



第四章 数组、串与广义表

本章内容：

4.1 一维数组与多维数组

4.2 特殊矩阵

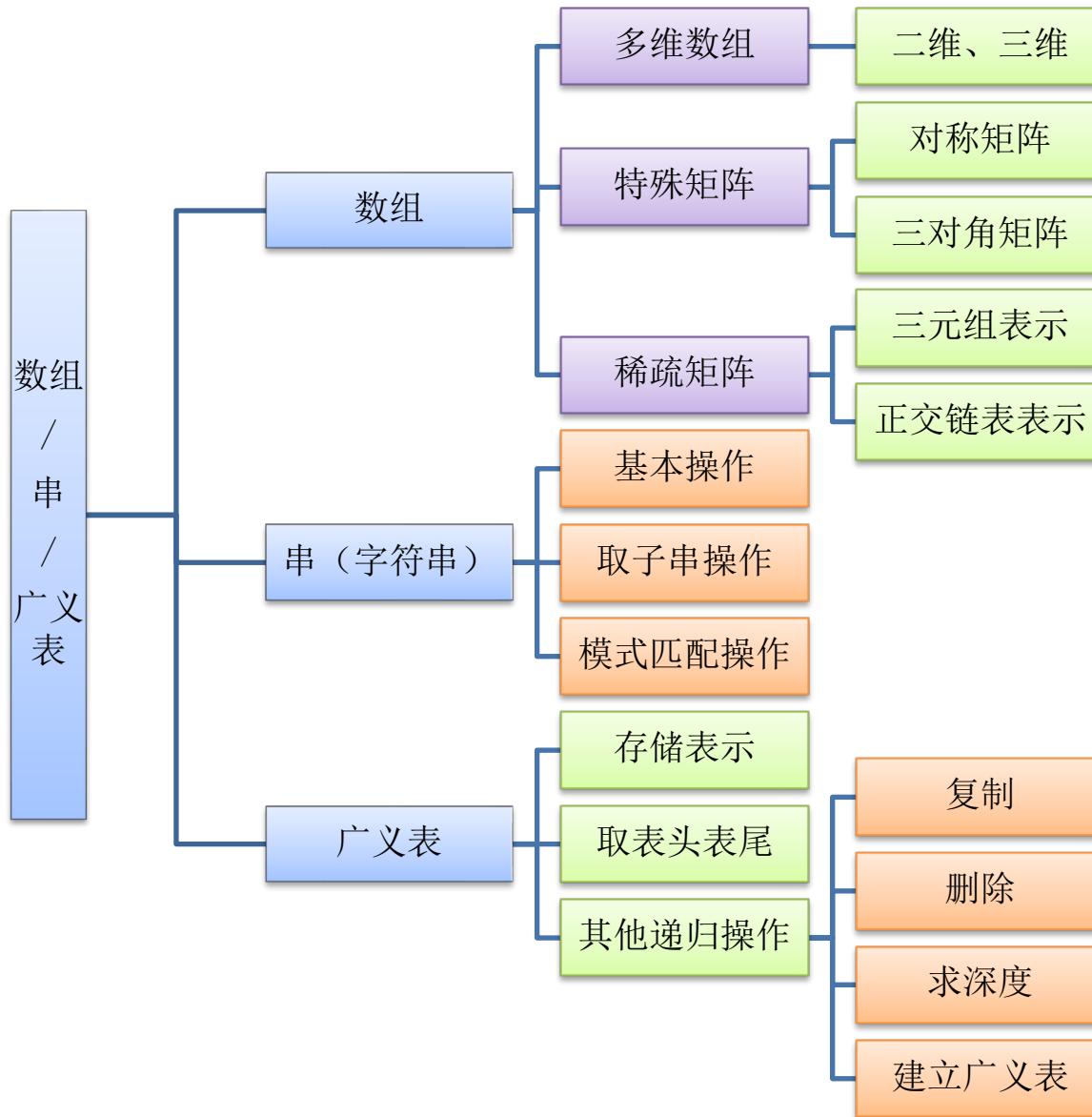
4.3 稀疏矩阵

4.4 字符串

4.5 广义表



知识导图





第四章 数组、串与广义表

本章内容：

4.1 一维数组与多维数组

4.2 特殊矩阵

4.3 稀疏矩阵

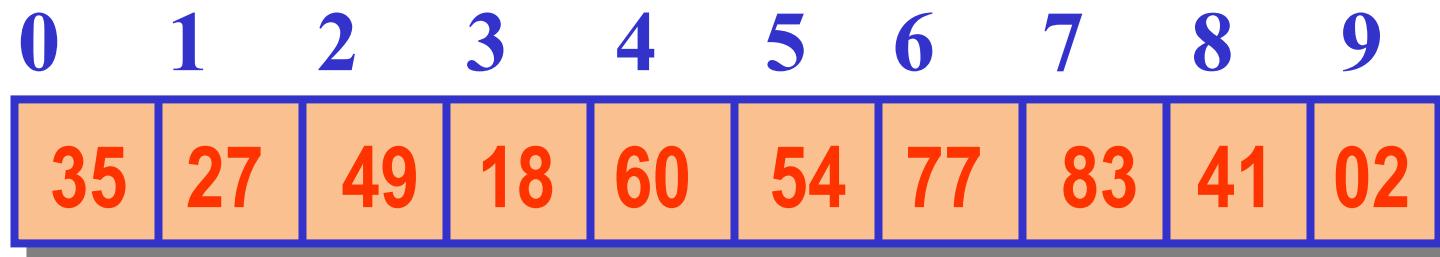
4.4 字符串

4.5 广义表



一维数组

- **定义：**数组是相同类型的数据元素的集合，而一维数组的每个数组元素是一个序对，由下标（index）和值（value）组成。
- 一维数组的示例



- 在高级语言中的一维数组只能按元素的下标直接存取数组元素的值。



一维数组的定义和初始化

```
#include <iostream.h>
main () {
    int a[3] = { 3, 5, 7 }, *elem, i; //静态数组
    for (i = 0; i < 3; i++)
        cout << a[i] << endl;
    elem = new int[3]; //动态数组
    for (i = 0; i < 3; i++)
        cin >> elem[i];
    while (elem)
        { cout << *elem << endl; elem++; }
}
```



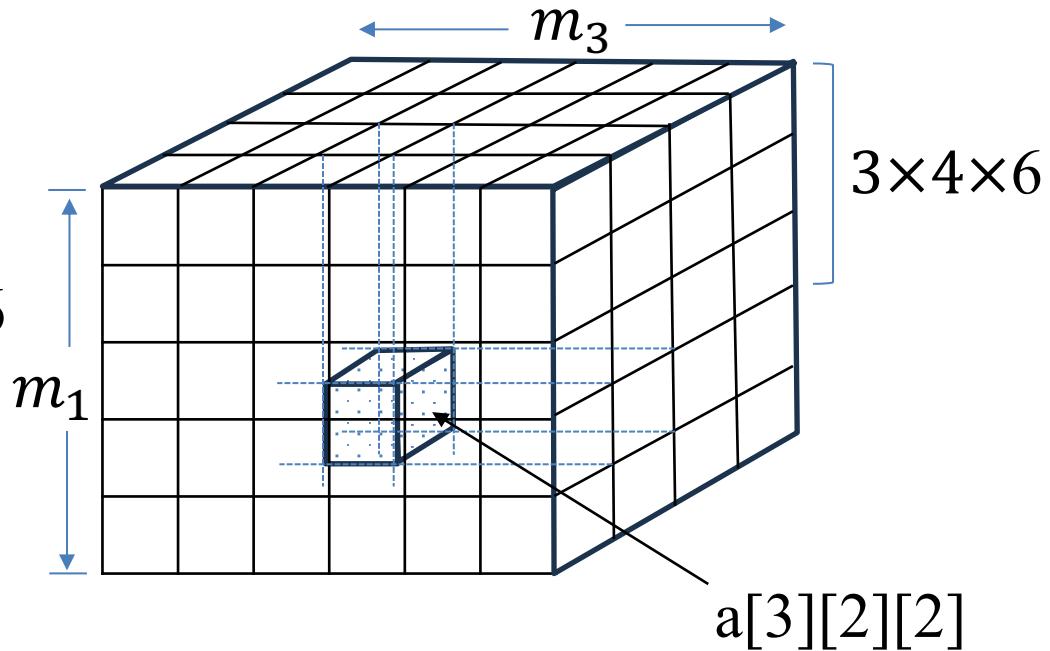
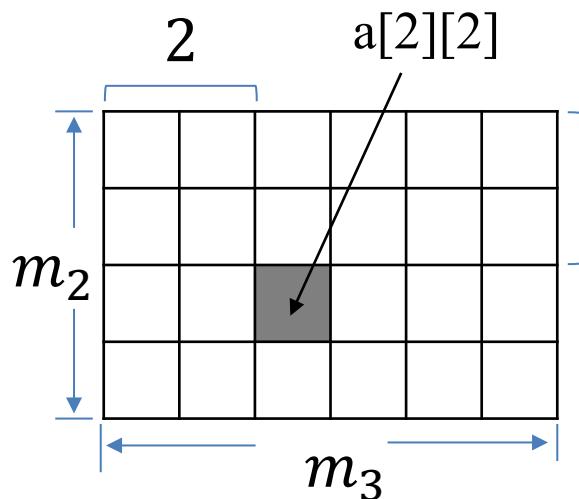
多维数组

- 多维数组是一维数组的推广。
- 多维数组的特点是每一个数据元素可以有多个直接前驱和多个直接后继。
- 数组元素的下标一般具有固定的下界和上界，因此它比其他复杂的非线性结构简单。



多维数组

$$m_1 = 5 \ m_2=4 \ m_3=6$$



二维数组：

- 行向量 下标 i
- 列向量 下标 j

三维数组：

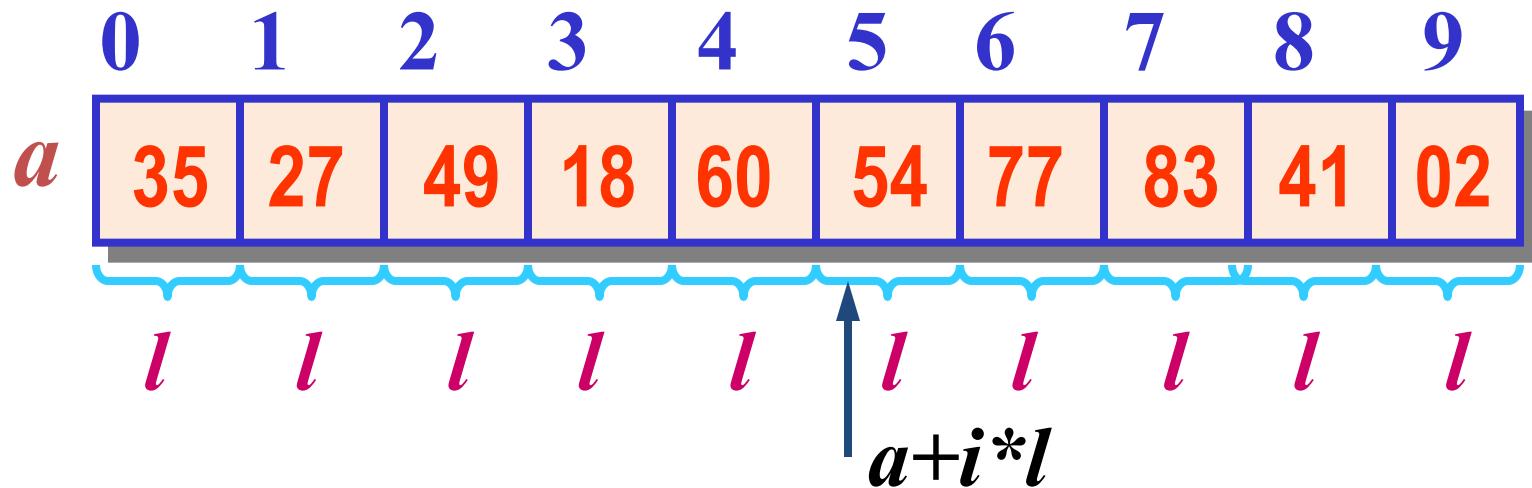
- 页向量 下标 i
- 行向量 下标 j
- 列向量 下标 k



数组的连续存储方式

一维数组

$$\text{LOC}(i) = \begin{cases} a, & i = 0 \\ \text{LOC}(i-1) + l = a + i * l, & i > 0 \end{cases}$$





