

基于 UML 的 MIS 分析与设计方法

A MIS Analysis and Design Method Based on UML

汪林林 华奇兵

(重庆邮电学院中韩合作 GIS 研究所 重庆400065)

Abstract Aiming at the disadvantages of traditional structured MIS analysis and design method based on E-R model, we present a MIS analysis and design method based on UML and summarize rules of mapping persistent object to RDBs. This method unifies the two processes of software design and database design, thus making MIS development more efficient.

Keywords MIS analysis and design, UML, OO method, RDB, Mapping

1 引言

MIS 系统开发一般须经过需求分析和系统设计两个阶段。传统需求分析对问题的认识和描述不是以问题域中的固有的事物作为基本单位并保持其原貌,而是在全局范围内以功能、数据或数据流为中心,例如数据流法把整个问题域看作一些数据流和处理,所以这些方法的分析结果不能直接地映射到问题域,而是经过了不同程度的转化和重新组合。设计阶段中数据库的设计是用实体-关系模型方法(E-R 方法),而在分析与设计这两个阶段进行转化时没有一个可靠的转换规则。如果在需求分析阶段采用面向对象的分析方法,设计阶段采用面向对象的设计方法,在编程阶段采用面向对象程序设计语言,则不仅填平了需求分析与设计阶段的鸿沟,使需求阶段到实现阶段,使用统一的对象概念,可以做到平滑的过渡,而且需求分析与设计阶段的对象模型可以很自然地转为数据库的结构。

面向对象的观点认为:任何一个系统是由若干个对象和这些对象间的相互作用构成。其中对象由数据和操作两部分组成。对象具有自主性、封装性和动态性。将面向对象作为方法学应用到软件工程的各个阶段,其实质就是寻找对象及对象间的相互作用关系。在需求分析阶段,面向对象分析是对真

实世界对象建模,其根本出发点是站在应用的角度对问题域进行刻画和描述,这样有利于对问题的理解。设计阶段是对原对象模型进一步描述,此阶段可用面向对象方法实现数据库的设计。

UML 是在著名的 Booch 方法、OMT 方法和 OOSE 方法的基础上,广泛征求意见,集众家之长,几经修改完成的。本文首先简要介绍了统一建模语言,由于目前流行的商用的数据库管理系统还是基于关系模型的,因此有必要对持久对象与关系数据库的映射进行一定的讨论。紧接着从面向对象分析与设计技术在 MIS 设计中的应用出发,提出一种基于 UML 的 MIS 分析与设计方法。应用基于 UML 的 MIS 分析与设计方法,一方面可以分析设计出软件程序,另一方面又可以完成数据库的设计,把这两个过程统一在整个 MIS 分析设计全过程中,相对传统 E-R 图来说有助于提高开发效率。

2 统一建模语言(UML)

2.1 统一建模语言的主要内容

UML 定义了用例图、静态图、行为图、交互图、实现图等五类模型图,共10种,其间的关系如图1所示(如果把该图当作一棵树的话,则所有的叶子结点即为 UML 定义的10种模型图)。

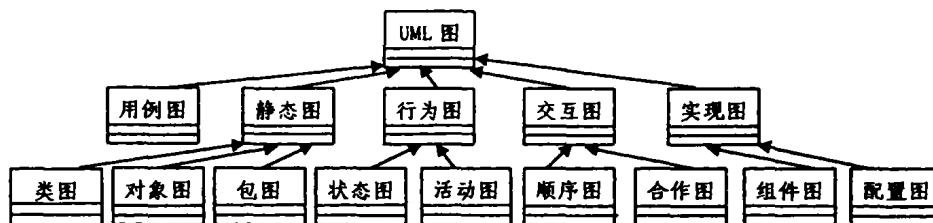


图1 UML 的10种模型图的关系图

从应用的角度看,当采用面向对象技术设计系统时,首先是描述需求;其次根据需求建立系统的静态模型,以构造系统的结构;第三步是描述系统的行为。

2.2 UML 的主要特点

统一建模语言 UML 的主要特点可以归结为以下三点:

