



北京航空航天大学  
BEIHANG UNIVERSITY

# 数据结构课程

## 第二讲 线性表

左源

新主楼A936

zuoyuan@buaa.edu.cn



## 第二讲 线性表

### 2.1 线性表的基本概念

### 2.2 线性表的顺序存储结构

### 2.3 线性链表及其操作

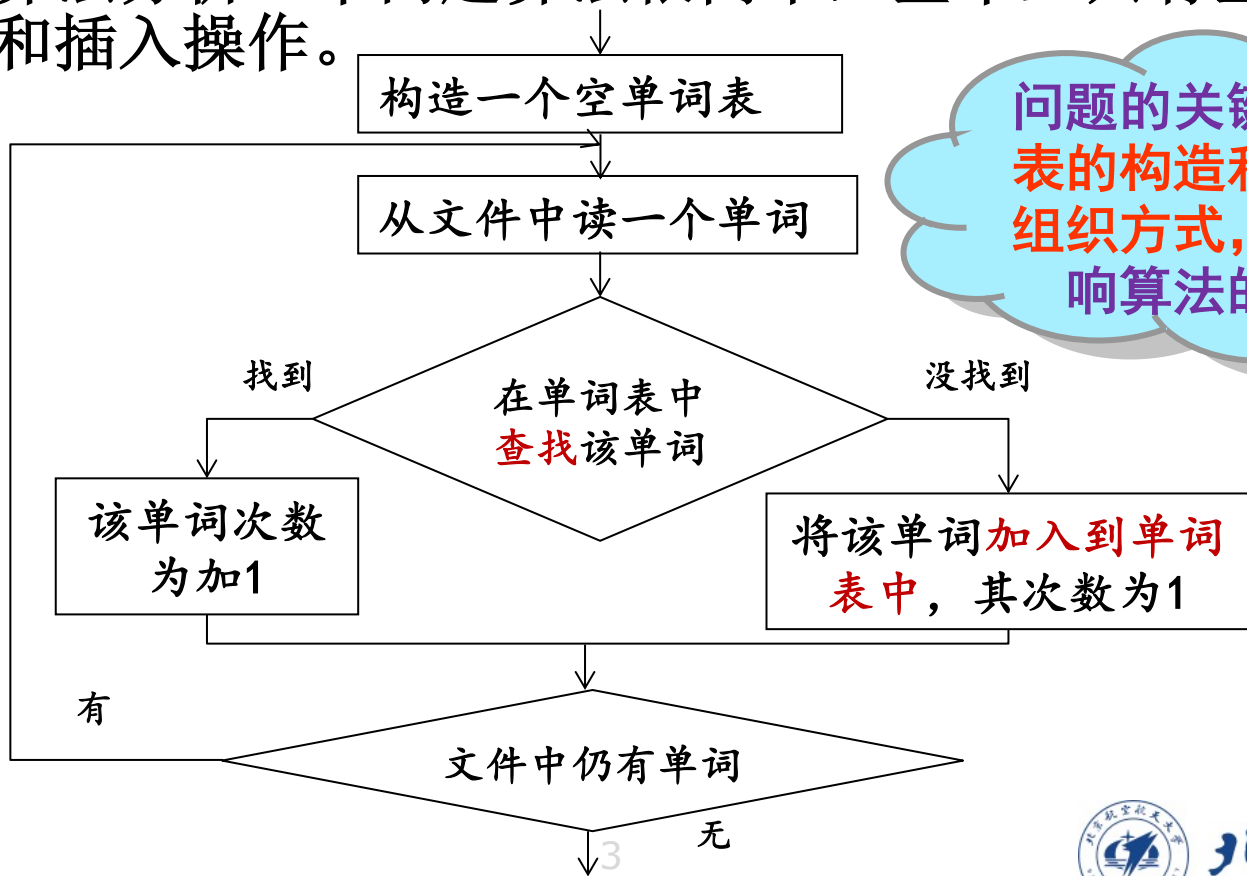
### 2.4 循环链表及其操作

### 2.5 双向链表及其操作

### 2.6 链表应用举例

## 问题 2.1 词频统计

- 问题：编写程序统计一个文本文件中每个单词的出现次数（词频统计），并按字典序输出每个单词及其频次。
- 算法分析：本问题算法很简单，基本上只有查找和插入操作。



问题的关键是单词表的构造和单词的组织方式，它将影响算法的效率

## 问题 2.1 词频统计

- 用何种数据结构来构造和组织单词表？

- 用数组？链表？还是…？来构造单词表
- 单词表是有序还是无序？



- 不同数据结构构造的单词表如何影响着算法的性能？



当单词表较大时（如一本长篇小说），单词的查找和插入会面临什么问题？

## 2.1 线性表的基本概念

### 2.1.1 线性表的定义

$A = (a_1, a_2, a_3, \dots, a_n)$

#### 1. 线性关系

(1) 当  $1 < i < n$  时， $a_i$  的直接前驱为  $a_{i-1}$ ， $a_i$  的直接后继为  $a_{i+1}$ ；

(2) 除了第一个元素与最后一个元素，序列中任何一个元素有且仅有一个直接前驱元素，有且仅有一个直接后继元素；

(3) 数据元素之间直接先后顺序为“一对一”的关系。

|          | 学号    | 姓名  | 性别  | 年龄  | 其他    |
|----------|-------|-----|-----|-----|-------|
| $a_1$    | 99001 | 张华  | 女   | 17  | ..... |
| $a_2$    | 99002 | 李军  | 男   | 18  | ..... |
| $a_3$    | 99003 | 王明  | 男   | 17  | ..... |
| ⋮        | ...   | ... | ... | ... | ..... |
| ⋮        | ...   | ... | ... | ... | ..... |
| ⋮        | ...   | ... | ... | ... | ..... |
| $a_{50}$ | 99050 | 刘东  | 女   | 19  | ..... |

## 2.1 线性表的基本概念

### 2. 线性表的定义

数据元素之间具有的逻辑关系为**线性关系**的数据元素集合称为**线性表**，数据元素的个数  $n$  为线性表的长度，长度为 0 的线性表称为空表。

线性表的特点：

(1) 同一性 (2) 有穷性 (3) 有序性

# 几个线性表的例子

$$A = (a_1, a_2, a_3, \dots, a_n)$$

一个数据元素  
为一个整数

① 数列: ( 25, 12, 78, 34, 100, 88 )  
           $a_1$     $a_2$     $a_3$     $a_4$     $a_5$     $a_6$

② 字母表: ( 'A', 'B', 'C', ..... , 'Z' )  
               $a_1$     $a_2$     $a_3$    ... ..  $a_{26}$

一个数据元素  
为一个字母



# 几个线性表的例子

## 3 数据表（文件）：

|          | 学号    | 姓名 | 性别 | 年龄 | 其他    |
|----------|-------|----|----|----|-------|
| $a_1$    | 99001 | 张华 | 女  | 17 | ..... |
| $a_2$    | 99002 | 李军 | 男  | 18 | ..... |
| $a_3$    | 99003 | 王明 | 男  | 17 | ..... |
|          |       |    |    |    |       |
| ⋮        |       |    |    |    |       |
|          |       |    |    |    |       |
| $a_{50}$ | 99050 | 刘东 | 女  | 19 | ..... |

一个数据元素为一个记录





## 2.1 线性表的基本概念

### 2.1.2 线性表的基本操作

1. **创建**一个新的线性表。
2. 求线性表的长度。
3. **检索**线性表中第  $i$  个数据元素。 ( $1 \leq i \leq n$ )
4. **根据**数据元素的某数据项(通常称为关键字)的值求该数据元素在线性表中的位置 (查找)。
5. 在线性表的第  $i$  个位置上**存入**一个新的数据元素。
6. 在线性表的第  $i$  个位置上**插入**一个新的数据元素。
7. **删除**线性表中第  $i$  个数据元素。
8. 对线性表中的数据元素按照某一个数据项的值的大小做升序或者降序**排序**。

## 2.1 线性表的基本概念

### 2.1.2 线性表的基本操作

9. **销毁**一个线性表。
10. **复制**一个线性表。
11. **按照一定的原则，将两个或两个以上的线性表合并**成为一个线性表。
12. **按照一定的原则，将一个线性表分解**为两个或两个以上的线性表。

.....