二维数组

- 一维数组常被称为**向量（Vector）**。
- 二维数组 $A[n][m]$ 可看成是由 n 个**行向量**组成的向量，也可看成是由 m 个**列向量**组成的向量。
- 一个二维数组类型可以定义为其分量类型为**一维数组类型的一维数组类型**：

```
typedef T array2[n][m]; //T为元素类型
```

等价于：

```
typedef T array1[m]; //行向量类型
```

```
typedef array1 array2[n]; //二维数组类型
```



静态定义的数组：

- 其维数和维界不再改变，在**编译时**静态分配存储空间。一旦数组空间用完则不能扩充。

动态定义的数组：

- 其维界不在说明语句中显式定义，而是在**程序运行中**创建数组对象时通过 **new** 动态分配和初始化，在对象销毁时通过 **delete** 动态释放。



二维数组的动态定义和初始化

```
#include <iostream.h>
```

```
.....
```

```
int **A;
```

```
int row = 3, col = 3; int i, j;
```

```
A = new int *[row];
```

```
for (i = 0; i < row; i++)
```

```
    A[i] = new int [col];
```

```
for (i = 0; i < row; i++)
```

```
    for (j = 0; j < col; j++) cin >> A[i][j];
```

```
.....
```

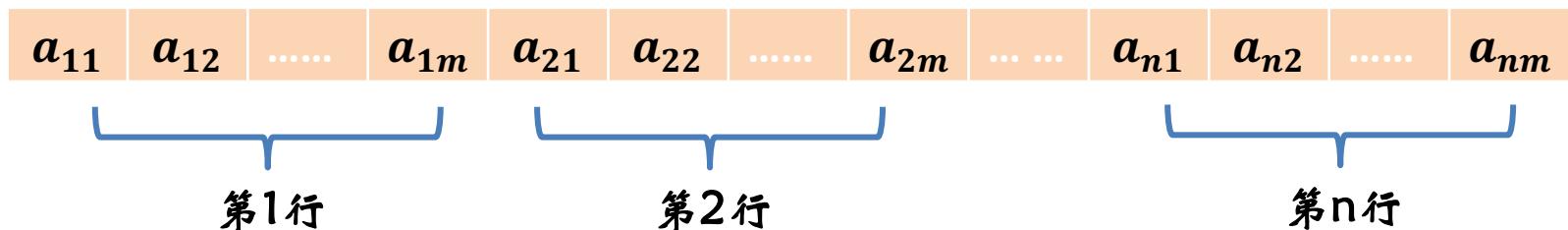


$$a = \begin{pmatrix} a[0][0] & a[0][1] & \cdots & a[0][m-1] \\ a[1][0] & a[1][1] & \cdots & a[1][m-1] \\ a[2][0] & a[2][1] & \cdots & a[2][m-1] \\ \vdots & \vdots & \ddots & \vdots \\ a[n-1][0] & a[n-1][1] & \cdots & a[n-1][m-1] \end{pmatrix}$$

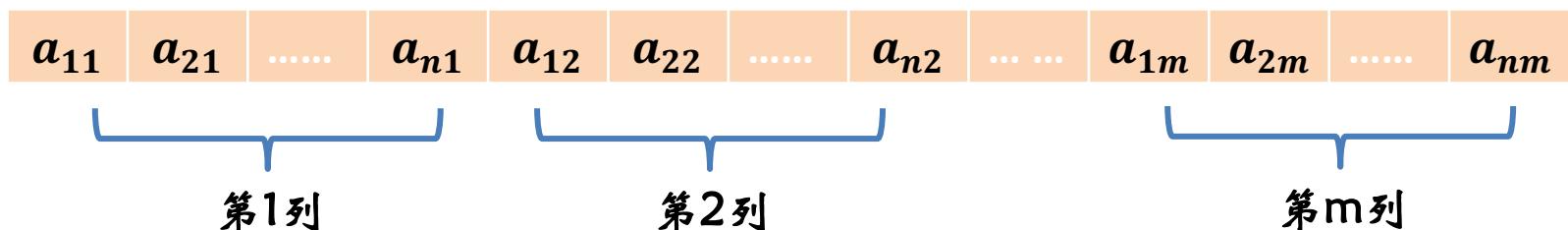
二维数组a在存储器中有两种地址映射方式：
行优先、列优先



- 二维数组在一维地址空间（存储器）中的
行优先存储



- 二维数组在一维地址空间（存储器）中的
列优先存储





二维数组中数组元素的顺序存放

$$a = \begin{pmatrix} a[0][0] & a[0][1] & \cdots & a[0][m-1] \\ a[1][0] & a[1][1] & \cdots & a[1][m-1] \\ a[2][0] & a[2][1] & \cdots & a[2][m-1] \\ \vdots & \vdots & \ddots & \vdots \\ a[n-1][0] & a[n-1][1] & \cdots & a[n-1][m-1] \end{pmatrix}$$

行优先存放：

设数组开始存放位置 $\text{LOC}(0, 0) = a$, 每个元素占用 l 个存储单元

$$\text{LOC}(j, k) = a + (j * m + k) * l$$



二维数组中数组元素的顺序存放

$$a = \begin{pmatrix} a[0][0] & a[0][1] & \cdots & a[0][m-1] \\ a[1][0] & a[1][1] & \cdots & a[1][m-1] \\ a[2][0] & a[2][1] & \cdots & a[2][m-1] \\ \vdots & \vdots & \ddots & \vdots \\ a[n-1][0] & a[n-1][1] & \cdots & a[n-1][m-1] \end{pmatrix}$$

列优先存放：

设数组开始存放位置 $\text{LOC}(0, 0) = a$, 每个元素占用 l 个存储单元

$$\text{LOC}(j, k) = a + (k * n + j) * l$$



三维数组

- 各维元素个数为 m_1, m_2, m_3
- 下标为 i_1, i_2, i_3 的数组元素的存储地址：
(按页 / 行 / 列存放)

$$\text{LOC}(i_1, i_2, i_3) = a + \underbrace{(i_1 * m_2 * m_3 + i_2 * m_3 + i_3) * l}_{\begin{array}{l} \text{前 } i_1 \text{ 页} \\ \text{总元素} \\ \text{个数} \end{array}} \quad \underbrace{\quad}_{\begin{array}{l} \text{第 } i_1 \text{ 页} \\ \text{前 } i_2 \text{ 行} \\ \text{总元素} \\ \text{个数} \end{array}} \quad \underbrace{l}_{\begin{array}{l} \text{第 } i_2 \text{ 行} \\ \text{前 } i_3 \text{ 列} \\ \text{元素个} \\ \text{数} \end{array}}$$



n维数组

- 各维元素个数为 $m_1, m_2, m_3, \dots, m_n$
- 下标为 $i_1, i_2, i_3, \dots, i_n$ 的数组元素的存储地址：

$$\text{LOC}(i_1, i_2, \dots, i_n) = a +$$

$$(i_1 * m_2 * m_3 * \dots * m_n + i_2 * m_3 * m_4 * \dots * m_n + \\ + \dots + i_{n-1} * m_n + i_n) * l$$

$$= a + \left(\sum_{j=1}^{n-1} i_j * \prod_{k=j+1}^n m_k + i_n \right) * l$$



第四章 数组、串与广义表

本章内容：

4.1 一维数组与多维数组

4.2 特殊矩阵

4.3 稀疏矩阵

4.4 字符串

4.5 广义表



特殊矩阵

- 特殊矩阵是指非零元素或零元素的分布有一定规律的矩阵。
 - ◆ 对称矩阵
 - ◆ 三对角矩阵
- 利用特殊矩阵的一些性质寻找特殊的存储方法，可以节省大量的存储空间和计算时间。



对称矩阵的压缩存储

设有一个 $n \times n$ 的对称矩阵 A 。

$$A = \begin{bmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0n-1} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1n-1} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-10} & a_{n-11} & a_{n-12} & \cdots & a_{n-1n-1} \end{bmatrix}$$

对称矩阵中的元素关于主对角线对称，
 $a_{ij} = a_{ji}, \quad 0 \leq i, j \leq n-1$



对称矩阵的压缩存储

为节约存储，只存对角线及对角线以上的元素，或者只存对角线或对角线以下的元素。前者称为上三角矩阵，后者称为下三角矩阵。

$$\begin{bmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0n-1} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1n-1} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-10} & a_{n-11} & a_{n-12} & \cdots & a_{n-1n-1} \end{bmatrix} \quad \text{下三角矩阵}$$



对称矩阵的压缩存储

a_{00}	a_{01}	a_{02}	...	a_{0n-1}
a_{10}	a_{11}	a_{12}	...	a_{1n-1}
a_{20}	a_{21}	a_{22}	...	a_{2n-1}
...
a_{n-10}	a_{n-11}	a_{n-12}	...	a_{n-1n-1}

上三角矩阵

- ▶ 把它们按行存放于一个一维数组 B 中，称之为对称矩阵 A 的压缩存储方式。
- ▶ 数组 B 共有 $n + (n - 1) + \dots + 1 = n*(n+1) / 2$ 个元素。



$$\begin{bmatrix} a_{00} & a_{01} & a_{02} & \cdots & a_{0n-1} \\ a_{10} & a_{11} & a_{12} & \cdots & a_{1n-1} \\ a_{20} & a_{21} & a_{22} & \cdots & a_{2n-1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n-10} & a_{n-11} & a_{n-12} & \cdots & a_{n-1n-1} \end{bmatrix}$$

下
三
角
矩
阵

0	1	2	3	4	5	6	7	8	n(n+1)/2-1		
B	a_{00}	a_{10}	a_{11}	a_{20}	a_{21}	a_{22}	a_{30}	a_{31}	a_{32}	a_{n-1n-1}

若 $i \geq j$, 数组元素 $A[i][j]$ 在数组B中的存放位置为
 $1 + 2 + \cdots + i + j = (i + 1) * i / 2 + j$

前*i*行元素总数

第*i*行第*j*个元素前元素个数



对称矩阵的压缩存储

- 若 $i < j$ ，数组元素 $A[i][j]$ 在矩阵的上三角部分，在数组 B 中没有存放，可以找它的对称元素 $A[j][i] = j * (j + 1) / 2 + i$ 。
- 若已知某矩阵元素位于数组 B 的第 k 个位置 ($k \geq 0$)，可寻找满足

$$i(i + 1) / 2 \leq k < (i + 1)*(i + 2) / 2$$

的 i ，此即为该元素的行号。

$$j = k - i * (i + 1) / 2$$

此即为该元素的列号。

例，当 $k = 8$, $3*4 / 2 = 6 \leq k < 4*5 / 2 = 10$,

取 $i = 3$ 。则 $j = 8 - 3*4 / 2 = 2$ 。



$n = 4$

a_{00}	a_{01}	a_{02}	a_{03}
a_{10}	a_{11}	a_{12}	a_{13}
a_{20}	a_{21}	a_{22}	a_{23}
a_{30}	a_{31}	a_{32}	a_{33}