(1)UML 是国际软件界广泛认可的系统建模语言的标准,它统一了对象方法的标记,使软件设计工具能发挥更大的

效用。现有的对象设计也能更容易地被重新使用。UML 不仅统一了 Booch、OMT 和 OOSE 等方法中的基本概念,还吸取了面向对象技术领域其他流派的长处,其中也包括非 OO 方法的影响。

(2)UML 取长补短,吸取了如 Booch、OMT 等多种表示方法的长处,适当地平衡了简洁性和具体化。UML 已经成为一个单独的演化系统,不致于像以前那样有多种标准而引起

汪林林 教授,研究领域为数据库系统、GIS 系统、计算机网络等。华奇兵 硕士,研究领域为数据库系统、GIS 系统、面向对象技术等。

误会。

(3)UML 在演变过程中还提出了一些新的概念,如构造型、线程、过程等概念。

因此,UML 是一种标准的、先进的建模语言。

2.3 统一建模语言的应用领域

UML 的目标是以面向对象图的方式来描述任何类型的系统,具有很广的应用领域。其中最常用的是建立软件系统的模型,但它同样可以用于描述非软件领域的系统,如机械系统、企业机构或业务过程,以及处理复杂数据的信息系统、具有实时要求的工业系统或工业过程等。总之,UML 是一个通用的标准建模语言,可以对任何具有静态结构和动态行为的系统进行建模。

3 持久对象与关系数据库的映射

3.1 三级模式的扩展

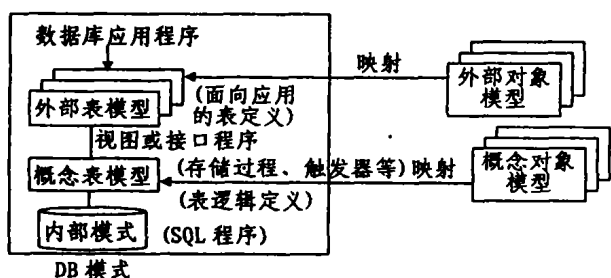


图2 三级数据库模式面向对象模型的扩展

数据库系统的三级模式是数据的三个抽象级别。通过内模式把数据的具体组织留给 DBMS 去完成,用户只要抽象、逻辑地处理数据,而不用关心这些数据如何在计算机中表示和存储,从而大大减轻了用户使用计算机的负担。最接近物理数据库的内部模式由 DBMS 提供的 SQL 来描述。概念模式是数据的全局逻辑表示。一般的概念模式是数据库物理模式作用域的边界,它能实现数据库的物理意义、特定 DBMS 的特殊操作对外部应用程序的信息隐蔽。外部模式是从特定用户应用角度看待的数据库模式,从不同的应用出发对同一概念模式可以给出多种不同的外部模式。当外部应用系统以对象模型进行抽象时,从各个应用出发抽象出的对象模型可以映射到外部模型上,对此我们不妨称之为外部对象模型。但是,外部模型只是概念模型的子集,所以数据库设计的核心在于系统对象模型(不妨称之为概念对象模型)向数据库概念模型的映射。图2为三级数据库模式面向对象模型的扩展。

3.2 对象模型向关系数据库表的映射规则

映射规则简单地总结如下:

(1)将类的属性类型(Attribute Type)映射成域(Domain)

(2)将类的属性(Attribute)映射为表的列(Column)

(3)在关系数据库中实现继承

用唯一表示类的对象 OID 来表示数据库中的键码,从而使得在关系数据库中实现类的继承、聚集和关联变得简单。在关系数据库中实现继承有三种基本的方法:

方法一:对整个类继承使用一个表。所谓整个类继承是指将类继承图当作一棵树时其中的所有结点之间的继承关系,将整个类继承映射为关系数据库中的一个表,其中类的所有属性都映射为该表的列。

