对用户输入的需求信息进行评估, 如果发现问题 (包括需求信息不完整, 或有矛盾), 即可向用户发出提示, 由用户补充和更正有关的需求信息.

2 类自然语言 BIDL

我们说把开发 MIS 的钥匙交到最终用户手里, 其关键是最终用户能以他自己所熟悉的方式提供需求信息. 帮助最终用户做到这一点的是类自然语言 BIDL. 在 BIDL 中没有数据流、类、对象这样的软件术语, 有的只是财务、人事、仓库、营业、服务这一类企业用语. 它外表很象自然语言, 企事业管理人员易学易用. 但它又是严格受限的, 不能任意发挥, 因此是一种程序设计语言, 可以被计算机编译和执行. 所谓编译就是需求信息 (知识) 的抽取. 所谓执行就是 MIS 的建立.

BIDL 的基本元素是关键字, 在程序中用下划线标出, 例如: 的任务是, 的领导是, 人员清单为, 服务的方式, 转帐, 打印财务报表, 记帐, 进库时需, 出库时需, 进行盘点, 共有固定资产, 的市场份额, 统计应付帐目, 生成付款单, 基本工资, 职务工资 等等.

利用关键字, 可以构成 BIDL 语句. 语句的格式可用 BNF 方式表示. 举例如下:

〈领导关系句〉 ::= 〈上级〉[部门]领导〈下级〉[部门]

〈部门设置句〉 ::= 〈部门名〉分为〈部门名序列〉[等共〈正整数〉个]

〈职务设置句〉 ::= 〈部门名〉的职务有〈职务序列〉[等共〈正整数〉种]|〈部门〉的领导是〈领导人〉

它们可以例化为如下的句子:

* 经理办公室领导餐厅部.

* 餐厅部分为中餐、西餐、清真餐厅等共 3 个.

* 餐厅部的职务有餐厅经理, 餐厅帮办, 厨师长, 公款吃喝稽查员, 反浪费宣传员等共 5 种.

还可以有这样的句子格式：

〈帐务管理句〉 ::= 〈建帐单位〉采用〈记帐法〉

〈库房句〉 ::= 〈部门名〉有〈库房序列〉等〈正整数〉间库房 | 〈库房〉和〈库房〉相距〈正整数〉〈距离〉

〈服务项目句〉 ::= [〈部门名〉]有〈服务序列〉等〈正整数〉种服务项目

〈报表制度句〉 ::= 〈报表名〉报表打印一式〈正整数〉份,分送〈部门名序列〉

它们可以例化为如下的句子：

* 光明餐厅采用复式记帐法。

* 正大商场有百货库、副食库等 2 间库房。

* 百货库和副食库相距 500 米。

* 坦诚公司有托运货物、特快专递、代客进口等 3 种服务项目。

* 资金月报表打印一式 5 份,分送区政府、区税务局、区工商局、总经理办公室、董事长办公室。

BIDL 程序的第 3 个层次是段落或称章。章是把一组同类的句子(描述需求信息的某一方面的句子)集合起来,形成知识模块。目前的 BIDL 有 11 章,它们是:机构章、人事章、财务章、仓库章、服务章、报表章、数据章、公关章、生产章、资产章、市场章。上面的例句分别取自机构章、财务章、仓库章、服务章、报表章。不难看出,企事业管理人员使用这样的程序设计语言不会有什么困难。

BIDL 的语义通过关键字及其组合体现,并用语义网络描述。我们使用的是语义网络语言 SNETL^[9],一个 SNETL 程序由一组关系表达式组成,每个关系表达式都写成三元组

(节点 关系 节点)

的形式。在某些情况下节点也可以是关系,因此 SNETL 具有部分高阶表达的能力。SNETL 有一组系统固有的关系,其中最重要的是 is-a, sub-of 和 part-of 3 种,分别表示从属关系、包含关系和部分关系。其它的关系有 time(时间), way(方式), place(地点), reason(原因), goal(目的), degree(程度), belief(可信度), object(对象), number(数量)等。三元组内的节点和关系前皆可以加标号(应用性出现),三元组本身之前也可以加标号(定义性出现)。这些标号把三元组集合联结成语义网络。用户可以定义自己的关系。他还可以说明关系的类型,例如关系是对称的、或反对称的、或传递的。

下面用例子说明如何通过 SNETL 描述 BIDL 的语义,这种描述可附加在语法公式上。

例 1:〈部门包含句〉 ::= 〈部门 1〉是〈部门 2〉的一个分部

SEMANTICS ((division1), is-a, DIVISION)

((division1), part-of, (division2))

BIDL 语句:服务台是天马宾馆的一个分部

转化成语义网络:(服务台, is-a, DIVISION)

(服务台, part-of, 天马宾馆)

例 2:〈距离句〉 ::= 〈位置 1〉和〈位置 2〉的距离是〈距离〉

SEMANTICS ((integer), value:L1, (distance):L2)

L1:(米, measure, value)

L2:((distance), distance, ((place1), <place2>))

BIDL 语句:服务台和财务科的距离是 50 米

转化成语义网络:(50,value:L1,距离:L2)

L1:(米,measure,value)

L2:(距离,distance,(服务台,财务科))

3 需求信息的获取

前已说过,PROMIS 的需求信息获取是通过用抽象概念模型匹配输入信息并进行加工以获取具体概念模型的. 经过对 BIDL 程序的初步加工之后,输入信息现在以 SNETL 语义网络的形式出现. 为了说明需求信息获取过程,我们应该先给出抽象概念模型的结构.