上三角矩阵

0 1 2 3 4 5 6 7 8 9

B

a_{00}	a_{01}	a_{02}	a_{03}	a_{11}	a_{12}	a_{13}	a_{22}	a_{23}	a_{33}
----------	----------	----------	----------	----------	----------	----------	----------	----------	----------

若 $i \leq j$ ，数组元素 $A[i][j]$ 在数组 B 中的存放位置
为 $n + (n-1) + (n-2) + \cdots + (n-i+1) + j - i$

前 i 行元素总数

第 i 行第 j 个元素前元素个数



- 若 $i \leq j$, 数组元素 $A[i][j]$ 在数组 B 中的存放位置为

$$\begin{aligned} & n + (n-1) + (n-2) + \cdots + (n-i+1) + j - i = \\ & = (2*n - i + 1) * i / 2 + j - i = \\ & = (2*n - i - 1) * i / 2 + j \end{aligned}$$

- 若 $i > j$, 数组元素 $A[i][j]$ 在矩阵的下三角部分, 在数组 B 中没有存放。因此, 找它的对称元素 $A[j][i]$ 。 $A[j][i]$ 在数组 B 的第 $(2*n - j - 1) * j / 2 + i$ 的位置中找到。



三对角矩阵的压缩存储

$$A = \begin{bmatrix} a_{00} & a_{01} & 0 & 0 & 0 & 0 \\ a_{10} & a_{11} & a_{12} & 0 & 0 & 0 \\ 0 & a_{21} & a_{22} & a_{23} & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & a_{n-2n-3} & a_{n-2n-2} & a_{n-2n-1} \\ 0 & 0 & 0 & 0 & a_{n-1n-2} & a_{n-1n-1} \end{bmatrix}$$

0 1 2 3 4 5 6 7 8 9 10

$$\mathbf{B} = \begin{matrix} & a_{00} & a_{01} & a_{10} & a_{11} & a_{12} & a_{21} & a_{22} & a_{23} & \dots & a_{n-1 n-2} & a_{n-1 n-1} \end{matrix}$$

$$\overbrace{2}^2 \quad \overbrace{3}^3 \quad \overbrace{3}^3 \quad \dots \quad \overbrace{\begin{array}{c} 2 \\ \text{加法} \end{array}}^2 \quad \text{加法}$$



三对角矩阵的压缩存储

- 三对角矩阵中总共有 $3n-2$ 个非零元素。

只有主对角线及在主对角线上下最近的两条对角线上的元素外是非0元素。

- 将三对角矩阵A中三条对角线上的元素按行存放在一维数组 B 中，且 a_{00} 存放于B[0]。
- 在三条对角线上的元素 a_{ij} 满足

$$0 \leq i \leq n-1, \quad i-1 \leq j \leq i+1$$

- 矩阵A中元素 A[i][j] 在 B 中位置为 $k = 2*i + j - i*(3-1+j-i+1)$ 。



三对角矩阵的压缩存储

- 若已知三对角矩阵中某元素 $A[i][j]$ 在数组 $B[]$ 存放于第 k 个位置，则有

$$i = \lfloor (k + 1) / 3 \rfloor$$

$$j = k - 2 * i$$

- 例：

当 $k = 8$ 时，

$$i = \lfloor (8+1) / 3 \rfloor = 3, j = 8 - 2*3 = 2$$

当 $k = 10$ 时，

$$i = \lfloor (10+1) / 3 \rfloor = 3, j = 10 - 2*3 = 4$$



特殊矩阵的例子

- 对称矩阵：

- 机器学习中，对称矩阵用于表示协方差矩阵，描述不同变量之间的相关性。
- 用于无向图表示中，节点和节点之间的关系。

- 三对角矩阵：

- 如在求解差分方程、热传导方程和振动方程等物理问题时，经常会遇到三对角矩阵方程。



第四章 数组、串与广义表

本章内容：

4.1 一维数组与多维数组

4.2 特殊矩阵

4.3 稀疏矩阵

4.4 字符串

4.5 广义表



稀疏矩阵 (Sparse Matrix)

$$A_{6 \times 7} = \begin{pmatrix} 0 & 0 & 0 & 22 & 0 & 0 & 15 \\ 0 & 11 & 0 & 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & -6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 39 & 0 \\ 91 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 28 & 0 & 0 & 0 & 0 \end{pmatrix}$$

- 设矩阵 A 中有 s 个非零元素，若 s 远远小于矩阵元素的总数（即 $s \leq m \times n$ ），则称 A 为稀疏矩阵。



稀疏矩阵 (Sparse Matrix)

- 设矩阵 A 中有 s 个非零元素。令 $e = s/(m * n)$, 称 e 为矩阵的稀疏因子（通常认为 $e \leq 0.05$ 时称之为稀疏矩阵）。
- 在存储稀疏矩阵时，为节省存储空间，应只存储非零元素。
- 稀疏矩阵的两种表示方法：
 - 三元组表示（数组存储）
 - ✓ 对于非零元素，每一个三元组记下它所在的行和列的位置 (i, j), 以及值 a_{ij}
 - ✓ 三元组 (i, j, a_{ij}) 唯一确定了矩阵 A 的一个非零元素
 - 正交链表表示（链表存储）



稀疏矩阵的定义——三元组表示

```
const int drows = 6, dcols = 7, dterms = 9;  
template<class E>  
struct Triple {           //三元组  
    int row, col;         //非零元素行号/列号  
    E value;              //非零元素的值  
    void operator = (Triple<E>& R) //赋值  
    { row = R.row; col = R.col; value =  
        R.value; }  
};  
template <class E>  
class SparseMatrix {
```



稀疏矩阵的定义——三元组表示

public:

```
SparseMatrix (int Rw = drows, int Cl = dcols,  
             int Tm = dterms); //构造函数  
void Transpose(SparseMatrix<E>& b); //转置  
void Add (SparseMatrix<E>& a,  
          SparseMatrix<E>& b); //a = a+b  
void Multiply (SparseMatrix<E>& a,  
                SparseMatrix<E>& b); //a = a*b
```

private:

```
int Rows, Cols, Terms; //行 / 列 / 非零元素数  
Triple<E> *smArray; //三元组表 数据成员
```

}; 稀疏矩阵可由表示非零元的一系列**三元组**及其**行列数**唯一确定。



稀疏矩阵的构造函数

```
template <class E>
SparseMatrix<E>::SparseMatrix (int Rw, int Cl,
int Tm) {
    Rows = Rw;  Cols = Cl;  Terms = Tm;
    smArray = new Triple[Terms];      //三元组表
    if (smArray == NULL) {
        cerr << “存储分配失败！” << endl;
        exit(1);
    }
};
```



稀疏矩阵的转置

- 一个 $m \times n$ 的矩阵 A , 它的转置矩阵 B 是一个 $n \times m$ 的矩阵, 且 $A[i][j] = B[j][i]$ 。即
 - 矩阵 A 的行成为矩阵 B 的列
 - 矩阵 A 的列成为矩阵 B 的行。
- 存储规则: 在稀疏矩阵的三元组表中, 非零矩阵元素按行存放。当行号相同时, 按列号递增的顺序存放。
- 稀疏矩阵的转置运算要转化为对应三元组表的转置。



稀疏矩阵的转置

稀疏矩阵

$$\begin{pmatrix} 0 & 0 & 0 & 22 & 0 & 0 & 15 \\ 0 & 11 & 0 & 0 & 0 & 17 & 0 \\ 0 & 0 & 0 & -6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 39 & 0 \\ 91 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 28 & 0 & 0 & 0 & 0 \end{pmatrix}$$

行 (row)	列 (col)	值 (value)
[0]	0	22
[1]	0	15
[2]	1	11
[3]	1	17
[4]	2	-6
[5]	3	39
[6]	5	91
[7]	2	28



稀疏矩阵的转置

转置矩阵

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 91 & 0 \\ 0 & 11 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 28 \\ 22 & 0 & -6 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 17 & 0 & 39 & 0 & 0 \\ 15 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

	行 (row)	列 (col)	值 (value)
[0]	0	4	91
[1]	1	1	11
[2]	2	5	28
[3]	3	0	22
[4]	3	2	/6
[5]	5	1	17
[6]	5	3	39
[7]	6	0	15



用三元组表表示的稀疏矩阵及其转置

原矩阵三元组表

	行 (row)	列 (col)	值 (value)
[0]	0	3	22
[1]	0	6	15
[2]	1	1	11
[3]	1	5	17
[4]	2	3	/6
[5]	3	5	39
[6]	4	0	91
[7]	5	2	28

转置矩阵三元组表

	行 (row)	列 (col)	值 (value)
[0]	0	4	91
[1]	1	1	11
[2]	2	5	28
[3]	3	0	22
[4]	3	2	/6
[5]	5	1	17
[6]	5	3	39
[7]	6	0	15



稀疏矩阵转置算法思想

- 设矩阵列数为 **Cols**, 对矩阵三元组表扫描**Cols** 次。第 **k** 次检测列号为 **k** 的项。
- 第 **k** 次扫描找寻所有列号为 **k** 的项，将其行号变列号、列号变行号，顺次存于转置矩阵三元组表。

原矩阵三元组表

	行 (row)	列 (col)	值 (value)
[0]	0	3	22
[1]	0	6	15
[2]	1	1	11
[3]	1	5	17
[4]	2	3	/6
[5]	3	5	39
[6]	4	0	91
[7]	5	2	28

转置矩阵三元组表

	行 (row)	列 (col)	值 (value)
[0]	0	4	91
[1]	1	1	11
[2]	2	5	28
[3]	3	0	22
[4]	3	2	/6
[5]	5	1	17
[6]	5	3	39
[7]	6	0	15



稀疏矩阵的转置

```
template <class E>
void SparseMatrix<E> ::  
Transpose (SparseMatrix<E>& B) {  
    //转置this矩阵，转置结果由B返回  
    B.Rows = Cols;  B.Cols = Rows;  
    B.Terms = Terms;  
    //转置矩阵的列数,行数和非零元素个数  
    if (Terms > 0) {  
        int CurrentB = 0; //转置三元组表存放指针  
        int i, k;
```



```
for (k = 0; k < Cols; k++)      //处理所有列号
    for (i = 0; i < Terms; i++) //第 k 次检测列 k 的项
        if (smArray[i].col == k) {
            B.smArray[CurrentB].row = k;
            B.smArray[CurrentB].col = smArray[i].row;
            B.smArray[CurrentB].value=
                smArray[i].value;
            CurrentB++;
        }
    }
};
```



快速转置算法

- 设矩阵三元组表总共有 t 项，上述算法的时间代价为 $O(n * t)$ 。当非零元素的个数 t 和 $m * n$ 同数量级时，算法transMatrix的时间复杂度为 $O(m * n^2)$ 。
- 若矩阵有 200 行，200 列，10,000 个非零元素，总共有 2,000,000 次处理。
- 快速转置算法的思想为：对原矩阵A 扫描一遍，按 A 中每一元素的列号，立即确定在转置矩阵 B 三元组表中的位置，并装入它。



快速转置算法

- 为加速转置速度，建立辅助数组 **rowSize** 和 **rowStart**:
 - ◆ **rowSize**记录矩阵转置前各列，即转置矩阵各行非零元素个数；
 - ◆ **rowStart**记录各行非零元素在转置三元组表中开始存放位置。
- 扫描矩阵三元组表，根据某项列号，确定它转置后的行号，查 **rowStart** 表，按查到的位置直接将该项存入转置三元组表中。



[0] [1] [2] [3] [4] [5] [6]