方法二:对整个类继承中的每一个非抽象类使用一个表。

在这种方法中,每一个非抽象类(抽象类的属性由子类继承)对应一个表,每个表的属性由一个非抽象类的属性及其继承的属性组成。

方法三:对整个继承的第一个类使用一个表。每一个类(无论是抽象类还是非抽象类)对应一个表,表的属性由其对应的类的属性与 OID 组成。

(4)关联的映射

·一或零对一关联的映射 如果是零对一关联,将两个关联的类分别映射为两个表,将“零”的对象类映射的表,加上一个指向另外一个表的外键;如果是一对一关联,将两个关联的类分别映射为两个表,并任选一个表,加上一个指向另外一个表的外键。

·一对多关联的映射 将两个关联的类分别映射为两个表,将“多”的对象类映射的表,加上一个指向另外一个表的外键。

·多对多关联的映射 实现多对多关联,需要引入关联表的概念。传统实现中,关联表的属性包含关系中两个表的主键,并且关联表的主键往往是它们的组合。另一种实现方法是将关联表视为普通表,使用自身的主键 OID,然后加入实现关系所必需的外键。

三元和多元关系与二元关联类似。

4 基于 UML 的 MIS 分析与设计方法

虽然在整个软件生命周期中所用的概念都是一致的,但是为了强调不同的分工与侧重点,下面总结提出的方法仍然分为分析与设计两个阶段。在分析阶段侧重于问题域,建立一个独立于实现的模型,而设计阶段重点考虑与实现相关的因素。下面应用面向对象分析与设计的原则与方法,结合 MIS 的特点,总结提出一种基于 UML 的 MIS 分析与设计的方法。

4.1 基于 UML 的 MIS 分析

基于 UML 的 MIS 分析方法,可以分为以下几个步骤:

1)在用户参与的情况下,对需求进行调查和分析,定义执行者和用例,得到用例模型。

2)从问题陈述和用例模型的分析中,建立初始的分析对象模型。这个模型概要地给出了描述问题域应用语义的类以及这些类间的关系和原始属性。

3)重新定义和精化类与类间的关系,产生一个精化了的分析对象模型。

4)参照分析对象模型,可以从用例图、行为图和交互图来描述系统的动态特性。通过重复分析过程同时精化对象模型和动态分析模型。

由上可以看出,在这一阶段将涉及到三个重要的模型:用例模型、分析对象模型、动态分析模型。分析过程中各种模型的关系如图3所示。

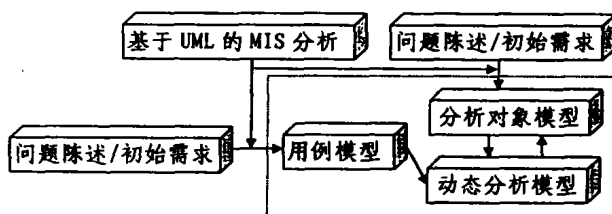


图3 分析过程中各种模型的关系

下面对用例模型、分析对象模型、动态分析模型三个模型

作进一步的阐述。

·用例模型 在系统分析阶段,由系统开发人员与用户不断地交互讨论,导出一个双方均认同的需求的文档,但通常情况下需求文档并不十分清晰,综合性也不太好。为此,建立复杂一些的用例模型,可以更好地理解用户需求,并驱动分析与设计,甚至测试。

用例模型可以用用例图来进行刻画。从需求规范到用例模型有四个步骤:1)定义出系统的确切范围;2)发现执行者(或者用户);3)寻找用例;4)定义用例间的关系。

·分析对象模型 在确定了用例模型之后,对用例进行详细的分析,开始标识对象。对象是描述问题域以及构造应用的基础。分析对象模型主要包括:构造系统的对象类、对象之间的关联、对象维护的信息(对象的属性)、对象提供的服务(对象的行为)等。可以用UML中的对象图或类图进行描述。

·动态分析模型 描述了应用的动态行为。主要包括:1)应用中的事件流;2)对象的状态改变。动态分析模型可以用UML的行为图与交互图进行描述。

我们通常用分析对象模型来刻画对象的静态结构,而用

动态分析模型来刻画应用的动态行为。这两种模型是紧密相关的,它们反映了同一应用的不同方面。所以在建立这两种模型时不一定是顺序的,这两种模型可以交替进行,同时开发,互相补充。