## 2.1 线性表的基本概念

### 2.1.2 线性表的基本操作

- `initList(nodeType *list, int length);` //创建一个空表
- `destroyList (nodeType *list, int length);` //销毁一个表
- `printList(nodeType *list, int length);` //输出一个表
- `getNode (nodeType *list, int pos);` //获取表中指定位置元素
- `searchNode(nodeType *list, Type node);` //在表中查找某一元素
- `insertNode(nodeType *list, int pos, nodeType node);` // 在表中指定位置插入一个结点
- `deleteNode(nodeType *list, int pos);` //在表中指定位置删除一个结点



## 第二讲 线性表

2.1 线性表的基本概念

2.2 线性表的顺序存储结构

2.3 线性链表及其操作

2.4 循环链表及其操作

2.5 双向链表及其操作

2.6 链表应用举例

## 2.2 线性表的顺序存储结构

### 2.2.1 构造原理

用一组地址连续的存储单元依次存储线性表的数据元素，数据元素之间的逻辑关系通过数据元素的存储位置直接反映。

$( a_1, a_2, a_3, \dots \dots, a_n )$

k个单元



$LOC(a_i)$

所谓一个元素的地址是指该元素占用的k个(连续的)存储单元的第一个单元的地址。

## 2.2 线性表的顺序存储结构

### 结论

若假设每个数据元素占用  $k$  个存储单元，并且已知第一个元素的存储位置  $LOC(a_1)$ ，则有

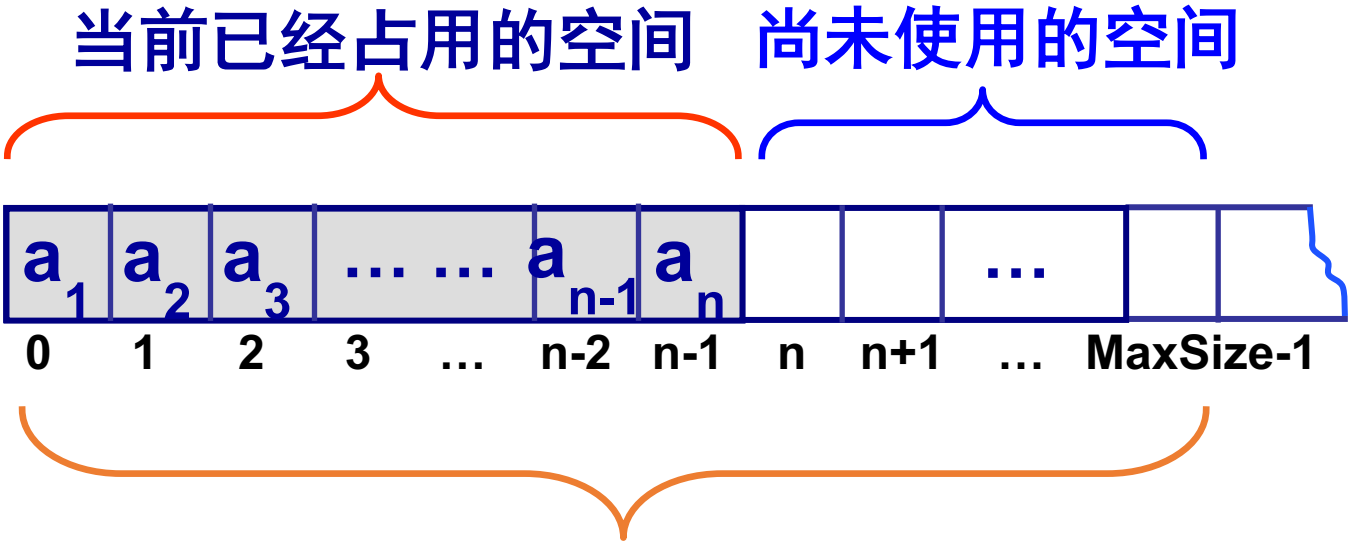
$$LOC(a_i) = LOC(a_1) + (i-1) \times k$$

### 思考

这个如此简单的公式，说明了顺序存储的线性表具有一个什么样的巨大优势呢？

# 顺序存储结构示意图

$(a_1, a_2, a_3, \dots, a_{n-1}, a_n)$



**$n < \text{MaxSize}$**

事先分配给线性表的空间



$(a_1, a_2, a_3, \dots, a_{n-1}, a_n)$

在C语言中

预先分配给线性表的空间大小

```
#define MaxSize 100  
ElemType A[MaxSize];  
int n;
```

表的长度

数组-顺序表



# 2.2 线性表的顺序存储结构

## 2.2.2 基本算法

1. 查找：确定元素item在长度为n的顺序表list中的位置

( a<sub>1</sub>, a<sub>2</sub>, a<sub>3</sub>, ..., a<sub>i-1</sub>, a<sub>i</sub>, a<sub>i+1</sub>, ..., a<sub>n-1</sub>, a<sub>n</sub> )

算法

顺序查找法

```
int searchElem(ElemType list[ ], int n, ElemType item)
{
    int i;
    for(i=0;i<n;i++)
        if(list[i]==item)
            return i; /* 查找成功, 返回在表中位置 */
    return -1; /* 查找失败, 返回信息-1 */
}
```

时间复杂度O(n)



## 2.2 线性表的顺序存储结构

### 2.2.2 基本算法

#### 如何在有序顺序表中查找元素？

##### ■ 折半查找算法（The binary search）：

在有序数据集中查找指定数据项（或数据项插入位置）最常用及最快的算法是折半查找算法。当数据集较大时，折半查找平均效率高。

# 折半查找算法 (binary search)

- 假设数据集按由小到大排列，折半查找算法的核心思想是：
  1. 将要查找的有序数据集的中间元素与指定数据项相比较；
  2. 如果指定数据项小于该中间元素，则将数据集的前半部分指定为要查找的数据集，然后转步骤1；
  3. 如果指定数据项大于该中间元素，则将数据集的后半部分指定为要查找的数据集，然后转步骤1；
  4. 如果指定数据项等于中间元素，则查找成功结束。
  5. 最后如果数据集中没有元素再可进行查找，则查找失败。

下面以在一个有序整型数据集中查找给定整数为例来说明折半查找原理。

# 折半查找算法 (续)

例：在下面有序数据集中查找数据项62

|   |   |    |    |    |    |    |    |    |    |
|---|---|----|----|----|----|----|----|----|----|
| 0 | 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| 5 | 7 | 16 | 24 | 25 | 50 | 45 | 50 | 62 | 65 |

← 查找范围 →  
low      **item > data[mid], 即 62 > 25**      high

|   |   |    |    |    |    |    |    |    |    |
|---|---|----|----|----|----|----|----|----|----|
| 0 | 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| 5 | 7 | 16 | 24 | 25 | 50 | 45 | 50 | 62 | 65 |

← 查找范围 →  
low      **item > data[mid], 即 62 > 50**      high

|   |   |    |    |    |    |    |    |    |    |
|---|---|----|----|----|----|----|----|----|----|
| 0 | 1 | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
| 5 | 7 | 16 | 24 | 25 | 50 | 45 | 50 | 62 | 65 |

← 查找范围 →  
**item = data[mid], 即 62 = 62**      low high

item = 62 (查找项)  
low = 0 (查找范围开始)  
high = 9 (查找范围结束)  
mid = (low+high)/2=4 (查找范围中间)

low = mid+1=5  
high = 9  
mid = (low+high)/2=7

low = mid+1=8  
high = 9  
mid = (low+high)/2=8

# 折半查找算法（续）

算法

在有序（递增）顺序表list中查找给定元素如下：

```
int searchElem(ElemType list[ ], int n, ElemType item)
{
    int low=0, high=n-1, mid;
    while(low <= high){
        mid = (high + low) / 2;
        if( item < list[mid])
            high = mid - 1;
        else if ( item > list[mid])
            low = mid + 1;
        else
            return (mid);
    }
    return -1;
}
```

有关查找算法将在后面章节作专门介绍

时间复杂度 $O(\log_2 n)$

也就是说假如一个顺序表有1024个元素，顺序查找算法平均查找次数为512次，而折半查找算法最多只须10次。

## 2. 插入：在长度为n的顺序表list的第i个位置上插入一个新的数据元素item

在线性表的第  $i-1$  个数据元素与第  $i$  个数据元素之间插入一个由符号  $item$  表示的数据元素，使得长度为  $n$  的线性表

$( a_1, a_2, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{n-1}, a_n )$

$n$  个数据元素

转换成长度为  $n+1$  的线性表

$( a_1, a_2, \dots, a_{i-1}, item, a_i, \dots, a_{n-1}, a_n )$

$n+1$  个数据元素



$( a_1, a_2, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{n-1}, a_n )$

$n-i+1$ 个元素

依次后移一个位置

item

$A[j+1]=A[j];$

正常情况下需要做的工作：

- (1) 将第  $i$  个元素至第  $n$  个元素依次后移一个位置；
- (2) 将被插入元素插入表的第  $i$  个位置；
- (3) 修改表的长度(表长增1)。(  $n++;$  )

插入操作需要考虑的异常情况：

- (1) 是否表满？  $n=MaxSize$  ?
- (2) 插入位置是否合适？(正常位置: $0 \leq i \leq n$ )