语义

rowSize

1 1 1 2 0 2 1

矩阵 A 各列非零元素个数

rowStart

0 1 2 3 5 5 7

矩阵 B 各行开始存放位置

A三元组	(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
行row	0	0	1	1	2	3	4	5
列col	3	6	1	5	3	5	0	2
值value	22	15	11	17	-6	39	91	28



稀疏矩阵的快速转置算法

```
template <class E>
void SparseMatrix<E>::
    FastTranspos (SparseMatrix<E>& B) {
    int *rowSize = new int[Cols];      //列元素数数组
    int *rowStart = new int[Cols];     //转置位置数组
    B.Rows = Cols;   B.Cols = Rows;
    B.Terms = Terms;
    if (Terms > 0) {
        int i, j;
        for (i = 0; i < Cols; i++) rowSize[i] = 0;
```



稀疏矩阵的快速转置算法

```
for (i = 0; i < Terms; i++)
    rowSize[smArray[i].col]++;
rowStart[0] = 0;
for (i = 1; i < Cols; i++)
    rowStart[i] = rowStart[i-1]+rowSize[i-1];
for (i = 0; i < Terms; i++) {
    j = rowStart [smArray[i].col];
    B.smArray[j].row = smArray[i].col;
    B.smArray[j].col = smArray[i].row;
    B.smArray[j].value = smArray[i].value;
    rowStart [smArray[i].col]++; //A中相同列号处理
}
}// end of if
delete [ ] rowSize; delete [ ] rowStart; }
```



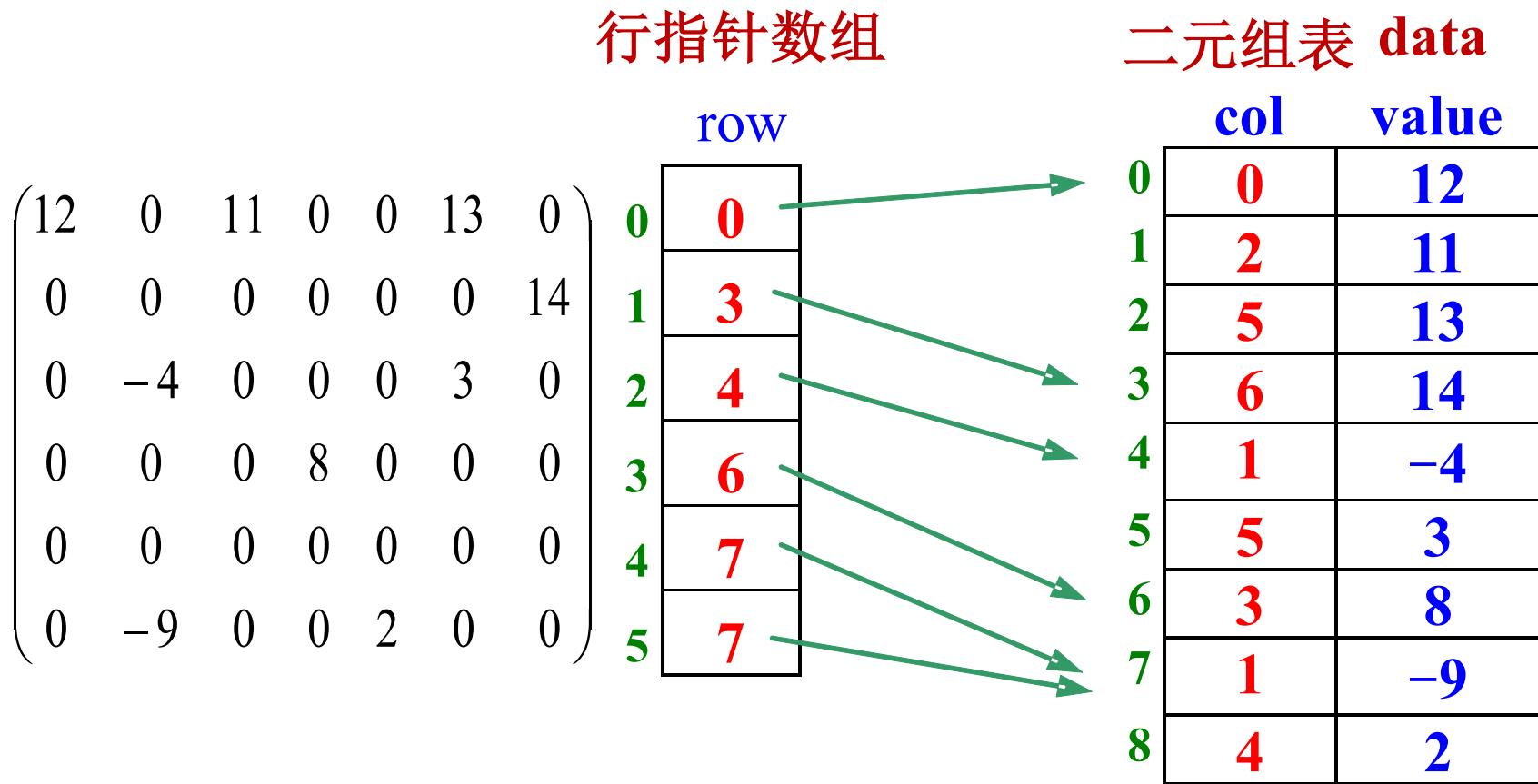
带行指针数组的二元组

- 稀疏矩阵的三元组表可以用**带行指针数组的二元组表代替**。
- 在行指针数组中元素个数与矩阵行数相等。第 i 个元素的下标 i 代表矩阵的第 i 行，**元素的内容**即为稀疏矩阵第 i 行的第一个非零元素在二元组表中的**存放位置**。



带行指针数组的二元组

二元组表中每个二元组只记录非零元素的列号和元素值，且各二元组按列号递增的顺序排列。





用正交链表来表示稀疏矩阵

- 稀疏矩阵表示为行链表与列链表的十字交叉——称为正交链表(十字链表)
- 优点：适应矩阵操作(+、 -、 *)时矩阵非零元素的动态变化



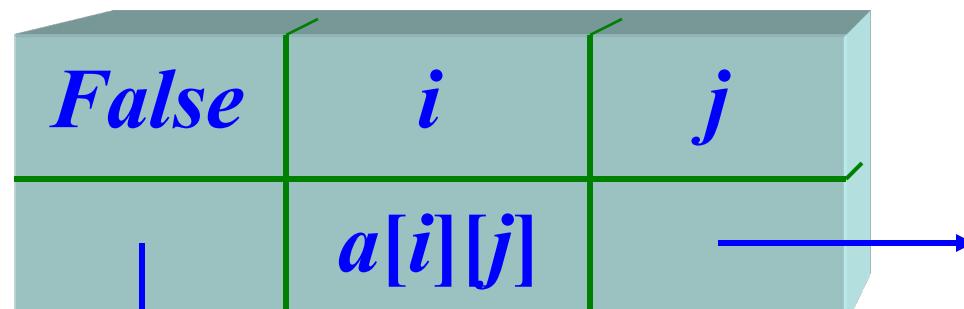
稀疏矩阵的结点

$head$	$next$
$down$	$right$

(a) 表头结点

$head$	row	col
$down$	$value$	$right$

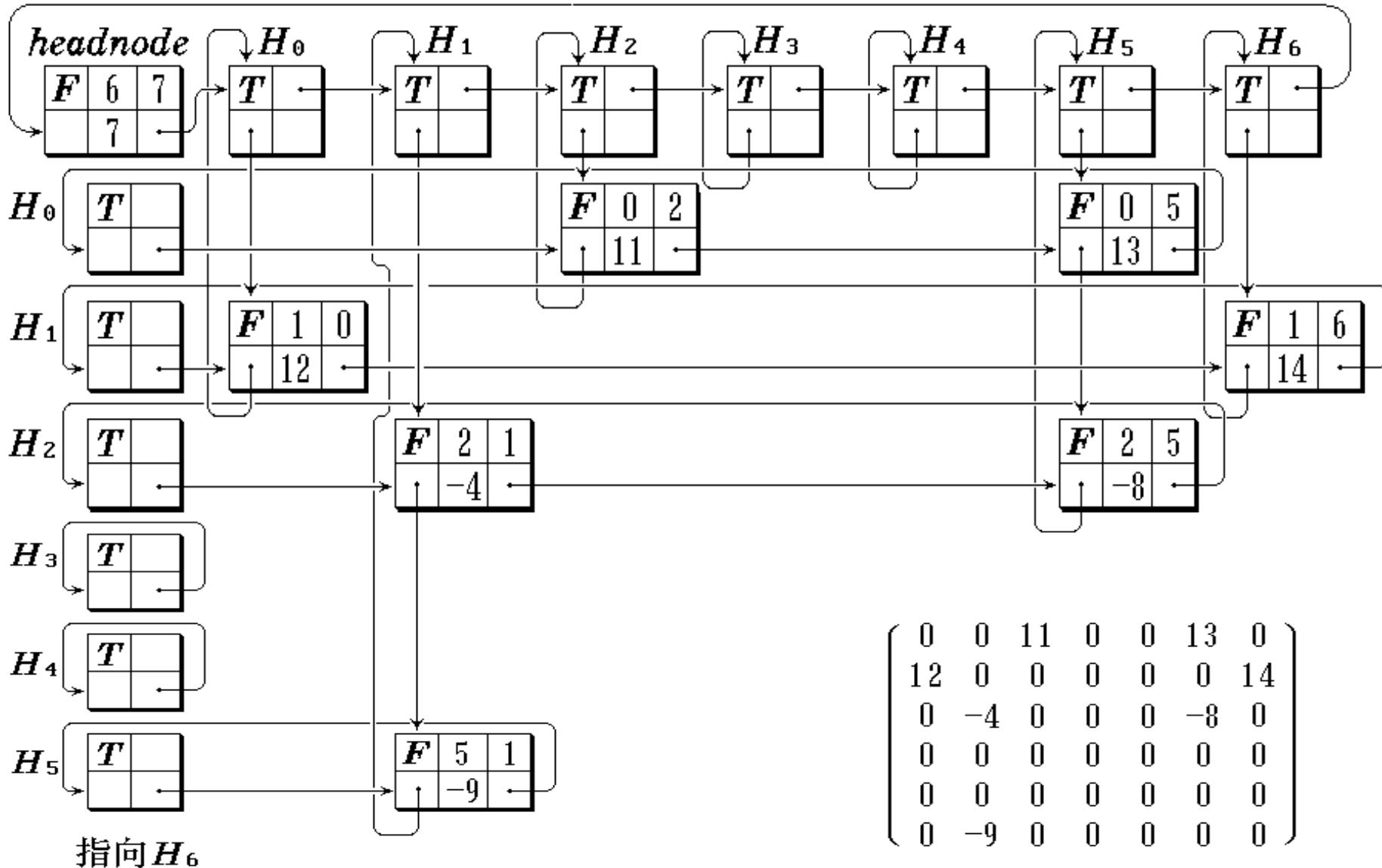
(b) 非零元素结点



(c) 建立 $a[i][j]$ 结点



稀疏矩阵的正交链表表示的示例





第四章 数组、串与广义表

本章内容：

4.1 一维数组与多维数组

4.2 特殊矩阵

4.3 稀疏矩阵

4.4 字符串

4.5 广义表



字符串 (String)

- 字符串是 n (≥ 0) 个字符的有限序列，
记作 $S : "c_1c_2c_3\dots c_n"$
其中， S 是串名字
“ $c_1c_2c_3\dots c_n$ ”是串值
 c_i 是串中字符
 n 是串的长度， $n = 0$ 称为空串。
- 例如， $S : "Nanjing University"$ 。
- 注意： 空串和空白串不同， 例如 “ ” 和 “”
分别表示长度为1的空白串和长度为0的空串。



字符串 (String)

