

Chapter 4 数组，串与广义表

数组

数组：相同类型的数据元素的集合，对于一维数组是下标与值组成的序对的集合。多维数组可由一维数组扩充得到，二维数组是最简单的非线性表示

多维数组的存储表示：

- **行优先**：第 $i + 1$ 个行向量紧跟在第 i 个行向量后面，得到数组元素的一维序列
- **列优先**：第 $i + 1$ 个列向量紧跟在第 i 个列向量后面，得到数组元素的一维序列

三维数组一般使用 页/行/列 的优先级表示

特殊矩阵的压缩存储

对称矩阵

对称矩阵：矩阵元素关于主对角线对称

对于阶数为 n 的对称矩阵，可以仅存储其上三角矩阵或下三角矩阵，需保存的元素为 $\frac{n(n+1)}{2}$ 个

下三角：按 $a_{00}, a_{10}, a_{11}, a_{20}, a_{21} \dots$ 的序列存储，矩阵元素 a_{ij} 在数组中的偏移量为 $\frac{i(i+1)}{2} + j$ （仅针对 $i \geq j$ 的情况， $i < j$ 时按 a_{ji} 计算）

上三角：按 $a_{00}, a_{01}, a_{02}, \dots, a_{0n-1}, a_{10}, \dots$ 的序列存储，矩阵元素 a_{ij} 在数组中的偏移量为 $\frac{i(2n-(i+1))}{2} + j$ （仅针对 $i \leq j$ 的情况， $i > j$ 时按 a_{ji} 计算）

多对角线矩阵

三对角矩阵：当 $|i - j| > 1$ 时有 $a_{ij} = 0$ ， n 阶矩阵共有 $3n - 2$ 个非零元素

按行优先，存储序列为 $a_{00}, a_{01}, a_{10}, a_{11}, a_{12}, a_{21}, \dots$ ，矩阵元素 a_{ij} 在数组中的偏移量为 $2i + j - 3$

稀疏矩阵

稀疏矩阵：非零元素数量远小于矩阵元素数量，若使用二维数组存储，浪费大量存储单元和运算资源

稀疏矩阵的三元组表示：对于每个非零元素，以行有限顺序顺序存储其（行/列/元素值）的三元组

稀疏矩阵的快速转置：建立辅助数组， $a[0]$ 为 0， $a[i]$ 存储矩阵中列号小于 i 的元素数量之和（用一个额外数组存储每个列号对应的元素数量，然后累加），其值便是矩阵转置后第 i 行开始存储的位置（下标），时间复杂度为 $O(\max\{col, terms\})$ ，空间复杂度上要多增加两个大小为 col 的数组

稀疏矩阵加法：类比多项式加法

稀疏矩阵乘法：类比快速转置，对于 $A \times B$ ，为 B 建立辅助矩阵，标志每行开始的位置

稀疏矩阵的正交表示：行链表和列链表都是带表头结点的循环链表，用表头结点表征第几行/第几列

字符串

字符串：简称为串，是 n 个字符的一个有限序列。

若字符串不为空，从字符串中连续取出若干个字符组成的串叫做原串的**子串**，称子串的第 0 个字符在原字符串中的位置为子串在串中的位置，空串是任意串的子串，一个串是它本身的子串

字符串的存储：

- 顺序存储
- 链式存储
- 索引存储

广义表

广义表：简称表，是线性表的推广，是 n 个表元素组成的序列，表元素可以是数据元素或是子表。当表非空时，称第一个表元素为**表头**，其余表元素为**表尾**

非空广义表的表头可以是数据元素或广义表，但表尾一定是广义表

广义表的性质：

- 有次序性：各表元素在表中以线性排列，每个表元素最多有一个直接前驱，一个直接后继
- 有长度：广义表中元素个数一定，不能是无限的
- 有深度：广义表中括号的层数即为表的深度（表的嵌套层数）
- 可递归：广义表本身可以是自己的子表，具有此性质的表为递归表

- 可共享：广义表可以被其他广义表共享，称为共享表或再入表

递归表 > 再入表 > 纯表 > 线性表

广义表的存储表示

通常用链表表示广义表，每个结点包含三个域，type, data, link，type 标注该元素是原子还是表，data 存储数据/子表指针，link 存储指向下一个表元素的指针

由于广义表的定义是递归的，于是广义表的算法也是递归的：复制/求长度/求深度/判断相等/删除/建立