# 约定

若插入成功，算法返回1，  
否则，算法返回-1。





# 算法

/\* N是顺序表的长度，为一个全局变量\*/

```
int insertElem(ElemType list[], int i, ElemType item )
```

```
{
```

```
int k;
```

测试空间满否

测试插入位置合适否

```
if (N==MaxSize || i<0 || i>N)
```

```
return -1;
```

/\* 插入失败 \*/

```
for( k=N-1; k>=i; k-- )
```

```
list[k+1]=list[k];
```

/\* 元素依次后移一个位置 \*/

```
list[i]=item;
```

/\* 将item插入表的第i个位置 \*/

```
N++;
```

/\* 线性表的长度加1 \*/

```
return 1;
```

/\* 插入成功 \*/

```
}
```

该算法的时间复杂度是： **$O(n)$**

衡量插入和删除算法时间效率的一个**重要指标**：  
**元素移动次数的平均值**

若设  $p_i$  为插入一个元素于线性表第  $i$  个位置的**概率(概率相等)**，则在长度为  $n$  的线性表中插入一个元素需要移动其他的元素的平均次数为

$$T_{is} = \sum_{i=1}^n p_i(n-i+1) = \sum_{i=1}^n (n-i+1)/(n+1) = n/2$$

### 3. 删除：删除长度为n的顺序表list的某个数据元素

把线性表的第  $i$  个数据元素从线性表中去掉，使得长度为  $n$  的线性表

$( a_1, a_2, \dots, a_{i-1}, a_i, a_{i+1}, \dots, a_{n-1}, a_n )$

$n$  个数据元素

转换成长度为  $n-1$  的线性表

$( a_1, a_2, \dots, a_{i-1}, a_{i+1}, \dots, a_{n-1}, a_n )$

$n-1$  个数据元素



(  $a_1, a_2, \dots, a_{i-1}$ ,  $a_i$ ,  $a_{i+1}, \dots, a_{n-1}, a_n$  )

$n-i$ 个元素

依次前移一个位置

`list[j-1]=list[j];`

正常情况下需要做的工作:

- (1) 将删除元素的下一元素至第 $n$ 个元素依次前移一个位置;
- (2) 修改表的长度(表长减1)。 (`n--;`)

删除操作需要考虑的异常情况:

(1) 是否表空?

`n=0` ?

(2) 删除位置是否合适? ( 正常位置:  $0 \leq i \leq n-1$  )



# 约定

若删除成功，算法返回1，  
否则，算法返回-1。



# 算法

```
/* N是表的长度，为一个全局变量 */
int deleteElem( ElemType list[ ], int i )
{
    int k;
    if( i<0 || i>N-1 )
        return -1;
    for( k=i+1; k<N; k++ )
        list[k-1]=list[k];
    N--;
    return 1;
}
```

测试表空和位置合适与否


/\* 删除失败 \*/

/\* 元素依次前移一个位置 \*/

/\* 线性表的长度减1 \*/

/\* 删除成功 \*/

该算法的时间复杂度是： **$O(n)$**



衡量插入和删除算法时间效率  
的一个重要指标：  
元素移动次数的平均值

若  $p_i$  为删除线性表中第  $i$  个数据元素的概率  
(设概率相等), 在长度为  $n$  的线性表中删除第  $i$   
个数据元素需要移动其他的元素的平均次数为

$$T_{ds} = \sum_{i=1}^n p_i(n-i) = \sum_{i=1}^n (n-i)/n = (n-1)/2$$

## 2.2.3 顺序存储结构的特点

### 1. 优点

$$LOC(a_i) = LOC(a_1) + (i-1) \times k$$

1. 构造原理简单、直观，易理解。
2. 元素的存储地址可以通过一个简单的解析式计算出来。是一种**顺序存储结构**，读取速度快。
3. 由于只需存放数据元素本身的信息，而无其他空间开销，相对链式存储结构而言，存储空间开销小
4. 对于有序表，可使用折半查找等快速查找算法，**查找效率高。**

### 2. 缺点

对于**动态表**（即需要频繁插入和删除操作的表）往往由于问题规模不知，需要事先分配很大的空间，造成空间浪费或空间不足。

1. **存储分配需要事先进行。**
2. 需要一块地址连续的存储空间。
3. 基本操作(如插入、删除)的时间效率较低。

需要频繁的移动数据

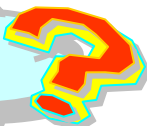
$O(n)$



例

已知长度为  $n$  的非空线性表  $list$  采用顺序存储结构,并且数据元素按值的大小非递减排列(有序),写一算法,在该线性表中插入一个数据元素  $item$ ,使得线性表仍然保持按值非递减排列。

如何找到插入位置



$$a_i \leq a_{i+1} \quad 1 \leq i \leq n-1$$

$item$   
 $a_1, a_2, a_3, \dots, \underline{a_j}, \underline{a_{j+1}}, \underline{a_{j+2}}, \dots, a_n$   
↑ 依次后移一个位置

插入

```
for(j=n-1;j>=i;j--)  
    list[j+1]=list[j];  
list[j]=item;  
n++;
```

## 需要做的工作

1. 寻找插入位置：  
从表的第一个元素开始进行比较，若有关系  
 $item < a_i$   
则找到插入位置为表的第*i*个位置。
2. 将第*i*个元素至第*n*个元素依次后移一个位置；
3. 将item插入表的第*i*个位置；
4. 表的长度增1。

例

1, 3, 5, 5, 8, 10, 13, 15, 20, 25

item=12

1, 3, 5, 5, 8, 10, 12, 13, 15, 20, 25

# 算法

$a_1, a_2, a_3, \dots, a_i, a_{i+1}, a_{i+2}, \dots, a_n, \text{item}$

↑ 特殊情况

/\* N是表长度，是一个全局变量 \*/

```
int insertElem(ElemType list[], ElemType item
```

能处理吗？

```
{  
    int i,j;
```

```
    if(N == MAXSIZE) return -1;
```

```
    for(i=0; i<N&&item>=list[i]; i++)
```

/\* 寻找item的合适位置 \*/

```
    ;
```

```
    for(j=N-1; j>=i; j--)
```

```
        list[j+1]=list[j];
```

确定插入位置

```
    list[i]=item;
```

```
    N++;
```

/\* 将item插入表中 \*/

```
    return 1;
```

表长加1

$O(n)$



# 主函数

```
#include <stdio.h>
#define MAXSIZE 1000
typedef int ElemType;
int N=0;
int main()
{
    int i;
    ElemType data,list[MAXSIZE];
    scanf("%d", &N);
    for(i=0; i< N; i++)
        scanf("%d", &list[i]);
    scanf("%d", &data);
    if(insertElem(list, data ) == 1)
        printf("OK\n");
    else
        printf("Fail\n");
    return 0;
}
```



# 算法

```
int insertElem(ElemType list[ ], ElemType item)
{
    int i=0,j;

    if (N == MAXSIZE) return -1;
    i = searchElem(list, item);

    for(j=N-1; j>=i; j--)
        list[j+1]=list[j];

    list[i]=item;
    N++;
    return 1;
}
```

算法总复杂度为：

$O(\log_2 n + n)$

$= O(n)$

折半查找算法如下：

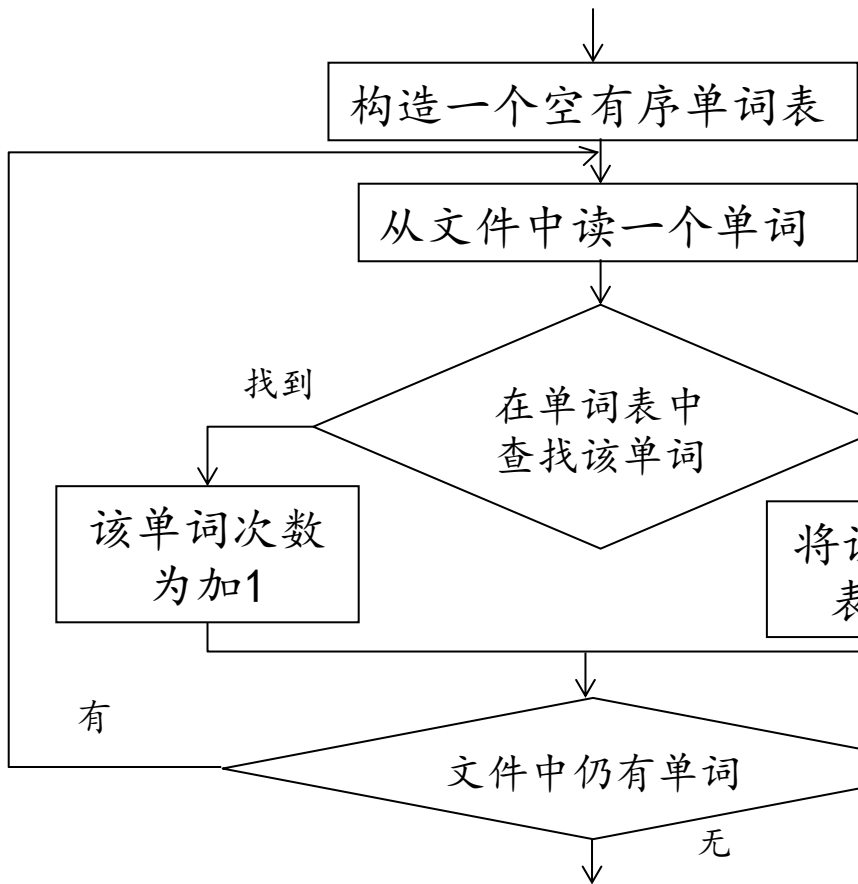
```
int searchElem(ElemType list[ ], ElemType item)
{
    int low=0, high=n-1, mid;
    while(low <= high){
        mid = (high + low) / 2;
        if(( item < list[mid])
            high = mid - 1;
        else if ( item > list[mid])
            low = mid + 1;
        else
            return (mid);
    }
    return low ;
}
```