注意,在分析阶段,分析对象模型和动态分析模型还处于草图状态,具体的细化要到设计部分进行。

4.2 基于UML的MIS设计

在以数据库为核心的MIS设计中,数据库的设计是系统设计的一个中心环节,数据库的结构方式以及效率问题会影响整个系统的应用与扩充,因此设计阶段的其中一个重要任务是利用前面基于UML的MIS分析结果,继续细化,形成设计模型,完成数据库的设计及软件的设计工作。基于UML的MIS设计分以下步骤进行:

1)架构设计。需要定义包(子系统),包间的相关性和基本的通信机制,系统平台选择等;

2)在保证可追踪问题域的前提下,把分析阶段的分析对象模型进一步细化成与实现相关的设计对象模型;

(下转第137页)

(上接第54页)

表3 GI统计及其Z值

编号	名称	Gi(d)	E(Gi)	VAR(Gi)	Zvalue
1	黑龙江	0.053	0.069	0.000	-0.720
2	内蒙古	0.266	0.275	0.000	-0.428
3	新疆	0.116	0.103	0.000	0.734
4	吉林	0.760	0.103	0.000	-1.304
5	辽宁	0.092	0.103	0.000	-0.621
6	甘肃	0.212	0.207	0.000	0.298
7	河北	0.242	0.241	0.000	0.042
8	北京	0.073	0.069	0.000	0.281
9	山西	0.144	0.138	0.000	0.401
10	青海	0.140	0.138	0.000	0.152
11	天津	0.082	0.069	0.000	1.229
12	陕西	0.244	0.241	0.000	0.154
13	宁夏	0.103	0.103	0.000	0.002
14	西藏	0.133	0.138	0.000	-0.231
15	山东	0.141	0.138	0.000	0.236
16	河南	0.213	0.207	0.000	0.307
17	江苏	0.116	0.138	0.000	-1.217
18	安徽	0.188	0.207	0.000	-1.130
19	四川	0.282	0.276	0.000	0.166
20	湖北	0.197	0.207	0.000	-0.564
21	上海	0.046	0.069	0.000	-2.012
22	浙江	0.169	0.172	0.000	-0.098
23	江西	0.204	0.207	0.000	-0.143
24	湖南	0.209	0.207	0.000	0.148
25	云南	0.138	0.138	0.000	0.035
26	贵州	0.129	0.138	0.000	-0.630
27	福建	0.099	0.103	0.000	-0.276
28	广西	0.148	0.138	0.000	0.545
29	广东	0.192	0.172	0.000	0.710
30	海南	0.045	0.034	0.000	1.243

区域,中西部高人口增长率区域,中东部低人口增长率区域和南部高人口增长率区域。

结束语 本文将空间统计方法引入到数据挖掘领域,空间自相关分析可以用来研究数据的空间分布特征,空间关联分析可以分析空间数据的空间关联性。同时本文用这两种空间统计分析方法研究了两个时期人口增长率的空间关联关系。1982—1990年中国可分为北部人口低增长、中西部高人口增长率、中东部低人口增长率和南部高人口增长率等四个区

域;1990—1998年中国可分为北部低人口增长率和南部高人口增长率两个区域。两个时期的空间聚类虽然不完全相同,但它们有共同的特点,南部和西部的人口增长率都比较高。这些结论比较符合中国的实际情况,这就说明空间自相关和G统计分析是有效的数据挖掘方法。