抽象概念模型也称领域模型,是该应用领域中信息处理知识的集中和概括. 领域模型分为过程模型和数据模型 2 部分. 其中过程模型又分为 2 部分. 一部分是由领域专家输入生成的,称为 Templates(样板),它们将被用户输入的需求信息例化并转化为 4GL 代码,构成 MIS 程序的基本部分. 另一部分是由软件专家输入的,所利用的语言称为 PBPDL(primitive based process description language),其中包含一系列的 MIS 操作原语(Primitives),可以被加工成数据库操作进程,作为对由 Templates 生成的基本 4GL 程序的补充,是对 MIS 功能的“精雕细刻”.

在过程模型中,每个 Template 都包含 2 部分内容. 一部分是对 MIS 的某个子系统的程序功能描述,又称功能骨架. 另一部分是当需要例化这个功能骨架时在功能骨架上可能进行的操作. 这后一部分内容体现了 Template 的自修改功能. Template 的基本结构如下:

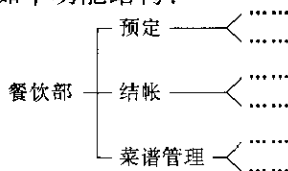
```

TEMPLATE <名字>
  TYPE <类型>
  ATTRIBUTES <当前对象的可观察性质>
  RELATIONS <关系样板>
  SUBOBJECTS <子对象关系>
  EVENTS <对当前对象中的属性的操作>
  OPERATIONS <可能施行于当前对象的操作>
  BEHAVIORS <当前对象的行为描述>
  PROCEDURES <描述对象行为的过程>
  CONSTRAINTS <对象属性的静态约束条件及其它外部条件>
END TEMPLATE

```

鉴于篇幅限制,此处不拟对 TEMPLATE 中各项作详细说明,仅举一例以示其概要.

假设天马宾馆的餐饮部有如下功能结构:



则描述餐饮部 MIS 子系统的 TEMPLATE 结构大致如下:

```

TEMPLATE 餐饮部
  TYPE 功能实施型
  ATTRIBUTES 名字:string;资金:Real;.....
  RELATIONS LOCAL order_rel(名字);..... # 点菜 #
    INPUT price_rel(名字);..... # 物价 #
    OUTPUT account_rel(名字);..... # 结帐 #
  EVENTS [create(N)]名字=N; # 建立新对象 #
  .....
  OPERATIONS new Behavior ..... # 建立新操作 #
    change Behavior ..... # 更改操作 #
    disable Behavior ..... # 删除操作 #
  BEHAVIORS order =
    (append(N)→order_rel(Name).create(N)
    |modify(N)→order_rel(Name).modify(N));
    # 点菜登记及修改 #
    .....
  PROCEDURES [compute(M)]=SELECT(SUM(price×amount),
    (order_rel(Name).object(N),
    menu_rel(Name).object(L)),
    foodno in order_rel(Name).object(N)=
    foodno in menu_rel(Name).object(L))
    .....
  CONSTRAINTS price:ATTRIBUTE(menu_rel(Name));
    amount:ATTRIBUTE(order_rel(Name));
    foodno:ATTRIBUTE(menu_rel(Name)) & ATTRIBUTE(order_
    rel(Name));
END TEMPLATE

```

不难看出,BEHAVIOR 集中了该 TEMPLATE 的固有操作,PROCEDURES 是这些操作的详细描述,CONSTRAINTS 是对它们的限制,EVENTS 把 TEMPLATE 例化为对象,OPERATIONS 则在例化时对 TEMPLATE 的操作进行裁剪.

PBPD 一共有8类原语,它们是基本变量赋值;基本变量输出;元组变量赋值;集合变量赋值;集合变量输出;表元素添加;表元素删除;表内容修改.它们基本上概括了通常的数据库操作.至于这些原语的具体书写格式,这里不再赘述.