## 问题2.1：词频统计 – 顺序表

- 问题：编写程序统计一个文件中每个单词的出现次数（词频统计），并按字典序输出每个单词及出现次数。
- 算法分析：
  1. 首先构造一个空的有序（字典序）单词表；
  2. 每次从文件中读入一个单词；
  3. 在单词表中（折半）查找该单词，若找到，则单词次数加1，否则将该单词插入到单词表中相应位置，并设置出现次数为1；
  4. 重复步骤2，直到文件结束。

# 问题2.1: 词频统计 - 顺序

```
/*用数组构造一个顺序表，  
表中单词按字典序组织*/  
struct Inode {  
    char word[MAXWORD];  
    int count;  
};  
struct Inode wordlist[MAXSIZE];  
int N=0; //一个空表
```



用函数getWord从文件中每次读入一个单词

用函数searchWord在单词表中查找一个单词

用函数insertWord在单词表中插入一个单词



## 问题2.1：词频统计 – 代码实现

```
#include <stdio.h>
#include <string.h>
#include <ctype.h>
#define MAXWORD 32
#define MAXSIZE 1024
struct Inode {
    char word[MAXWORD];
    int count;
};
int getWord(FILE *bfp, char *w);
int searchWord(struct Inode list[], char *w);
int insertWord(struct Inode list[], int pos, char *w);
int N=0; //单词表中单词的实际个数
```

为什么要将N  
定义为一个全局变量！

```
int main()
{
    struct Inode wordlist[MAXSIZE];           /*单词表*/
    int i;
    char filename[MAXWORD], word[MAXWORD];
    FILE *bfp;

    scanf("%s", filename);
    if((bfp = fopen(bname, "r")) == NULL){
        fprintf(stderr, "%s can't open!\n", filename);
        return -1;
    }
    while( getWord(bfp, word) != EOF)
        if(searchWord(wordlist, word) == -1) {
            fprintf(stderr, "Wordlist is full!\n");
            return -1;
        }
    for(i=0; i<= N-1; i++)
        printf("%s %d\n", wordlist[i].word, wordlist[i].count);
    return 0;
}
```



## 问题2.1: 词频统计 – 代码实现\*

```
/*从文件中读入一个单词（仅由字母组成的串），并转换成小写字母*/
int getWord(FILE *fp, char *w)
{
    int c;

    while(!isalpha(c=fgetc(fp)))
        if(c == EOF) return c;
        else continue;
    do {
        *w++ = tolower(c);
    } while(isalpha(c=fgetc(fp)));
    *w='\0';
    return 1;
}
```

## 问题2.1: 词频统计 - 代码实现\*

/\*在表中相应位置插入一个单词，同时置次数为1\*/

```
int insertWord(struct lnode list[], int pos, char *w)
```

```
{
    int i;

    if (N == MAXSIZE) return -1;

    for(i=N-1; i>=pos; i--){
        strcpy(list[i+1].word, list[i].word);
        list[i+1].count = list[i].count;
    }
    strcpy(list[pos].word, w);
    list[pos].count = 1;
    N++;
    return 1;
}
```

/\*在表中查找一单词，若找到，则次数加1；  
否则将该单词插入到有序表中相应位置，同  
时次数置1\*/

```
int searchWord(struct lnode list[], char *w)
```

```
{
    int low=0, high=N-1, mid=0;
    while(low <= high){
        mid = (high + low) / 2;
        if(strcmp( w, list[mid].word)<0)
            high = mid -1;
        else if(strcmp( w, list[mid].word)>0)
            low = mid + 1;
        else {
            list[mid].count++;
            return 0;
        }
    }
}
```

```
return insertWord(list, low, w);
```

在本程序中：

查找算法复杂度为 $O(\log_2 n)$

插入算法复杂度为 $O(n)$

## 问题2.1：词频统计

# 思考1

- 由于顺序表（数组）的大小需要事先确定，用顺序表作为单词表会有什么问题？

由于事先不知道被统计的文件大小（可能是本很厚的书），单词表的大小如何定：

- 1) 太小，对于大文件会造成单词表溢出；
- 2) 太大，对一般文件处理会造成很大的空间浪费。

- 词频统计需要在单词表中频繁的查找和插入单词，有序顺序表结构的单词表有什么特点？

- 1) 单词查找效率很高（可用折半查算法，一次查找算法复杂度为 $O(\log_2 n)$ ）；
- 2) 单词表中单词需要频繁移动（对于一般的英文材料来说，插入操作较多，一次插入算法的复杂度为 $O(n)$ ）；

- 无序（输入序）顺序表结构构造的单词表又有什么特点？

- 1) 单词查找效率低（顺序查找算法，一次查找算法复杂度为 $O(n)$ ）；
- 2) 单词不需要移动（新单词总是放在表尾），但需要对最终的总表要进行排序，一般排序算法的复杂度为 $O(N^2)$ ；

# 思考2

## 问题2.1：词频统计

- 请用**无序（输入序）**顺序表来重新实现问题2.1，并比较一下两个程序的实际性能。  
即单词表中的单词按其读入顺序排放，新加入的单词总是放在表的末尾。

如何获取程序的实际运行时间？

# 作业

已知顺序表中的每个数据元素为互不相等的整数。设计算法将其调整为两部分，使得左边所有元素为奇数，右边为偶数。分析算法的时间复杂度。

**要求：**

- (1) 数据结构需采用顺序存储结构的线性表。设计并给出算法流程。
- (2) 分析算法的大O时间复杂度
- (3) 用C语言实现

**注：作业待线性表全部讲完时，和线性表部分的其他作业一并提交。提交方式后续会给出说明！**

# 线性表顺序存储结构的优缺点

## ■ 优点

- 构造原理简单（如用数组）
- 无须为表中元素之间的逻辑关系而增加额外的存储空间
- 可以快速地存取表中任一位置的元素。
- 对于有序表可以用折半查找算法查找元素，查找效率高。

## ■ 缺点

- 存储空间需要事先分配
- 需要一片地址连续的存储空间（容易造成表空间不够或浪费）
- 插入和删除操作需要移动大量元素，效率低。



## 第二讲 线性表

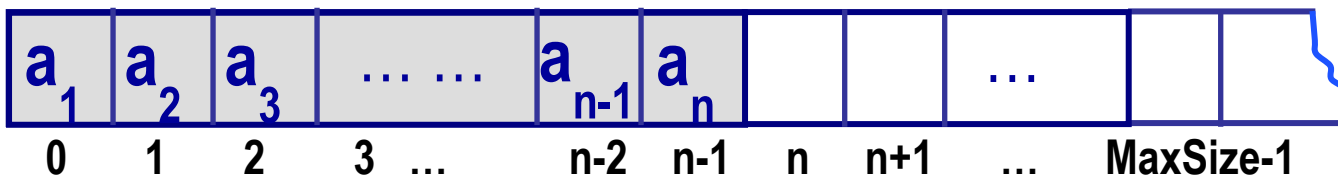
- 2.1 线性表的基本概念
- 2.2 线性表的顺序存储结构
- 2.3 线性链表及其操作
- 2.4 循环链表及其操作
- 2.5 双向链表及其操作
- 2.6 链表应用举例