- 串中任意个**连续字符**组成的子序列称为该串的**子串**，包含子串的串相应地称为主串。
- 设A和B分别为

A = “This is a string” B = “is”

则 B 是 A 的子串，A 为主串。B 在 A 中出现了两次，首次出现所对应的主串位置是**2**（从0开始）。因此，称**B 在 A 中的位置为2**。

- 特别地，空串是任意串的子串，任意串是其自身的子串。



字符串的类定义

```
#ifndef ASTRING_H //定义在文件 "Astring.h"  
//  
#define ASTRING_H  
#define defaultSize = 128; //字符串的最大长度  
class AString {  
    //对象: 零个或多个字符的一个有限序列。  
private:  
    char *ch; //串存放数组  
    int curLength; //串的实际长度  
    int maxSize; //存放数组的最大长度  
public:
```



```
AString(int sz = defaultSize);           //构造函数
AString(const char *init );              //构造函数
AString(const AString& ob);             //复制构造函数
~AString() {delete [ ]ch; }             //析构函数
int Length() const { return curLength; } //求长度
AString& operator() (int pos, int len); //求子串
bool operator == (AString& ob) const
{ return strcmp (ch, ob.ch) == 0; }
//判断串相等. 若串相等, 则函数返回true
bool operator != (AString& ob) const
{ return strcmp (ch, ob.ch) != 0; }
//判断串不等. 若串不相等, 则函数返回true
```



```
bool operator !() const { return curLength == 0; }
```

//判串空否。若串空，则函数返回true

```
AString& operator = (AString& ob); //串赋值
```

```
AString& operator += (AString& ob); //串连接
```

```
char& operator [ ] (int i); //取第 i 个字符
```

```
int Find (AString& pat, int k) const; //串匹配
```

```
};
```



字符串的构造函数

```
AString::AString(int sz) {  
    //构造函数：创建一个空串  
    maxSize = sz;  
    ch = new char[maxSize+1];      //创建串数组  
    if (ch == NULL)  
        { cerr << “存储分配错!\n”; exit(1); }  
    curLength = 0;  
    ch[0] = ‘\0’;  
};
```



字符串的构造函数

```
AString::AString(const char *init) {  
    //复制构造函数: 从已有字符数组*init复制  
    int len = strlen(init);  
    maxSize = (len > defaultSize) ? len : defaultSize;  
    ch = new char[maxSize+1];    //创建串数组  
    if (ch == NULL)  
        { cerr << “存储分配错！\n”; exit(1); }  
    curLength = len;            //复制串长度  
    strcpy(ch, init);          //复制串值  
}  
;
```



字符串的复制构造函数

```
AString :: AString(const AString& ob) {  
    //复制构造函数：从已有串ob复制  
    maxSize = ob.maxSize;           //复制串最大长度  
    ch = new char[curLength+1];     //创建串数组  
    if (ch == NULL)  
        { cerr << “存储分配错! \n”; exit(1); }  
    curLength = ob.curLength;        //复制串长度  
    strcpy(ch, ob.ch);              //复制串值  
};
```



字符串重载操作的使用示例

序号	重载操作	操作	使用示例 (设使用操作的当前串为S： 'good')
1	() (int pos, int len)	取子串	S1 = S(1, 3), //S1结果为 'ood'
2	== (const AString& ob)	判两串相等	S == S1 , //若S与S1相等，结果为true，否则为false
3	!= (const AString& ob)	判两串不等	S != S1 , //若S与S1不等，结果为true，否则为false
4	! 0	判串空否	!S , //若串S为空，结果为 true，否则为false



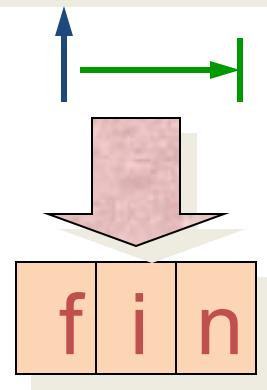
字符串重载操作的使用示例

序号	重载操作	操作	使用示例 (设使用操作的当前串为S : 'good')
5	= (const AString& ob)	串赋值	S1 = S, //S1结果为 'good'
6	+= (const AString& ob)	串连接	若设S1为 ' morning', 执行 S += S1, //S结果为 'good morning'
7	[] (int i)	取第 i 个字符	S[3], //取出字符为 'd'



提取子串的算法示例

pos = 2, len = 3

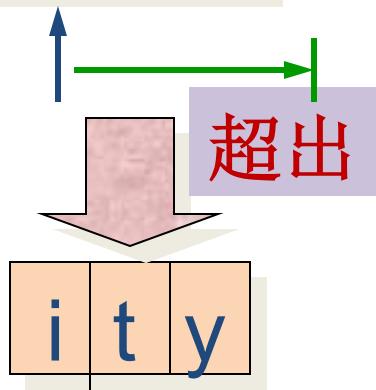


pos+len-1

$\leq \text{curLength}-1$

可以全部提取

pos = 5, len = 4



pos+len -1

$\geq \text{curLength}$

只能从pos取到串尾

需调整Len为 curLength-pos





串重载操作：提取子串

```
AString AString::operator () (int pos, int len) {  
    //从串中第 pos 个位置起连续提取 len 个字符形成  
    //子串返回  
    AString temp;                                //建立空串对象  
    if (pos >= 0 && pos+len-1 < maxSize &&  
len > 0)  
    {                                              //提取子串  
        if (pos+len-1 >= curLength)  
            len = curLength - pos; //调整提取字符数  
        temp.curLength = len;      //子串长度
```



串重载操作：提取子串

```
for (int i = 0, j = pos; i < len; i++, j++)
    temp.ch[i] = ch[j];      //传送串数组
temp.ch[len] = '\0';          //子串结束
}
else
{ temp.curLength = 0;   temp.ch[0] = '\0'; }
return temp;
};
```

- 例：串 `st = “university”`, `pos = 3`, `len = 4`
使用示例 `subSt = st(3, 4)`
提取子串 `subSt = “vers”`



串重载操作: 串赋值

```
AString& AString::operator = (const AString& ob)
{
    if (&ob != this) { //若两个串相等为自我赋值
        delete []ch;
        ch = new char[maxSize+1]; //重新分配
        if (ch == NULL)
            { cerr << “存储分配失败!\n” ; exit(1); }
        curLength = ob.curLength; strcpy(ch,ob.ch);
    }
    else cout << “字符串自身赋值出错!\n”;
    return this;
};
```



串重载操作：取串的第i个字符

```
char AString::operator [ ] (int i) {  
    //串重载操作：取当前串*this的第i个字符  
    if (i < 0 && i >= curLength)  
        { cout << “字符串下标超界!\n”; exit(1); }  
    return ch[i];  
};
```

- 例：串 **st = “university”**,
使用示例 **newSt = st;** **newChar = st[1];**
数组赋值结果 **newSt = “university”**
提取字符结果 **newChar = ‘n’**



串重载操作: 串连接

```
AString& AString::operator += (const AString& ob)
{
    char *temp = ch;                                //暂存原串数组
    int n = curLength + ob.curLength;                //串长度累加
    int m = (maxSize >= n) ? maxSize : n;           //新空间大小
    ch = new char[m];
    if (ch == NULL)
        { cerr << "存储分配错!\n"; exit(1); }
    maxSize = m; curLength = n;
    strcpy(ch, temp);                               //拷贝原串数组
    strcat(ch, ob.ch);                            //连接ob串数组
```



串重载操作: 串连接

```
delete []temp;  
return this;  
};
```

- 例: 串 st1 = “Nanjing ”,
 st2 = “university”,
使用示例 st1 += st2;
连接结果 st1 = “Nanjing university”
 st2 = “university”



串的模式匹配

- 定义 在主串中寻找子串（第一个字符）在串中的位置
- 词汇 在模式匹配中，子串称为模式，主串称为目标。
- 示例 目标 T：“Beijing”
模式 P：“jin”
匹配结果 = 3



朴素的模式匹配（B-F算法）

第1趟

T a b **b** a b a
 P a b **a**

$\downarrow i=2$
 $\uparrow j=2$

第2趟

T a **b** b a b a
 P a **b** a

$\downarrow i=1$
 $\uparrow j=0$

第3趟

T a b **b** a b a
 P a **b** a

$\downarrow i=2$
 $\uparrow j=0$

第4趟

T a b b a b a
 P a b a

$\downarrow i=6$
 $\uparrow j=3$



朴素的模式匹配算法

```
int AString::Find(AString& pat, int k) const {
    //在当前串中从第 k 个字符开始寻找模式 pat 在当
    //前串中匹配的位置。若匹配成功，则函数返回首
    //次匹配的位置，否则返回-1。
    int i, j, n = curLength, m = pat.curLength;
    for (i = k; i <= n-m; i++) {
        for (j = 0; j < m; j++)
            if (ch[i+j] != pat.ch[j]) break; //本次失配
        if (j == m) return i;
            //pat扫描完，匹配成功
    }
    return -1; //pat为空或在*this中找不到它
};
```

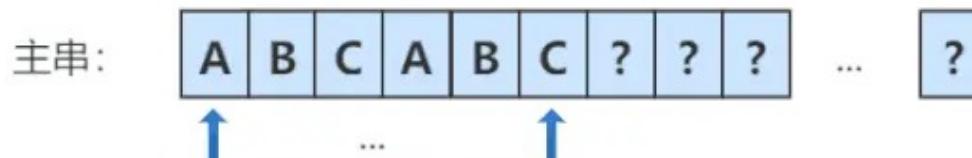


- 在最坏情况下，若设 n 为目标串长度， m 为模式串长度，则匹配算法最多比较 $n-m+1$ 趟，每趟比较都在比较到模式串尾部才出现不等，要做 m 次比较，总比较次数将达到 $(n-m+1)*m$ 。在多数场合下 m 远小于 n ，因此，算法的运行时间为 $O(n*m)$ 。
- 低效的原因在于每趟重新比较时，目标串的检测指针要回退。

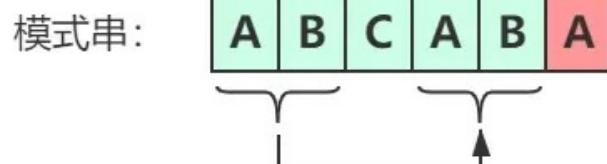


观察

主串地址: 0 1 2 3 4 5 6 7 8 ... n



主串地址: 0 1 2 3 4 5 6 7 8 ... n



直接跳跃到该步骤:



知吾道



改进的模式匹配 (KMP算法)

- 只要消除了每趟失配后为实施下一趟比较时目标指针的**回退/回溯**，可以提高模式匹配效率。
- 这种处理思想是由D.E.Knuth、J.H.Morris和V.R.Pratt同时提出来的，故称为**KMP算法**。
- 实施**KMP算法**的时间代价：
 - 若每趟第一个不匹配，比较 $n-m+1$ 趟，总比较次数最坏达 $(n-m)+m = n$ 。
 - 若每趟第 m 个不匹配，总比较次数最坏亦达到 n 。



改进的模式匹配

目标 T	t_0	t_1	t_2	t_{m-1}	...	t_{n-1}			
	↑	↑	↑				↑			
模式 pat	p_0	p_1	p_2	p_{m-1}					
目标 T	t_0	t_1	t_2	t_{m-1}	t_m	...	t_{n-1}		
	↑	↑			↑	↑		↑		
模式 pat	p_0	p_1	p_{m-2}	p_{m-1}					
目标 T	t_0	t_1	t_i	t_{i+1}	t_{i+m-2}	t_{i+m-1}	...	t_{n-1}
模式 pat	p_0	p_1	p_{m-2}	p_{m-1}					