参考文献

- 1 Agrawal R, Imielinski T, Iyer B, et al. Mining association rules between sets of items in large databases. In: Proc. of ACM SIGMOD Conf. on Management of Data, 1993. 207~216
- 2 Agrawal R, Srikant R. Fast algorithms for mining association rules. In: the Proc. of VLDB—1994, Chile, Sep. 1994. 487~499
- 3 Agrawal R, Srikant R. Mining generalized association rules. In: the Proc. of VLDB—1995, Switzerland, 1995. 407~419
- 4 Fayyad U, Piatetsky-shapiro G, Smyth P. From data mining to knowledge discovery: An overview. Advances in knowledge discovery and data mining, MIT Press, 1996. 1~34
- 5 Park M, Chen M, Yu P. An effective hash based algorithm for mining association rules. In: the Proc. of ACM SIGMOD 1995. 175~186
- 6 Han J, Fu Y. Discovery of multiple-level association rules from large databases. In: Proc. of VLDB—1995, Switzerland, 1995. 420~431
- 7 Cliff & Ord. Spatial processes: Models and applications. London: Pion, 1981
- 8 Goodchild. Spatial autocorrelation (CATMOG47). Norwich, UK: Geobooks, 1986
- 9 Getis and Ord. The analysis of spatial association by the use of distance statistics. Geographical analysis, 1992, 24: 189~206
- 10 Ord & Getis. Distributional issues concerning distance statistics. Working paper, The Pennsylvania State University and San Diego State University, 1994
- 11 Ding & Fotheringham. The integration of spatial analysis and GIS. Computers, Environment an urban systems, 1992, 16(1): 3~19
- 12 Kehris. Spatial autocorrelation statistics in ARC/INFO: [Research Report NO. 16]. North west regional research laboratory, Lancaster University, 1991

根据不同的应用会有多种复杂拓扑关系,以上所列出的只是其中一小部分,但已体现时空对象的动态性了。

3 时空查询语言

目前我们正在研究支持移动空间对象的子系统 ROSTE (RObust Spatial-Temporal Extension) 包括移动空间对象的数据类型、移动空间对象的存储结构、移动空间对象间的操作、移动空间对象与空间对象的操作等。并且正在把该系统与可扩充 DBMS-AMOS^[7]集成,生成时空 DBMS——STDBS。

STDBS 的查询语言为 STQL, 可以进行一些非常复杂的时空分析操作。设有模式如下:

```
planes(airline: charstring, id: charstring, Flight: Point)
cities(name: charstring, id: charstring, position: region)
weather(kind: charstring, Position: Region)
```

例1 找出所有两架飞机航线间的距离小于1公里的飞机。

```
SELECT p. airline, p. id, q. airline, q. id
FROM planes p, planes q
WHERE min(Distance(p. Flight, q. Flight)) < 1
```

例2 查询7:00到7:20时间内,国航300飞机所在的城市。

```
SELECT cities. name FROM planes, cities
WHERE Pr_Inside(planes. Flight, cities. position, 7:00, 7:20) AND
planes. airline = "Airchina300"
```

例3 查询曾穿过暴风雪或浓雾区的飞机数,并分组表示。

```
SELECT COUNT(*) FROM planes, weather
WHERE weather. kind = "snowstorm" OR weather. kind = "thickfog"
GROUP BY PR_Cross(planes. Flight, weather. Position)
```

结论及未来有关工作 本文讨论了时空数据库中的各种时空查询,主要对时空拓扑谓词进行了定义。为了满足具体应用的要求,文章还重点探讨了时空对象的复杂拓扑谓词,最后分类例示了具体的 STQL 查询。本文给出的拓扑谓词可用作时空 DBMS 中时空对象拓扑关系操作。我们已设计和实现了基于 Realms^[8]的时空对象表示、存储与管理,并且实现了大部分基于平面扫描算法的各种时空拓扑操作算法,算法的效