## 2.3 线性表的链式存储结构

### 将要讨论的内容

1. 线性链表的构造原理
2. 几个常用符号的说明
3. 线性链表的有关(操作)算法

### 顺序表



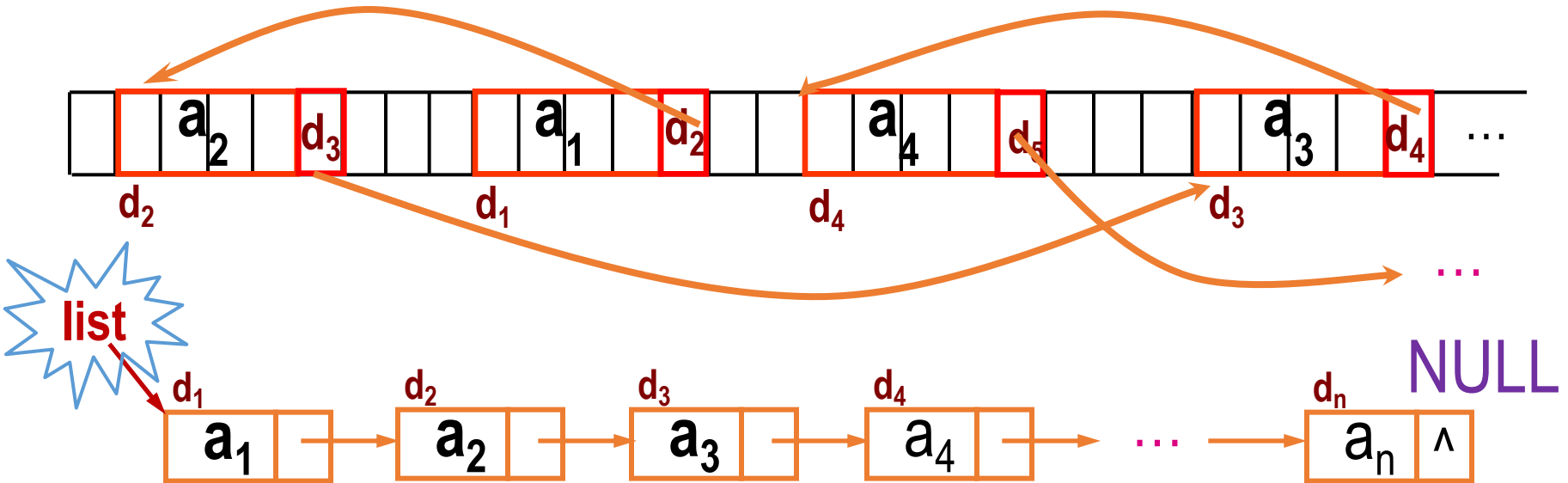
- (1) 静态分配内存
- (2) 需要连续空间
- (3) 插入和删除操作时效较低



# 线性链表的构造原理

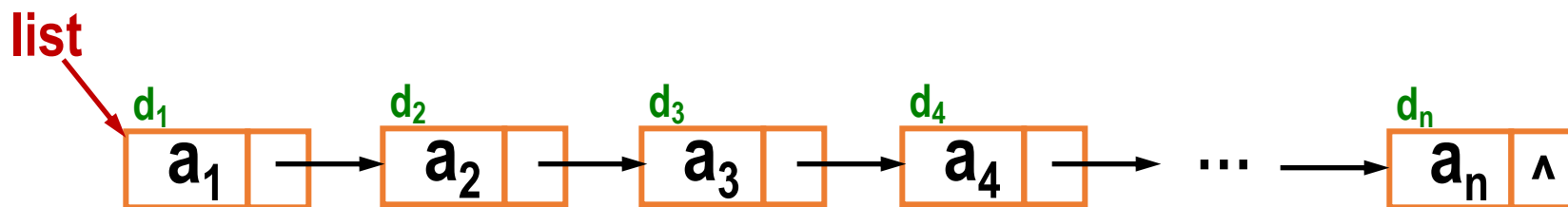
用一组地址任意的存储单元(连续的或不连续的)依次存储表中各个数据元素, 数据元素之间的逻辑关系通过**指针**间接地反映出来。

$(a_1, a_2, a_3, a_4, \dots, a_{n-1}, a_n)$





线性表的这种存储结构称为 **线性链表**，或者 **单链表**，其一般形式为：



k个存储单元

一个链结点

数据域

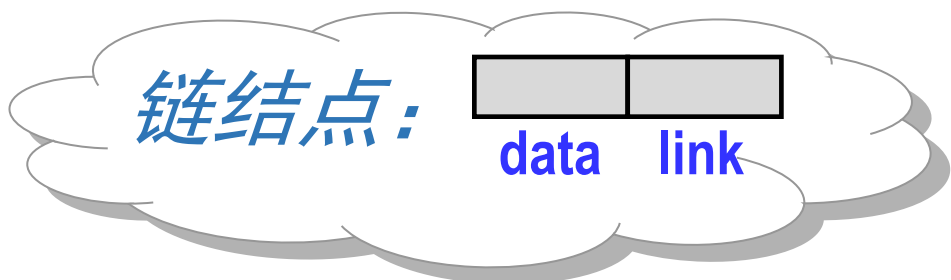
指针域

data

link



# 线性链表的定义



链表定义

```
struct node{
    ElemType data;
    struct node *link;
};
struct node *list, *p;
```

类型定义

```
struct node{
    ElemType data;
    struct node *link;
};
typedef struct node *Nodeptr;
typedef struct node Node;
Nodeptr list, p;
```

|                 | Num   | Name | Age |
|-----------------|-------|------|-----|
| a <sub>1</sub>  | 60101 | 张三   | 17  |
| a <sub>2</sub>  | 60102 | 李四   | 16  |
| a <sub>3</sub>  | 60103 | 王五   | 20  |
|                 |       |      |     |
| a <sub>30</sub> |       |      |     |



```
typedef struct {
    int Num;
    char Name[10];
    int Age;
} ElemType;
```

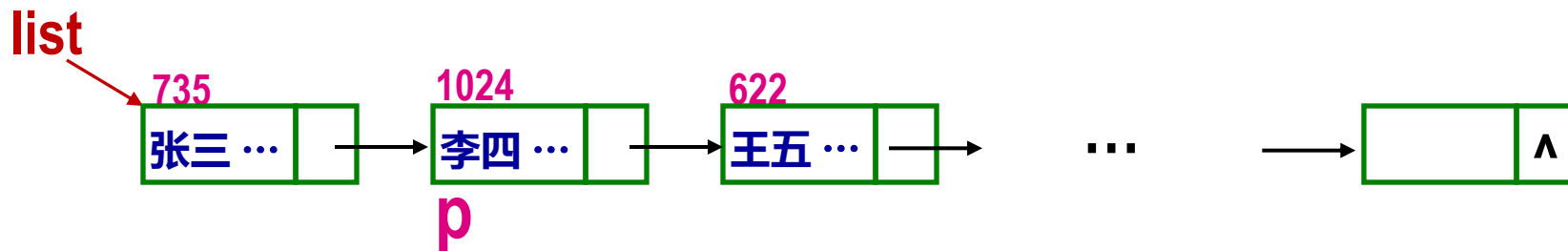


## 链表结点的基本操作

若指针变量  $p$  为指向链表中某结点的指针(即  $p$  的内容为链表中某链结点的地址), 则

$p \rightarrow data$

表示由  $p$  指向的链结点的数据域



$p \rightarrow link$

表示由  $p$  指向的链结点的指针域, 即  $p$  所指的链结点的下一个链结点的指针(地址)

## 2.3.3 链表的基本操作

- **建立**一个线性链表。
- 求线性链表的长度。
- 在非空线性链表的第一个结点前**插入**一个数据信息为 `item` 的新结点。
- 在线性链表中由指针 `q` 指出的结点之后**插入**一个数据信息为 `item` 的链结点。
- 在线性链表中满足某条件的结点后面插入一个数据信息为 `item` 的链结点。
- 从非空线性链表中**删除**链结点 `q` (`q` 为指向被删除链结点的指针)。

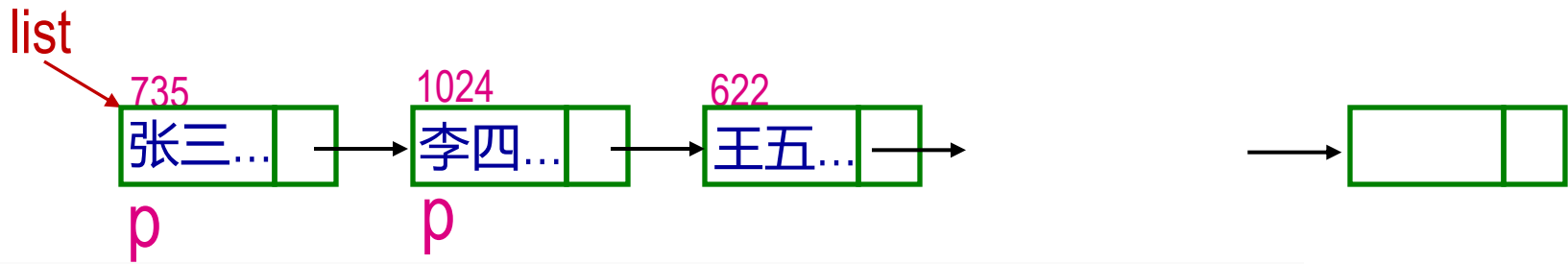
## 2.3.3 链表的基本操作

- **删除线性链表中满足某个条件的链结点。**
- **线性链表的逆转。**
- **将两个线性链表合并为一个线性链表。**
- **检索线性链表中的第 $i$ 个链结点。**

.....

