

对象的 RSBO 表示模型*

奚建清

(华南理工大学计算机科学系 广州 510641)

E-mail: csjxq@scut.edu.cn

摘要 对象的内部表示决定了对象的存储方式和访问方式,文章介绍了 RSBO(refined synchronous buffers of objects)表示方法,是 SBO(synchronous buffers of objects)对象表示法的改进,RSBO 表示法利用解析对象结构方式,把对象的复杂结构分解为具有相对简单结构的对象节,RSBO 在 SBO 表示法的基础上引入了指针节的机制,指针节结构是对象指针的扩展,并能同时支持在复杂对象的集合上的访问和计算.另外介绍了基于 RSBO 对象的访问方式和索引结构,传统数据库中的索引技术仍能够继续在基于 RSBO 表示机制的对象库中使用.

关键词 存储管理,数据库,面向对象数据库,对象表示,数据访问.

中图法分类号 TP311

面向对象数据库必须支持对象的联想计算和导航计算.文献[1]中给出了一种称为 SBO(synchronous buffers of objects)的对象表示方法,其基本出发点是在内存中用尽量少的指针表示对象间的关联关系(如 part-of 关系),因为指针只支持导航计算,而不支持集合计算.指针转换及其指针跟踪会极大地降低联想计算的效率.在 SBO 表示中,对象值的一个部分称为一个对象节.用一组内存缓冲数组来存储对象节,并用数组元素下标之间的一对一的对应关系来表达对象节间的同步关联.也即在多个关联的缓冲区的元素指针间实施同步.节的内存地址可在编译时静态确定,或运行时动态分配确定.通过与这种内部表示相适应的访问控制,可提高对象的访问效率.SBO 表示的缺点是不能表达复杂的对象结构,当对象节之间不能直接建立起一对一的对应关系时,SBO 表示就失效了.例如,下面这些情形中,对象节之间不能建立一对一的同步关系:(1)多个对象共享同一子对象;(2)对象节关联自身或关联同一对象集合节中的另一个对象节.因此,SBO 表示在表达能力上存在着缺陷.本文引入了指针节的概念,作为 SBO 表示的一种扩充,使得能在复杂对象关联情况下,建立起对象节之间一对一的同步关系,从而支持在复杂结构对象上的高效访问.这种扩充的 SBO 表示称为 RSBO(refined synchronous buffers of objects)对象表示.

1 RSBO 表示

先用一个例子来说明 RSBO 表示模式.设一个类 *Person* 定义了如下的属性“姓名 年龄 住址 相片”(为简单起见,对象标识符和操作省略了).这个类定义中没有定义复杂的对象关联关系,因此,利用 SBO 表示法,可将 *Person* 对象表示为如下的对象节($s_1 s_2$),其中 $s_1 = (\text{姓名 年龄 住址})$, $s_2 = (\text{相片})$.显然, s_1 和 s_2 之间能建立起一对一的关联关系.逻辑上,一个对象集合的节可表示为一个数组.

在 *Person* 中加入两个新的属性:“朋友 宿舍”,其中“朋友”属性的值域是 *Person* 实例的集合,“宿舍”的值域是另一个类 *ROOM* 的实例集合.如果多个 *Person* 对象可以共享朋友或宿舍,并且一个 *Person* 对象的朋友可以是自己,则这两个属性引入了复杂的对象关联关系.假设为新的属性设两个对象节 $s_3 = (\text{朋友})$, $s_4 = (\text{宿舍})$,则

* 本文研究得到广东省自然科学基金资助.作者奚建清,1962年生,博士后,副教授,主要研究领域为数据库,面向对象语言.

本文通讯联系人:奚建清,广州 510641,华南理工大学计算机科学系

本文 1997-09-15 收到原稿,1998-02-10 收到修改稿

$s_1[i]$ 和 $s_2[i]$ 及 $s_3[i]$ 和 $s_4[i]$ 之间不一定具有逻辑关联关系. 为了在 s_1, s_2, s_3, s_4 之间建立这种关系, 另引入两个对象节 s'_3 和 s'_4 , s'_3 和 s'_4 中的元素定义如下:

令 L 是对象集合节的长度, “&”和“*”与C语言中的地址运算符相同. 对任意 $i, j, k \leq L$, 如果有 $Person$ 类的对象 o , 使 $o = (s_1[i] s_2[i] s_3[j] s_4[k])$, 则令 $s'_3[i] = \&s_3[j], s'_4[i] = \&s_4[k]$.

因而, 如果 o 存在, 则 $o = (s_1[i] s_2[i] s_3[j] s_4[k]) = (s_1[i] s_2[i] * s'_3[i] * s'_4[i])$, 把 $(s_1[i] s_2[i] s'_3[i] s'_4[i])$ 称为 o 的RSBO表示. 下面较形式化地定义这种新的表示, 有关对象节的定义, 在文献[1]中已给出, 这里不再赘述.