率是令人满意的,作者将另文讨论时空分析操作算法的设计和实现。

参考文献

- 1 秦小麟. 移动空间数据类型和操作的初步研究. 计算机科学, 2000, 27(1)
- 2 Egenhofer M J, Herring J R. Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases: [Technical Report]. Department of Surveying Engineering, University of Maine, 1990
- 3 Erwig M, Güting R H, Schneider M, Vazirgianis M. Abstract and Discrete Modeling of Spatio-Temporal Data Types. In: Proc. of the 6th ACM Intl. Workshop on Geographical Information Systems (ACM-GIS), 1998
- 4 Erwig M, Schneider M. Spatio-Temporal Predicates: [Informatik-Report]. FernUniversity Hagen, Dec. 1999
- 5 Güting R H, Böhlen M H, et al. A Foundation for Representing and Querying Moving Objects: [Informatik-Report 238]. FernUniversität Hagen, 1998
- 6 Erwig M, Schneider M. Developments in Spatio-Temporal Query Languages. IEEE Int. Workshop on Spatio-Temporal Data Models and Languages, 1999
- 7 Karlsson J S, et al. AMOS User's Guide. Linköping University, Sweden, 1996
- 8 Güting R H, Schneider M. Realms: A Foundation for Spatial Data Types in Database Systems. In: Proc. 3rd Intl. Symposium on Large Spatial Databases, Singapore, June 1993. 14~35
- 9 Al-Taha K, Snodgrass R T, Soo M D. Bibliography on Spatio-Temporal Databases. A CM SIGMOD Record, 1994, 22(1): 59~67
- 10 Böhlen M H, Jensen C. Seamless Integration of Time into SQL: [Technical Report R-96-49]. Dept. of Computer Science, Aalborg University, Dec. 1996
- 11 Erwig M, et al. Temporal and Spatio-Temporal Data Models and Their Expressive Power. FernUniversity Hagen: [Informatik-Report]. Dec. 1997
- 12 Claramunt C, Thedriault M. Managing Time in GIS: An Event-Oriented Approach, Recent Advances in Temporal Databases. Springer-Verlag, 1995. 23~42
- 13 Güting R H. An Introduction to Spatial Database Systems. VLDB Journal, 1994, 3(4): 357~399

(上接第47页)

3) 为对象行为和属性的细节建模, 精化类的结构, 建立动态设计模型;

4) 数据库设计。按照前面所介绍的对象映射为关系数据库的规则进行数据设计;

5) 设计与问题域无关的用于系统实现的服务类, 补充完善设计对象模型, 例如数据库访问、用户接口等。

4.3 基于UML的MIS分析与设计方法的优点

上面的基于UML的MIS分析与设计方法相对传统的基于E-R图的结构化分析与设计方法具有以下优点:

- 对象与客观世界的实体相似, 便于为客观世界建立模型。
- UML 提供了大量的易于理解的模型图, 方便统一建模。
- 从需求阶段到实现阶段, 使用统一的对象概念, 开发过程中阶段的改变, 不需要方法或范型的转换, 可以做到平滑的过度。
- 结合 COM+ / CORBA / EJB 等分布式计算技术, 有利于建立大型分布式的 MIS 的应用。
- 能充分利用 OOP 开发工具的潜力, 开发出高质量的代

码, 有利于代码重用及后期的维护。

• 软件的分析设计成果得到了较大程度的复用。

小结 本文提出了一种基于UML的MIS分析与设计方法, 并对其中的一个关键问题——持久对象与关系数据库的映射规则进行了概括性总结。这种方法克服了传统的基于E-R模型的结构化分析与设计方法的缺点, 在MIS分析与设计的具体实践中取得了比较好的结果。

参考文献

- 1 Booch, Jacobson, Rumbaugh. The Unified Modeling Language Reference Manual. Addison-Wesley, 1998. 11
- 2 Priestley M. Practical Object-Oriented Design with UML. 清华大学出版社(影印版), 2000. 12
- 3 刘超, 张莉编著. 可视化面向对象建模技术. 北京航空航天大学出版社, 1999. 7
- 4 Cord P, Yourdon E 著, 邵维忠等译. 面向对象的分析. 北京大学出版社, 1991. 5
- 5 Cord P, Yourdon E 著, 邵维忠等译. 面向对象的设计. 北京大学出版社, 1994. 12
- 6 Blaha M. Using UML to Design Database Applications OMT Associates, Inc.; William Premerlani, General Electric R & D, 1999
- 7 邵维忠, 杨美清. 面向对象的系统分析. 广西科技出版社, 1998. 12