## 2.3.3 链表的基本操作

- `createList(int n);` //创建一个具有n个结点的链表
- `getLength(Nodeptr list);` //获得链表的长度
- `destroyList (Nodeprt list);` //销毁一个表
- `printList(Nodeptr list);` //输出一个表
- `insertFrst (Nodeptr list, ElemType elem );` //在链表头插入一个元素
- `insertLast(Nodeptr list , ElemType elem );` //在链表尾插入一个元素
- `insertNode(Nodeptr list , Nodeptr p, ElemType elem );` //在链表某一结点后插入包含某一个元素的结点
- `searchNode(Nodeptr list , ElemType elem );` //在链表中查找某一元素
- `deleteNode(Nodeptr list , ElemType elem );` //在链表中删除包含某一元素结点



指向下一个结点:

```
p = p->link;
```

插入一个结点:

```
q->link = p->link;
```

```
p->link = q;
```

(在p后插入q)

删除一个结点:

```
q = p->link;
```

```
p->link = p->link->link;
```

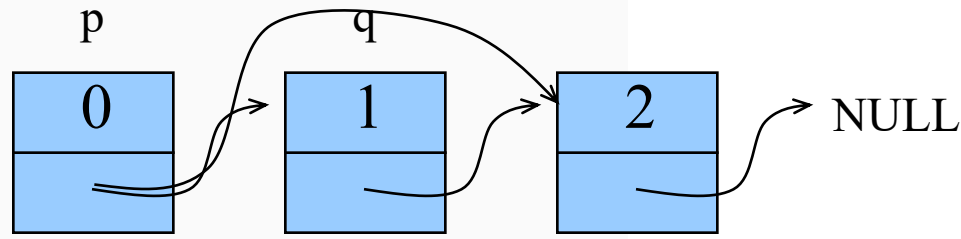
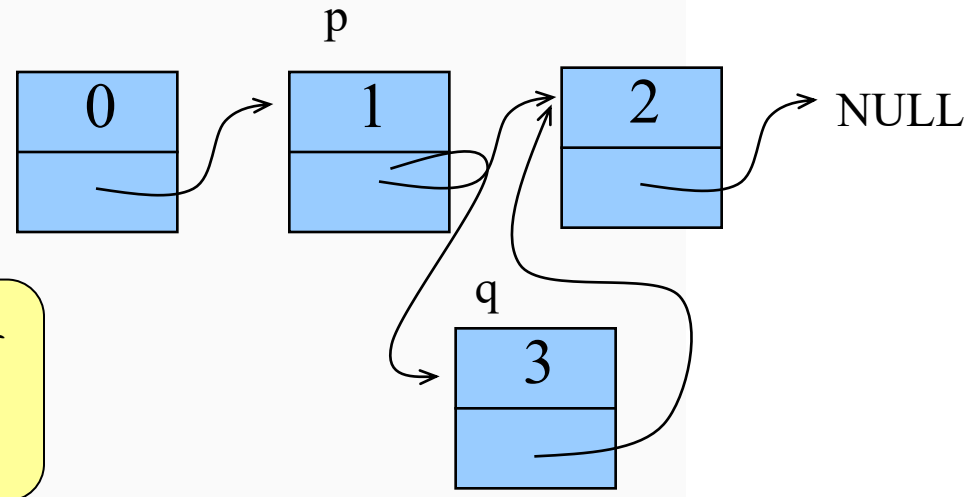
free(q); (删除p的下一结点)

遍历一个链表:

```
for(p=list; p != NULL; p=p->link)
```

....

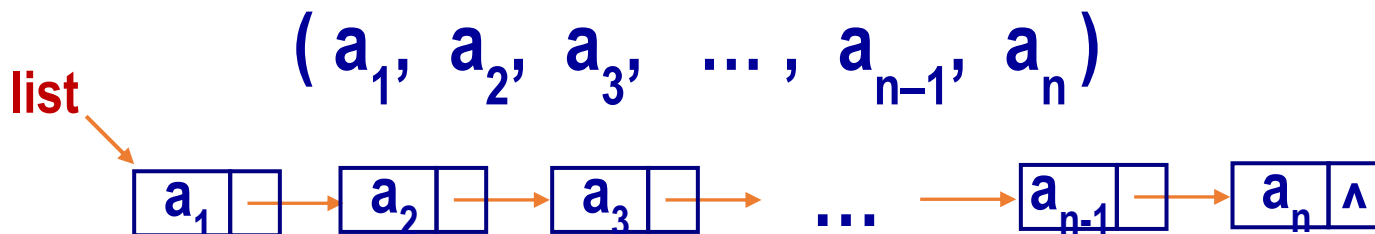
为什么首先要  
执行:  
q = p->link?







# 1. 建立一个线性链表 (C语言)



```

struct node{
    ElemType  data;
    struct node  *link;
};
  
```

```

typedef struct node *Nodeptr;
typedef struct node Node;
Nodeptr  list, p;
  
```

申请一个链结点的空间

**`p=(Nodeptr)malloc(sizeof(Node));`**

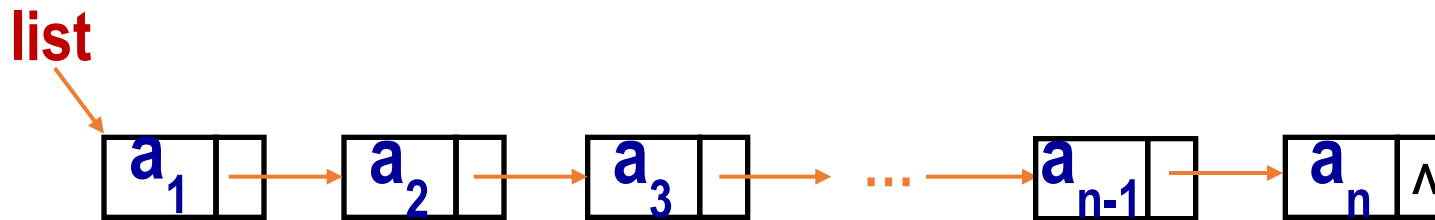
释放一个链结点的空间

**`free(p);`**

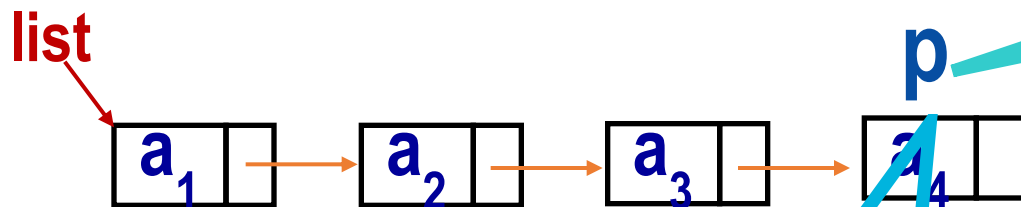
头文件: `#include <stdlib.h>`



$( a_1, a_2, a_3, a_4, \dots, a_{n-1}, a_n )$



一个结点的插入过程



# 算法

```
Nodeptr createList( int n ) /*创建一个具有n个结点的链表 */
{
    /* list是链表头指针, q指向新申请的结点, p指向最后一个结点*/
    Nodeptr p, q, list=NULL;
    int i;
    for(i=0;i<n;i++){
        q=(Nodeptr)malloc(sizeof(Node));
        q->data=read();          /* 取一个数据元素 */
        q->link=NULL;

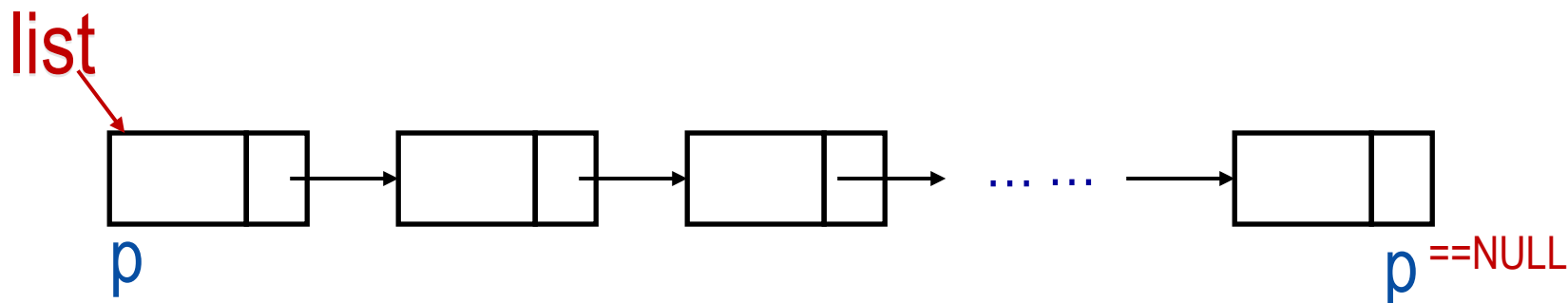
        if (list==NULL) /*链表为空*/
            list=p=q;
        else
            p->link=q;    /* 将新结点链接在链表尾部 */

        p=q;
    }
    return list;
}
```