设对象集合 O 被分成多个对象集合节 (O_1, O_2, \dots, O_n) , 则对于 O 中对象 o , 有正整数 i_1, i_2, \dots, i_n , 使得 $o = (O_1(i_1), O_2(i_2), \dots, O_n(i_n))$.

定义 1. 一个对象节如果含有对象OID, 则称为一个主对象节. 在一个对象集合 O 中, 对任意 $(o = (O_1(i_1), O_2(i_2), \dots, O_n(i_n)))$, 其中 $O_1(i_1)$ 是 o 的主对象节, 如果 $i_k = i_1$, 则 $O_k(i_k)$ 称为 o 的普通节, O_k 为 O 的普通节, 否则称为异常节.

定义 2. 对对象集合 $O = (O_1, \dots, O_n)$, 设 O_1 是主对象节对应的集合节(称为 O 的主节), 对 O 的任一异常节 O_j , 构造一新的对象节 O'_j 如下: 对任意 i , 若 $O_1(i)$ 和 $O_j(k)$ 同属一对象, 则令 $O'_j(i) = \&O_j(k)$. 称如此构造的 O'_j 是 O 的一个指针节.

定义 3. 在对象集合 O 的SBO表示 (O_1, O_2, \dots, O_n) 中, 如果把所有的异常节都换成相应的指针节, 则 O 表示为 (O_1, O'_2, \dots, O'_n) , 其中 O'_i 或是 O 的普通节, 或是指针节, 则这种表示称为 O 的RSBO表示(改进的SBO表示). 为直观起见, 可把一个对象的主节说成是该对象中异常节的共享者. 在一个对象集合的RSBO表示中, 每个指针节中的元素和主节中同下标的元素是一一对应的, 因此, SBO表示中的同步滑性质仍能在RSBO表示中得到保持.

2 指针节结构

指针节的实现有3种策略.

(1) 在内存中用普通指针实现指针节中的元素, 例如, 在上述例子中, 使 $s'_3[i]$ 等于 $s_3[j]$ 的地址. 这种方式支持快速的内存访问, 一次间址即可访问到相应的异常对象节. 这种方案适合于实现诸如 S_i 和 S_j 元素之间是一一对一的关联关系, 其缺点是未考虑到复杂共享和存储管理等问题.

(2) 用一个称为“译码器”的结点地址作为指针节元素, 其结构如图1所示.

一个对象的每个异常节有一个译码器, 译码器结点具有一个指向异常节的指针, 还含有一个计数器和共享者的地址队列. 计数器记录共享该异常节的内存对象个数, 设立该计数器的主要目的是为了在释放内存分配区时防止指针悬空. 与第1种方法相比, 这种方法需两次间址才能完成访问过程. 为了方便应用程序的处理, 译码器中还可以设置一个处理指示器, 表示对应的异常节被处理过的次数, 以避免在一次执行过程中一个异常节被重复处理的问题.

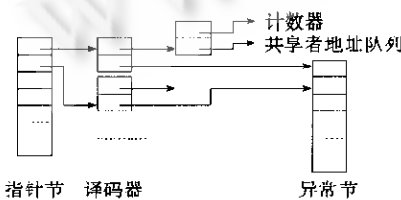


图1 指针节和异常节

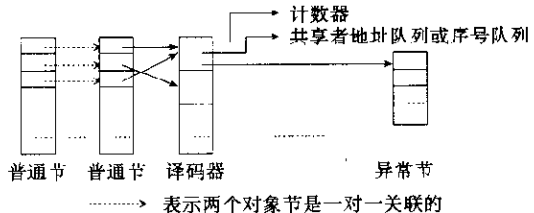


图2 译码器是一个数组

(3) 在上述方法2的基础上, 把各个译码器组织成一个数组, 另设一个普通节, 专门存放译码器数组元素的下标, 从而能使指针节和其他对象节之间实现同步. 如果共享者来自同一个对象集合节, 则译码器的共享者地址队列中也只需存放共享者在节中的序号, 如图2所示.

3 对象访问过程

下面以第2节中的方法3为例,讨论基于RSBO表示下的对象访问过程.在引入指针节结构后,对象的I/O可能会增加指针节的构造,访问对象值的过程为:

- (a) 通过对象标识符或对象结构属性名访问所属类及其类划分表;
- (b) 确定所需要访问的节、索引和外存地址;
- (c) 确定节的类型(普通节或指针节),分配相应的内存空间;
- (d) 输入相应的节至分配的节空间,确定指针节中的各个指针元素的值;
- (e) 通过属性偏差表和节的类型访问对象节,得到相应的对象属性值的直接或间接地址.

上述过程中,如果对象直接类可以静态确定,则通过编译器完成(a)、(b)、(c)、(e)步,从而执行时可以得到最高的访问效率.对象的访问方式既可以是导航方式,也可以是联想方式,也即上述的访问过程与计算模式无关.RSBO表示下,对象的具体访问方法类似于SBO访问方法.^[1]但对于异常节的访问,必须通过指针节间址,这种访问的方式类似于文献[2]中讨论的情形.