改进的模式匹配

T $t_0 t_1 \dots t_{s-1} t_s \quad t_{s+1} t_{s+2} \dots t_{s+j-1} t_{s+j} t_{s+j+1} \dots t_{n-1}$
|| || || || || ×

P $p_0 p_1 p_2 \dots p_{j-1} p_j$

则有 $t_s t_{s+1} t_{s+2} \dots t_{s+j-1} = p_0 p_1 p_2 \dots p_{j-1}$ (1)

下一趟，为使模式 P 与目标 T 匹配，必须满足

$p_0 p_1 p_2 \dots p_{j-1} \dots p_{m-1} = t_{s+1} t_{s+2} t_{s+3} \dots t_{s+j} \dots t_{s+m}$

如果 $p_0 p_1 \dots p_{j-2} \neq p_1 p_2 \dots p_{j-1}$ (2)

则立刻可以断定

$p_0 p_1 \dots p_{j-2} \neq t_{s+1} t_{s+2} \dots t_{s+j-1}$

下一趟[后移一位]必不匹配 $p_0 p_1 \dots p_{j-2}$



改进的模式匹配

同样，若 $p_0 p_1 \dots p_{j-3} \neq p_2 p_3 \dots p_{j-1}$

则再下一趟[后移2位]也不匹配，因为有

$p_0 p_1 \dots p_{j-3} \neq t_{s+2} t_{s+3} \dots t_{s+j-1}$

直到对于某一个“ k ”值，使得

$p_0 p_1 \dots p_{k+1} \neq p_{j-k-2} p_{j-k-1} \dots p_{j-1}$

且

$p_0 p_1 \dots p_k = p_{j-k-1} p_{j-k} \dots p_{j-1}$

则

$p_0 p_1 \dots p_k = t_{s+j-k-1} t_{s+j-k} \dots t_{s+j-1}$
 || || ||

$p_{j-k-1} p_{j-k} \dots p_{j-1}$

下一趟可以直接用 p_{k+1} 与 t_{s+j} 继续比较。



k 的确定方法

- Knuth 等人发现，对于不同的 j （失配位置）， k 的取值不同，它仅依赖于模式 P 本身前 j 个字符的构成，与目标无关。
- 可以用一个 next 特征向量来确定：当模式 P 中第 j 个字符与目标 S 中相应字符失配时，模式 P 中应当由哪个字符（设为第 $k+1$ 个）与目标中刚失配的字符重新继续进行比较。

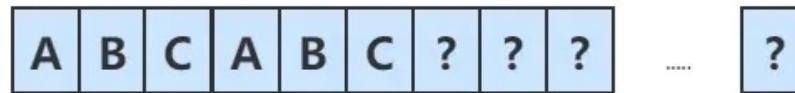


Next数组

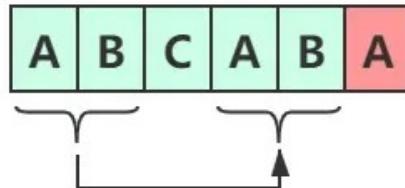
主串地址:

0 1 2 3 4 5 6 7 8 ... n

主串:



模式串:



直接跳跃到该步骤:



next数组的定义为: **next[i]**表示模式串**A[0]至A[i]**这个字串,使得前k个字符（前缀）等于后k个字符（后缀）的最大值。



Next数组

- 设模式 $P = p_0p_1\dots p_{m-2}p_{m-1}$, next定义如下:

$$\text{next}(j) = \begin{cases} -1, & j=0 \\ k+1, & 0 \leq k < j-1 \text{ 且使得 } p_0p_1\dots p_k = \\ & p_{j-k-1}p_{j-k}\dots p_{j-1} \text{ 的最大整数} \\ 0, & \text{其他情况} \end{cases}$$

- 示例

j	0	1	2	3	4	5	6	7
P	a	b	a	a	b	c	a	c
next(j)	-1	0	0	1	1	2	0	1



利用next特征向量进行匹配处理

- 若设若在进行某一趟匹配比较时在模式 P 的第 j 位失配：
 - 如果 $j > 0$, 那么在下一趟比较时模式串 P 的起始比较位置是 $p_{\text{next}(j)}$, 目标串 T 的指针不回溯, 仍指向下一趟失配的字符;
 - 如果 $j = 0$, 则目标串指针 T 进一, 模式串指针 P 回到 p_0 , 继续进行下一趟匹配比较。

运用KMP算法的匹配过程

第1趟 目标 $a \textcolor{red}{c} a b a a b a a b c a c a a b c$
模式 $a \textcolor{red}{b} a a b c a c$

$\times \ j=1 \Rightarrow \text{next}(1) = 0$, 下次 p_0

第2趟 目标 $a \textcolor{red}{c} a b a a b a a b c a c a a b c$
模式 $a b a a b c a c$

$\times \ j=0 \Rightarrow$ 下次 p_0 , 目标指针进 1

第3趟 目标 $a c a b a a b \textcolor{red}{a} a b c a c a a b c$
模式 $a b a a b \textcolor{red}{c} a c$

$\times \ j=5 \Rightarrow \text{next}(5) = 2$,
下次 p_2

第4趟 目标 $a c a b a a b \textcolor{red}{a} a b c a c a a b c$
模式 $(a b) \textcolor{red}{a} a b c a c \checkmark$



用KMP算法实现的快速匹配算法

```
int AString::fastFind(AString& pat, int k,  
    int next[]) const {
```

//从 k 开始寻找 pat 在 this 串中匹配的位置。若找到，函数返
//回 pat 在 this 串中开始位置，否则函数返回-1。
//数组next[] 存放 pat 的next[j] 值。

```
    int posP = 0, posT = k;
```

//两个串的扫描指针

```
    int lengthP = pat.curLength;
```

//模式串长度

```
    int lengthT = curLength;
```

//目标串长度

```
    while (posP < lengthP && posT < lengthT)
```

```
        if (posP == -1 || pat.ch[posP] == ch[posT])
```

//对应字符匹配

```
            { posP++; posT++; }
```

//求pat下趟比较位置

```
        else posP = next[posP];
```

//匹配失败

```
    if (posP < lengthP) return -1;
```

//匹配成功

```
    else return posT-lengthP;
```

```
}
```



此算法的时间复杂度取决于 `while` 循环。由于是无回溯的算法，执行循环时，**目标串字符比较有进无退**，要么执行 `posT` 和 `posP` 进 1，要么查找 `next[]` 数组进行模式位置的右移，然后继续向后比较。**字符的比较次数最多为 $O(\text{lengthT})$** ，不超过目标的长度。



next特征向量的计算

- 设模式 $P = p_0 p_1 p_2 \dots p_{m-1}$ 由 m 个字符组成，而 next 特征向量为 $\text{next} = n_0 n_1 n_2 \dots n_{m-1}$ ，表示了模式的字符分布特征。
- 目标串与模式串是同一个串的模式匹配。
- next 特征向量从 $0, 1, 2, \dots, m-1$ 逐项递推计算：
 - ① 当 $j=0$ 时， $n_0 = -1$ 。
设 $j > 0$ 时 $n_{j-1} = k$ ：
 - ② 当 $k = -1$ 或 $j > 0$ 且 $p_{j-1} = p_k$ ，则 $n_j = k+1$ 。
 - ③ 当 $p_{j-1} \neq p_k$ 且 $k \neq -1$ ，令 $k = n_k$ ，并让 ③ 循环直到条件不满足。
 - ④ 当 $p_{j-1} \neq p_k$ 且 $k = -1$ ，则 $n_j = 0$ 。

② 当 $k = -1$ 或 $j > 0$ 且 $p_{j-1} = p_k$, 则 $n_j = k+1$ 。

以前面的例子说明：

j	0	1	2	3	4	5	6	7
P	a	b	a	a	b	c	a	c
next [j]	-1	0	0	1	1	2	0	1
$j=0$	$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$	$j=7$	
$n_0 = -1$	$k = -1$	$k = 0$	$k = 0$	$k = 1$	$k = 1$	$k = 2$		$k = 0$
$n_1 =$ $=k+1$ $=0$	$p_1 \neq p_0$ $k = n_k =$ $=-1$	$p_2 = p_0$ $k = n_k = 0$ $=k+1$	$p_3 \neq p_1$ $p_3 = p_0$ $=k+1$	$p_4 = p_1$ $n_5 =$ $=k+1$	$p_5 \neq p_2$ $k = n_k = 0$ $=k+1$	$p_6 = p_0$ $p_5 \neq p_0$ $k = n_k = -1$		$n_7 = k+1 = 1$
$n_2 =$ $=k+1$ $=0$		$n_3 =$ $p_3 = p_0$ $=k+1$	$n_4 =$ $=k+1$	$n_5 =$ $=2$		$n_6 = k+1 = 0$		



对当前串计算next特征向量的算法

```
void AString::getNext(int next[]) {  
    int j = 0, k = -1, lengthP = curLength;  
    next[0] = -1;  
    while (j < lengthP)          //计算next[j]  
        if ( k == -1 || ch[j] == ch[k] ) {  
            j++; k++;  
            next[j] = k;  
        }  
        else k = next[k];  
};
```



第四章 数组、串与广义表

本章内容：

4.1 一维数组与多维数组

4.2 特殊矩阵

4.3 稀疏矩阵

4.4 字符串

4.5 广义表

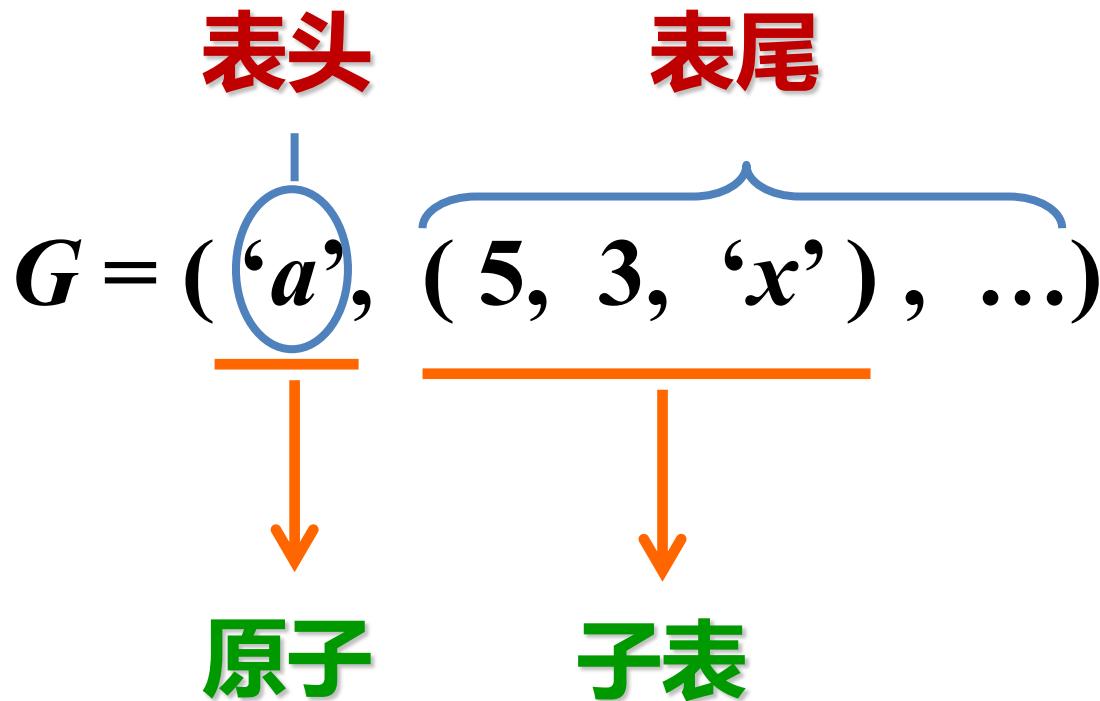


广义表 (General Lists)