申请一个新的链结点

时间复杂度 $O(n)$

## 2. 求线性链表的长度



初始：

```
n=0;  
p=list;
```

```
p=p->link;  
n++;
```

链表长度



list实际上就是一个指向链表头节点的指针。在实际使用时千万不要移动链表头结点指针！（有什么后果？）

# 算法

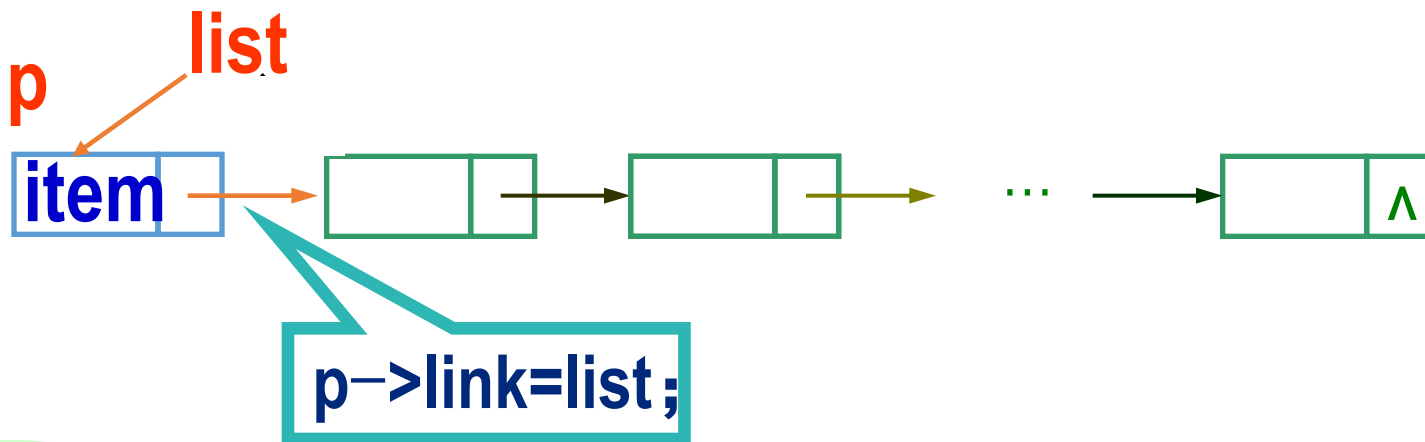
```
int getLength( Nodeptr list )           时间复杂度O(n)
{
    Nodeptr p;    /* p为遍历链表结点的指针 */
    int n=0;      /* 链表的长度置初值0 */
    for(p=list; p!=NULL; p=p->link)
        /* p依次指向链表的下一结点 */
        n++;     /* 对链表结点累计计数 */

    return n;    /* 返回链表的长度n */
}
```

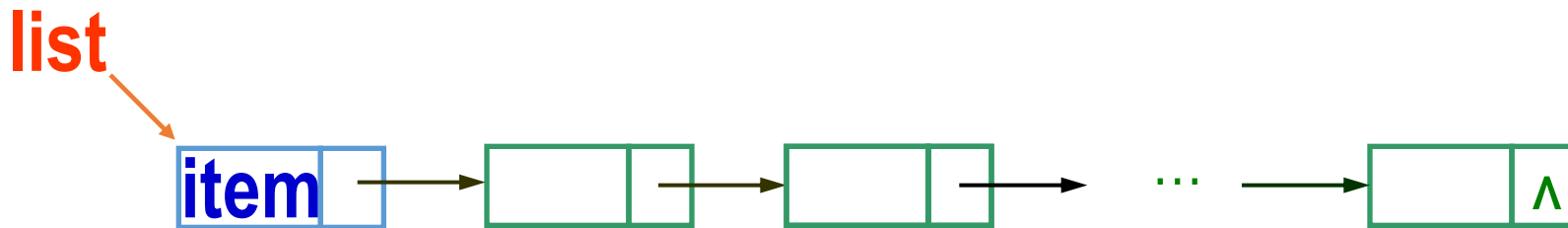
$(a_1, a_2, a_3, \dots, a_{n-1}, a_n)$  求n

### 3. 头插入

- 在非空线性链表的第一个结点前插入一个数据项为item的新结点



插入后



# 算法

有可能改变链表头的操作，都要返回新的链表头指针！！！！！！

```
Nodeptr insertFirst( Nodeptr list, ElemType item )
```

```
{
```

```
/* list指向链表第一个链结点 */
```

申请一个新结点

```
p=(Nodeptr)malloc(sizeof(Node));
```

```
p->data=item;
```

```
/* 将item赋给新结点数据域 */
```

```
p->link=list;
```

```
/* 将新结点指向原链表第一个结点*/
```

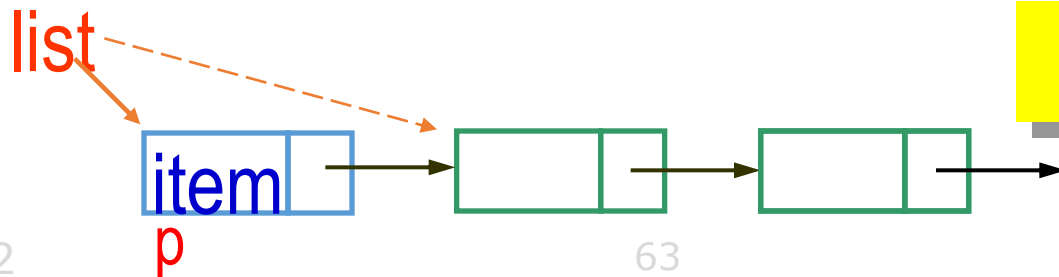
```
return p;
```

```
/* 将链表新头指针返回 */
```

```
}
```