4 集合属性的表示

对象的属性值可以是集合类型的(包括列表、元组).当集合属性值中的元素属于普通值类型,并且长度固定的时候,可以用普通节直接实现这种集合类型的值.如果属性值集合是可变长的,或其中的元素是对象,或含有被共享的部分,则可以用指针节来实现这种集合值,只需在每个指针节元素中存放集合的地址,如图3所示.

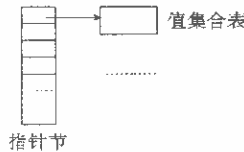


图3 集合属性的存放

5 索引结构

索引机制是数据库用于支持高效的集合访问和联想查找的主要设施.传统的关系库中,索引是用B+树、Hash表等技术来实现.在已有的一些面向对象数据库如Informix中,对对象集合的查询还使用了R+树索引技术.^[3,4]B+树中的索引键是属于一个关系模式中的属性项,而R+树中的索引键是对象结构空间中的一个区域,其中的属性项可以属于多个类模式.在一个数据库中,R+树支持在某一属性空间范围内对象的查找.单纯的B+树和R+树结构不能同时支持对象路径上的正反向查询,而在复杂应用中,从主对象到子对象的正向查询和从子对象到主对象的反向查询都是普遍的.^[2]

在RSBO表示模式下,对于一个对象集合 O ,若 $O=(S_1, \dots, S_n)$,其中 S_i 是 O 的节,则由于 S_i 在结构上的单线性,其外存上的物理存储技术可以采用传统的数据库的存储技术.传统的索引技术如B+树、Hash表和R+树等技术完全可以用于对象节的存储上,所不同的是索引树页结点中的指针.在传统的B+树中,一个索引指针对应一个元组(对象)的存储位置,但在RSBO表示中,一个索引项中的一个键值可能要对应多个对象节的存储地址.

由于RSBO表示模式能把对象的复杂共享或嵌套关联的结构加以析构,在物理组织上线性化,因此,可以在子对象级别上建立索引,支持对对象集合的浏览访问和查询.同时,由于主对象集合节和子对象集合节的顺序是一致的,或能通过指针节达到一致,因此,在一个节上建立的索引可以同时用于查询到其他同步节上的数据.例如,在“朋友”属性的异常节上建立的索引,不仅可以用于查询“朋友”子对象上的数据,还可以通过节间的同步关系访问“person”类的节 s_1 和 s_2 .

对象节间的同步关系使得能够利用指针节实现任何一条对象访问路径,而不仅仅是实现单个对象中局部的 part-of 关系,因此,文献[2]中提出的扩展对象关系可以统一地用 RSBO 表示中的指针节来实现,并在任何一条对象结构路径上建立索引。

6 总 结

RSBO 表示能有效地支持对象的结构复杂性,减少对象结构中指针的使用,对象节和指针节都支持集合的表示和单个对象的表示,因此,RSBO 结构上能够进行高效的导航计算和集合计算。特别是,把具有复杂关联或共享关系的属性表示成统一的指针节有助于在这种复杂属性上的联想搜索,这在已有的模型中是很难实现的。

参考文献

- 1 奚建清. 支持 OODB 高效计算的表示及其机制. 软件学报, 1997, 8(增刊): 74~79
(Xi Jian-qing. An object representation and mechanism supporting efficient computation in OODB. Journal of Software, 1997, 8(supplement): 74~79)
- 2 Kemper A, Amoerkotte G. Object-oriented Database Management. Englewood Cliffs, NJ: Prentice Hall International, Inc., 1994
- 3 Guttman A. R-trees: a dynamic index structure for spatial searching. In: Beatrice Yormark ed. Proceedings of ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1984. 47~57
- 4 Sellis T, Rousopoulos N, Faloutsos C. The R+ tree: a dynamic index for multi-dimensional objects. In: Stocker P M, Kent W, Hammersley P eds. Proceedings of the 13th International Conference on Very Large Databases. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1987. 507~518

RSBO: an Representation Model of Objects

XI Jian-qing

(Department of Computer Science South China University of Technology Guangzhou 510641)

Abstract The internal representation of objects in ODBMS (object-oriented database management system) determines the way to store and access the objects. In this paper, a kind of object representation method——RSBO (refined synchronous buffers of objects) is introduced. The method is a refinement of SBO (synchronous buffers of objects), which is a kind of object representation. With RSBO, an object having sophisticated structure is resolved into relatively simple object sectors (pieces of the object). Based on SBO, RSBO considers the mechanism of object pointer sectors, which is an expansion of object pointers. Pointer sectors not only have the similar properties of SBO sectors, but also support set-oriented computation. The access methods and possible index structures on RSBO are also discussed. The traditional index method can be still used on RSBO.

Key words Storage management, database, OODB (object-oriented database), object representation, data access.