- 广义表是 $n (\geq 0)$ 个表元素组成的有限序列，记作
$$LS (a_1, a_2, a_3, \dots, a_n)$$
- LS 是表名， a_i 是表元素，可以是表（称为子表），可以是数据元素（称为原子）。
- n 为表的长度。 $n = 0$ 的广义表为空表。



广义表 (General Lists)



$n > 0$ 时，表的第一个**表元素**称为广义表的**表头 (head)**，除此之外，其它表元素组成的表称为广义表的**表尾 (tail)**。



广义表的特性

- 有次序性
- 有深度
- 可递归
- 有长度
- 可共享

A()

A长度为0，深度为1

B(6, 2)

B长度为2，深度为1

C(‘a’, (5, 3, ‘x’))

C长度为2，深度为2

D(B, C, A)

D长度为3，深度为3

E(B, D)

E长度为？ 深度为？

F(4, F)

F长度为？ 深度为？



广义表的表头(head)与表尾(tail)

- 广义表的第一个表元素即为该表的表头，除表头元素外其他表元素组成的表即为该表的表尾。

A()	head(A) 和 tail(A) 不存在
B(6, 2)	head(B) = 6, tail(B) = (2)
C(‘a’, (5, 3, ‘x’))	head(C) =‘a’
D(B, C, A)	tail(C) = ((5,3,’x’))
E(B, D)	head(((5,3,’x’))) = (5,3,’x’)
F(4, F)	tail(((5,3,’x’))) = ()



A

空表

B

u v

线性表

C

a

纯表

E

B

u v

再入表

D

a

C

x y z

D

B

u

v

a

x

y

z

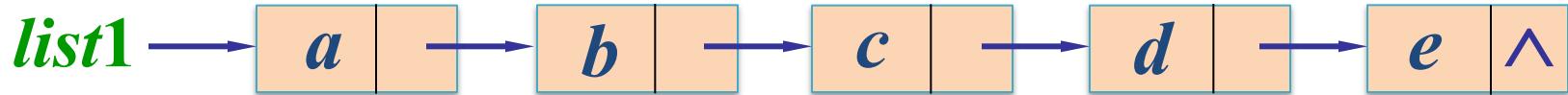
F

d

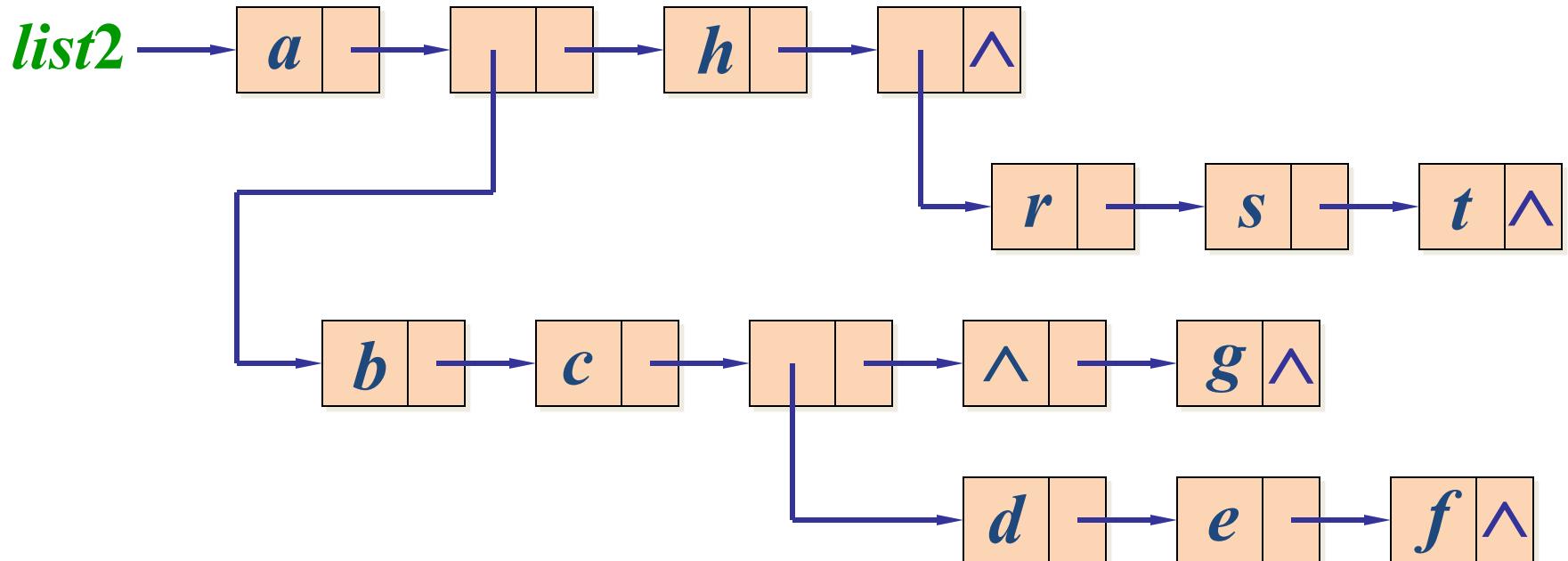
递归表



广义表的表示



$$list1 = (a, b, c, d, e)$$



$$list2 = (a, (b, c, (d, e, f), (), g), h, (r, s, t))$$



广义表结点定义

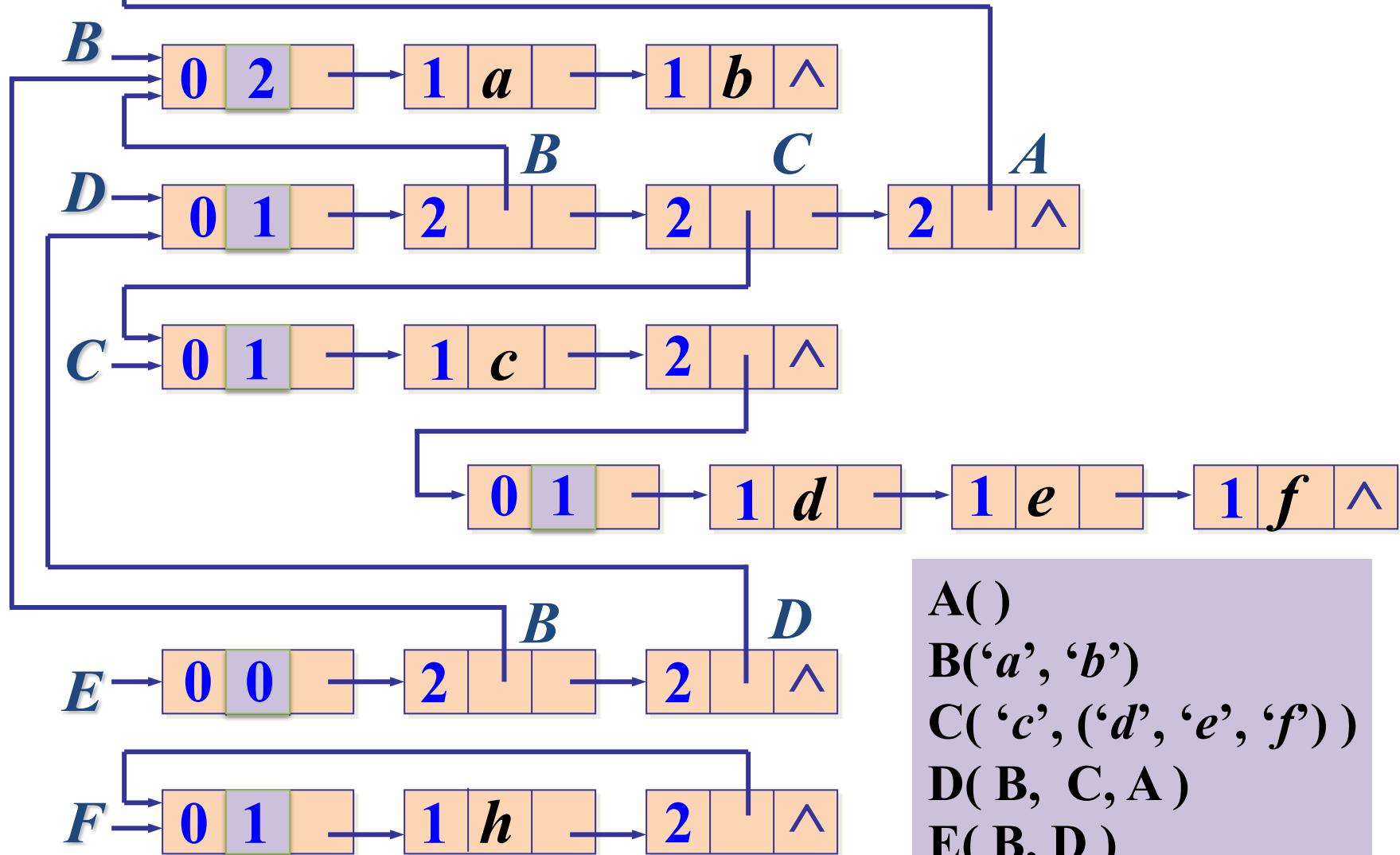
utype

info

tlink

- 结点类型 **utype**: = 0, 表头; = 1, 原子数据;
= 2, 子表
- 信息**info**: utype = 0时, 存放引用计数(ref);
utype = 1时, 存放数据值(value); utype = 2
时, 存放指向子表表头的指针(hlink)
- 尾指针**tlink**: utype = 0时, 指向该表第一个
结点; utype ≠ 0时, 指向同一层下一个结点

广义表的存储表示



A()
B('a', 'b')
C('c', ('d', 'e', 'f'))
D(B, C, A)
E(B, D)
F('h', F)



广义表的类定义

```
#include <iostream.h>
#include <stdlib.h>
template <class T>
struct Items {      //返回值的类结构定义
    int utype;        //结点类型 = 0/1/2
    union {           //联合
        int ref;         //utype=0,存放引用计数
        T value;          //utype=1,存放数值
        GenListNode<T> *hlink;
                           //utype=2,存放指向子表的指针
    } info;
```



```
Items() : utype(0), info.ref(0) {} //构造函数
Items(Items<T>& R)           //复制构造函数
{ utype = R.utype; info = R.info; }
};

template <class T>
struct GenListNode {           //广义表结点类定义
    int utype;                // = 0 / 1 / 2
    union {
        int ref;              //存放引用计数
        T value;              //存放数值
        GenListNode<T> *hlink; //存放子表指针
    } info;
    GenListNode<T> *tlink;    //同层下一结点指针
};
```



```
GenListNode()           //构造函数
    : utype(0), tlink(NULL), info.ref(0) {}
GenListNode(GenListNode<T>& R) {
//复制构造函数
    utype = R.utype;  tlink = R.tlink;
    info = R.info;
}
};

template <class T>
class GenList {           //广义表类定义
public:
```



```
Genlist();           //构造函数
~GenList();          //析构函数
bool Head (Items<T>& x);    //x 返回表头元素
bool Tail (GenList<T>& lt); //lt 返回表尾
GenListNode<T> *First();      //返回第一个元素
GenListNode<T> *Next (GenListNode<T> *elem);
//返回表元素elem的直接后继元素
void Copy ( const GenList<T>& R);
//广义表的复制
int Length();            //计算广义表长度
int depth();             //计算非递归表深度
```



private:

数据成员

GenListNode<T> *first; //广义表头指针

GenListNode<T> *Copy (GenListNode<T> *ls);

//复制一个ls指示的无共享非递归表

int Length (GenListNode<T> *ls);

//求由ls指示的广义表的长度

int depth (GenListNode<T> *ls);