PROMIS 领域模型中的数据模型也分为2部分:一组关系框架以及它们的实体-关系语义描述.这里我们对所使用的实体-关系表示略作说明,它分为4条线索,其中每条线索本身又是一个关系框架:

1. 两个实体间的关系及其序数

实体1	实体2	关系	序数1	序数2
-----	-----	----	-----	-----

2. 实体及其属性集

实体	(属性, 类型)集
----	-----------

3. 属性及其类型和实体集

属性	类型	实体集
----	----	-----

4. 属性及其限制

属性集	限制集
-----	-----

在使用实体—关系模型时,我们有一个基本假设,那就是它应该是领域完备的.即它的实体、关系和属性应能覆盖用户输入的需求信息,也即在生成 MIS 时不需要增加新的实体、关系和属性.如果出现了相反的情况,就要调用领域知识维护程序.

在与用户的需求信息匹配时,可施行于实体—关系模型上的操作(对实体—关系模型的例化)包括改变实体的属性;改变实体间的关系;改变关系的序数;改变属性值的限制;属性值的代真;序数的代真;一致性检查的传播等等.这些操作之间互相有联系.例如,删去一个实体可能导致某些属性或关系被删除,反之也是一样.

关系框架的操作涉及较多的技术细节,此处不赘述.

对照图1,不难看出,蕴含在诸 TEMPLATE 中的信息和操作可用来获取构造4GL 程序所需的信息.蕴含在关系框架和实体—关系模型中的信息可用来获取一个关系数据库.两者合在一起,构成了向最终用户获取需求信息的基础.

4 结 论

以知识为基础开发 MIS 和其它应用软件,这个思想在文献中已屡有所见.例如, M. Jarke 的 DAIDA 系统^[10]就是这方面的著名工作之一.这些工作有一个共同的特点:他们说的知识往往是指软件知识,再加上一些 MIS 领域共有的知识.至于特定的领域知识,如旅馆、商场、银行等的业务知识,均不包含在现有的 MIS 开发方法和工具之中.而 KISSME 方法和 PROMIS 工具,则除软件知识、MIS 领域通用知识外,还包含特定的领域知识.我们提出的方法和技术区别于其它方法和技术的明显特点之一,就是把特定领域的知识作为 MIS 开发的重要支柱.乍一看来,似乎目前流行的通用 MIS 开发方法概括性强,应用面广.而使用领域知识则将使知识库十分庞大,甚至失去控制.它“就事论事”,概括性差.实际上,用同一个工具应用于所有的 MIS 领域,这工具的能力必然是有限的,对开发人员的要求必然是很高的.我们的方法则是以不同的知识模块对付不同的应用领域.在同一个 PROMIS 工具中,领域知识模块是可换的,工具本身并不臃肿.而且正因为有了领域知识才得以把 PROMIS 交给最终用户直接使用.我们的方法还有一个额外的好处,就是可以使领域知识模块的开发独立于 PROMIS 工具的推广. PROMIS 向第3方知识软件商提供集成界面.最终用户可以直接向他们购买知识软件(知识包),从而使我国知识软件产业得到更好的发展. KISSME 方法和 PROMIS 系统已成功地应用于饭店、商场、税务等领域,并开发了相应的 MIS.

参 考 文 献

- 1 白光野. MIS 需求分析方法. 中国科学院软件研究所技术资料, 1994.
- 2 Cameron J R. An overview of JSD. *IEEE Transactions on Software Engineering*, 1986, **SE-12**(2): 222~240.
- 3 Spirey J M. *The Z-Notation: a reference manual*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- 4 Bjørner D, Hoare C A R, Langmaack H ed. *VDM and Z—formal methods in software development*. Springer Verlag, 1990.
- 5 Hoffmann B, Krieg-Brückner B ed. *Program development by specification and transformation*. Springer Verlag, 1993.
- 6 Lu Ruqian, Jin Zhi, Wan Ronglin. A knowledge based approach for automatically prototyping MIS. Proc. of AVIGNON'94, Paris, 1994.
- 7 Lu Ruqian, Jin Zhi, Wan Ronglin. Requirement specification in pseudo-natural language in PROMIS. Proc. of COMPSAC95, 1995. 96~101.
- 8 万荣林. PROMIS 的需求分析功能[硕士论文]. 中国科学院数学研究所, 1993.
- 9 夏幼明. 语义网络语言 SNETL[硕士论文]. 中国科学院数学研究所, 1987.
- 10 Jarke M, Mylopoulos J, Schmidt J W *et al.* DAIDA: an environment for evolving information systems. *ACM Transactions on Information Systems*, 1992, **10**(1): 1~50.

AN APPROACH OF ACQUIRING REQUIREMENT INFORMATION BASED ON DOMAIN KNOWLEDGE

Lu Ruqian Jin Zhi Wan Ronglin

(Institute of Mathematics The Chinese Academy of Sciences Beijing 100080)

Xia Youming

(Department of Mathematics Yunnan Normal University Kunming 650092)

Abstract A new approach of acquiring requirement information based on domain knowledge is proposed. PROMIS, the corresponding MIS development environment, has been applied to several application domains successfully.

Key words Knowledge-based systems, requirement specification, management information systems.