//计算由ls指示的非递归表的深度

bool equal (GenListNode<T> *s,

 GenListNode<T> *t);

//判以s和t为表头的两个表是否相等

void Remove (GenListNode<T> *ls);

//释放以ls为附加头结点的广义表



```
void Createlist (istream& in,
    GenListNode<T> *&ls);
//从输入流对象输入广义表的字符串描述,
//建立一个带头结点的广义表结构
friend istream& operator >> (istream& in,
    GenList<T>& L);
};
```



广义表类的构造函数

```
template <class T>
Genlist<T>::GenList( ) {           //构造函数
    GenListNode<T> * first = new GenListNode;
    if (first == NULL)
        { cerr << “存储分配失败！\n”; exit(1); }

};
```



```
template <class T>
bool GenList<T>::Head (Items<T>& x) {
//若广义表非空，则通过x返回其第一个元素的值
//否则函数没有定义
    if (first->tlink == NULL) return false;//空表
    else {                                //非空表
        x.utype = first->tlink->utype;
        x.info = first->tlink->info;
        return true;                      //x返回表头的值
    }
};
```



```
template <class T>
bool GenList<T>::Tail(GenList<T>& lt) {
    //若广义表非空，则通过lt返回广义表除表头元素
    //以外其他元素组成的表，否则函数没有定义
    if (first->tlink == NULL) return false; //空表
    else {                                //非空表
        lt.first->utype = 0;                //设置头结点
        lt.first->info.ref = 0;
        lt.first->tlink = Copy(first->tlink->tlink);
        return true;
    }
};
```



广义表的递归算法

- 一个递归算法有两种：一个是递归函数的外部调用；另一个是递归函数的内部调用。
- 通常，把外部调用设置为**公有函数**，把内部调用设置为**私有函数**。

广义表的复制算法

```
template <class T>           //公有函数
void GenList<T>::Copy(const GenList<T>& R)
    first = Copy(R.first);   //调用私有函数
};
```

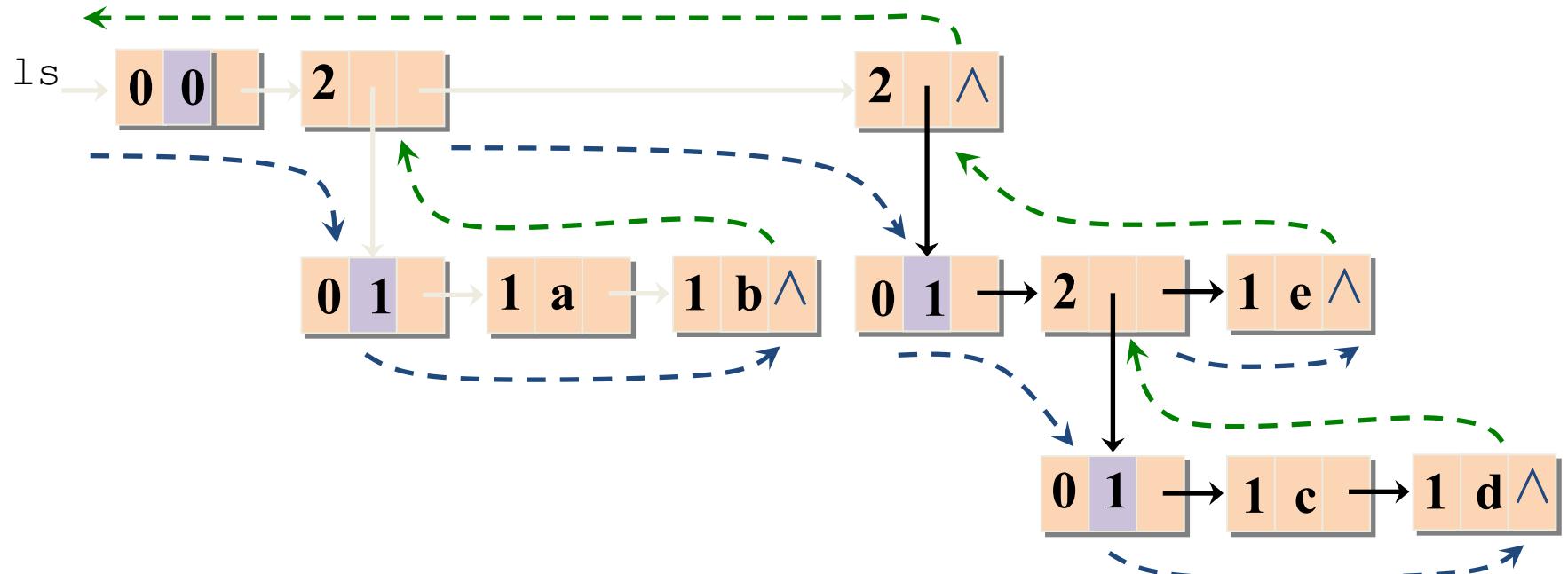


广义表的复制算法

```
template <class T>           //私有函数
GenListNode<T>* GenList<T>::Copy(GenListNode <T> *ls){
//复制一个 ls 指示的无共享子表的非递归表
    GenListNode<T> *q = NULL;
    if (ls != NULL) {
        q = new GenListNode<T>; //处理当前的结点q
        q->utype = ls->utype; //复制结点类型
        switch (ls->utype) { //根据utype传送信息
            case 0: q->info.ref = ls->info.ref; break;
            case 1: q->info.value = ls->info.value; break;
            case 2: q->info.hlink = Copy(ls->info.hlink);
                      break;
        }
        q->tlink = Copy(ls->tlink); //处理同层下一结点
    }
    return q;
};
```



广义表的复制算法





求广义表深度的算法

$$\text{Depth(LS)} = \begin{cases} 1, & \text{当LS为空表时} \\ 0, & \text{当LS为原子时} \\ 1 + \max_{0 \leq i \leq n-1} \{\text{Depth}(a_i)\}, & \text{其他, } n \geq 1 \end{cases}$$

- 例如，对于广义表

$E(B(a, b), D(B(a, b), C(u, (x, y, z))), A()))$

1 1 1 1
2
3
4



求广义表深度的算法

```
template <class T>
int GenList<T>::depth() {           //公有函数
//计算一个非递归表的深度
    return depth(first);
};
```

```
template <class T>                  //私有函数
int GenList<T>::depth(GenListNode<T> *ls) {
    if (ls->tlink == NULL) return 1;
    // ls->tlink ==NULL, 空表，深度为1
    GenListNode<T> *temp = ls->tlink;
    int m = 0, n;
```

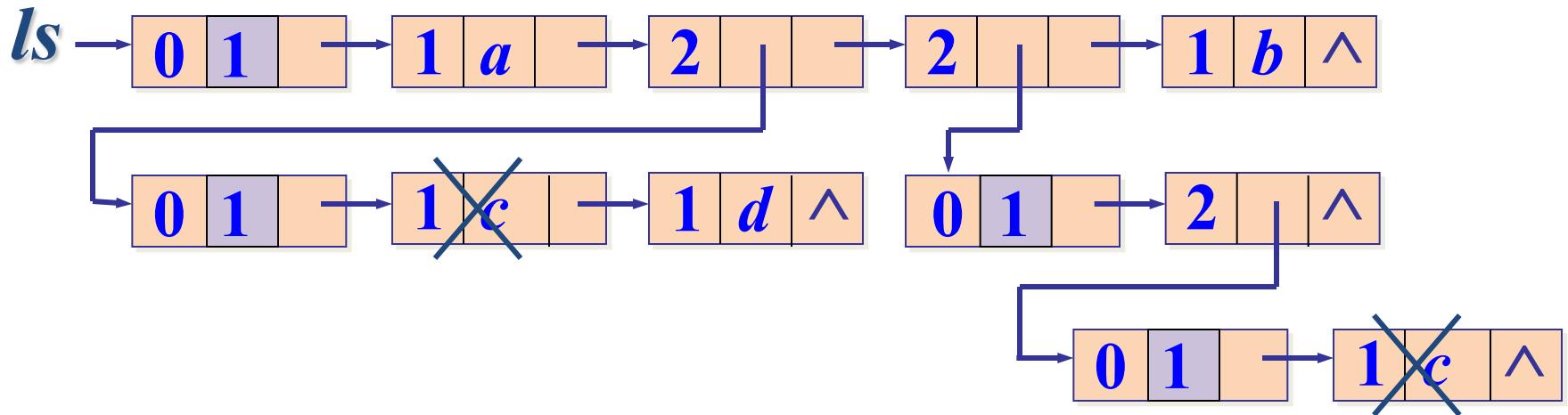


求广义表深度的算法

```
while (temp != NULL) { //在广义表顶层横扫
    if (temp->utype == 2) { //扫描到表结点
        n = depth(temp->info.hlink);
        //递归计算子表深度
        if (m < n) m = n;      //取最大深度
    }
    temp = temp->tlink;
}
return m+1;                  //返回深度
};
```



广义表的删除算法（结点删除）



删除过程：扫描子链表

- ◆ 若结点数据为 c , 删除。可能做循环连续删。
- ◆ 若结点数据不为 c , 不执行删除。
- ◆ 若结点为子表, 递归在子表执行删除。



广义表的删除算法（结点删除）

```
template <class T>
void delvalue(GenListNode<T> *ls, T x) {
    if (ls->tlink != NULL) {           //非空表
        GenListNode<T> * p = ls->tlink; //第一个结点
        while (p != NULL && (p->utype == 1 &&
            p->info.value == x)) {
            ls->tlink = p->tlink; delete p;
            p = ls->tlink;          //p指向同层下一结点
        }
        if (p != NULL) {
            if (p->utype == 2)      //递归在子表中删除
                delvalue(p->info.hlink, x);
            delvalue(p, x);          //在以p为表头的链表中递归删除
        }
    }
};
```



广义表的删除算法（结点删除）

- 对于函数中的**while**，起到的作用是同层连续出现内容为x的结点时，可以连续删除。
- 第一个**delvalue**递归调用，起到对子表转入删除的作用。
- 第二个**delvalue**递归调用，起到将当前结点看做表头，开始一个新的广义表的删除。



广义表的删除算法（表的删除）

- 对于共享表来说，如果一个表元素有多个地方使用它，贸然删去该表会造成其他地方使用出错。因此，当要做删除时，先把该表的头结点中的引用计数 `ref` 减1，当引用计数减到 0 时才能执行结点的真正释放。（缺省认为`ref`为1）

```
template <class T>
```

```
GenList<T>::~GenList() {
```

```
//广义表的析构函数, 每个头结点都有引用计  
数
```

```
    Remove(first);
```

```
};
```



广义表的删除算法（表的删除）

```
template <class T>
void GenList<T>::Remove(GenListNode<T> *ls) {
    //私有函数：释放以ls为表头指针的广义表
    ls->info.ref--;
    //头结点的引用计数减1
    if (ls->info.ref <= 0) {
        GenListNode<T> *q;
        while (ls->tlink != NULL) { //横扫表顶层
            q = ls->tlink;
            if (q->utype == 2) { //递归删除子表
                Remove(q->info.hlink);
                if (q->info.hlink->info.ref <= 0) //计数为0
                    delete q->info.hlink; //删子表头结点
            }
            ls->tlink = q->tlink; //删除当前节点
        }
    }
};
```



本章小结

- 多维数组的表示和实现
 - 多维数组（在一维数组中的存储）
 - 特殊矩阵（在一维数组中的存储）
 - 稀疏矩阵（三元组表示、正交链表表示）
- 字符串表示及操作
 - 字符串的存储表示
 - 取子串
 - 模式匹配（朴素匹配和KMP算法）
- 广义表的表示及实现
 - 表头、表尾
 - 广义表的复制、删除、求深度、建立等