应使用如下方式调用insertFirst函数：

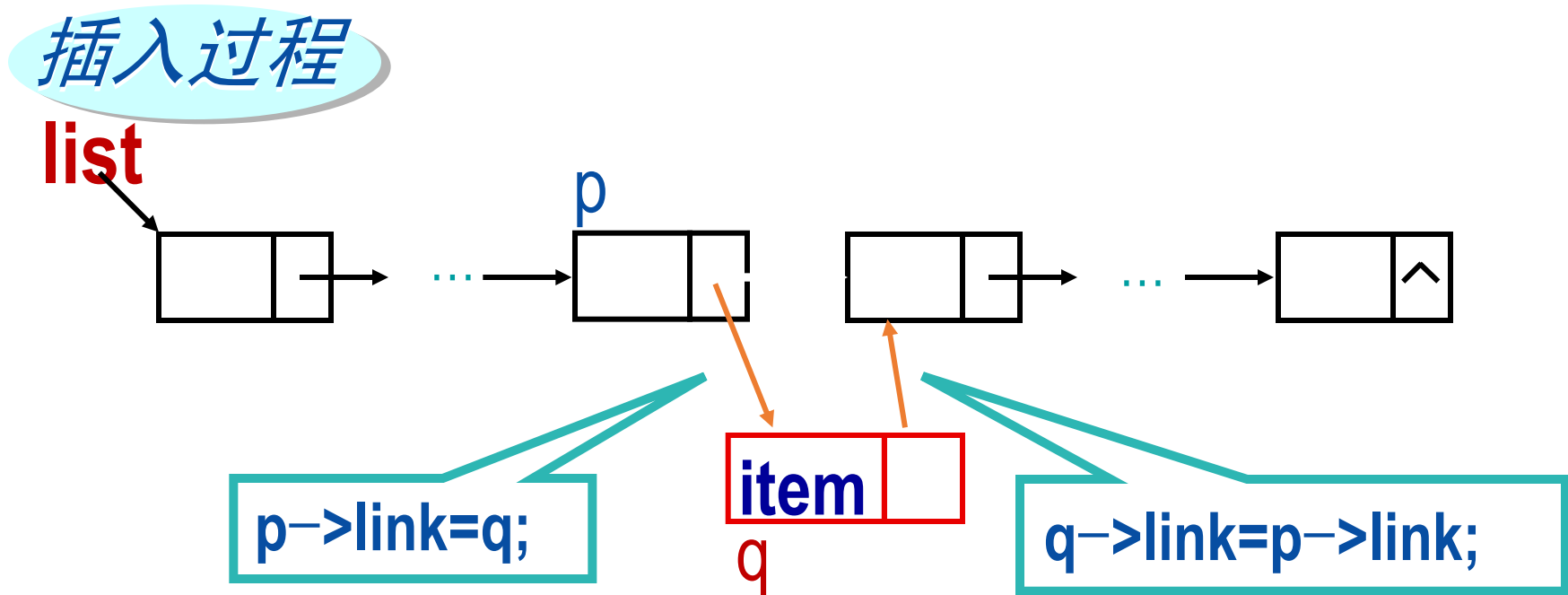
```
list = insertFirst(list, item);
```



时间复杂度： $O(1)$

# 4. 指定位置插入

- 在线性链表中由指针  $p$  指的链结点之后插入一个数据项为  $item$  的链结点





# 算法

```
void insertNode(Nodeptr p, ElemType item)
```

```
{
```

```
    Nodeptr q;
```

```
    q=(Nodeptr)malloc(sizeof(Node));
```

```
    q->data=item;          /* 将item送新结点数据域 */
```

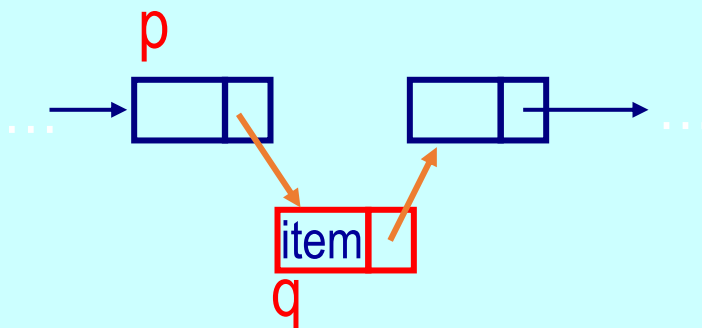
构造一个新结点

```
    q->link=p->link;
```

```
    p->link=q;
```

时间复杂度 $O(1)$

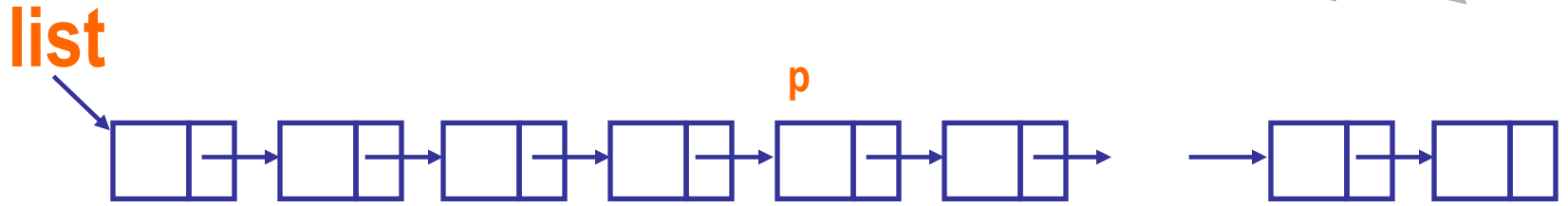
```
}
```



# 5. 指定位置插入

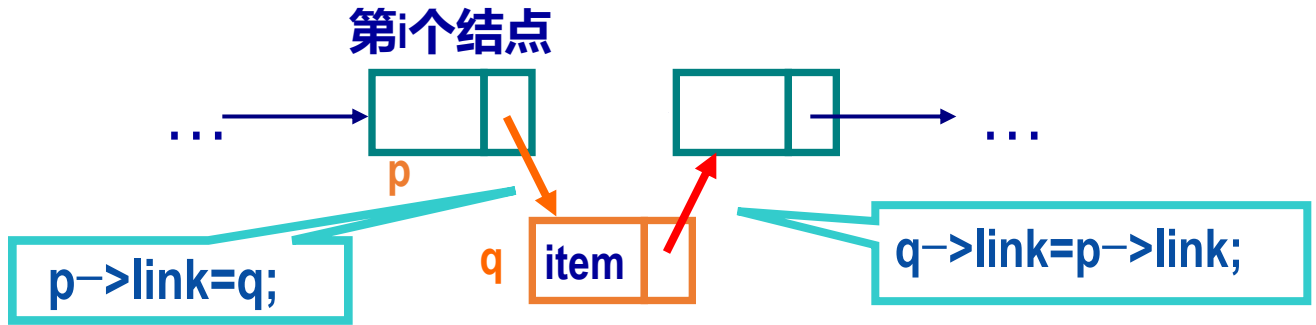
- 在线性链表中第 $n(n>0)$ 个结点后面插入一个数据项为 $item$ 的新结点

寻找第 $n$ 个结点



```
p = p -> link;
```

执行 $n-1$ 次



```
p -> link = q;
```

```
q -> link = p -> link;
```





```
void insertNode1( Nodeptr list, int n, ElemType item )
{
    Nodeptr p=list, q;
    int i;
    for(i=1;i<=n-1;i++){
        if(p->link==NULL)
            break;
        p=p->link;
    }
    q=(Nodeptr)malloc(sizeof(Node));
    q->data=item;
    q->link=p->link;
    p->link=q;
}
```

/\* 寻找第i个结点 \*/

/\* 不存在第i个结点 \*/

构造一个新结点

/\* 将新结点插入到第i个结点之后 \*/

时间复杂度O(n)



## 5a . 在有序线性链表中相应结点后面插入一个数据项为item的新结点

/\* 设list是一个有序增序链表，将元素elem插入到链表中 \*/

```
Nodeptr insertNode(Nodeptr list, ElemType elem)
```

```
{
```

```
    Nodeptr p,q, r;
```

```
    r = (Nodeptr)malloc(sizeof(Node)); //创建一个数据项为elem的新结点
```

```
    r->elem = elem; r->link = NULL;
```

```
    if(list == NULL) /* list是一个空表 */
```

```
        return r;
```

```
    for(p=list; elem > p->elem && p != NULL; q = p, p = p->link) /* 找到插入位置 */
```

```
    ;
```

```
    if( p == list){ /* 在头结点前插入 */
```

```
        r->link = p;
```

```
        return r;
```

```
    }
```

```
    else { /* 在结点q后插入一个结点 */
```

```
        q->link = r;
```

```
        r->link = p;
```

```
    }
```

```
    return list;
```

```
}
```

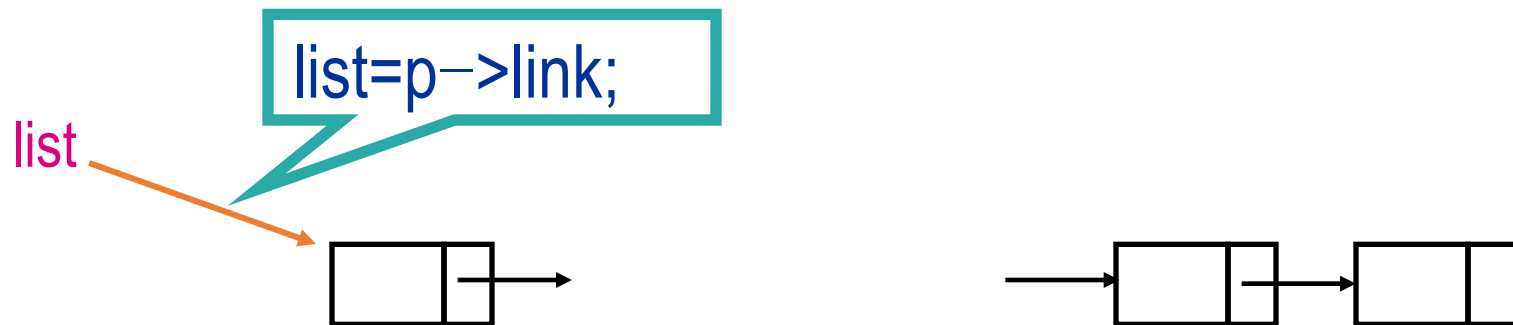
有可能改变链表头的操作，都要返回新的链表头指针！！！！！！

1. 在结点p前插入一个结点，必须要知道该结点的前序结点指针，在本程序中，q为p的前序结点指针；

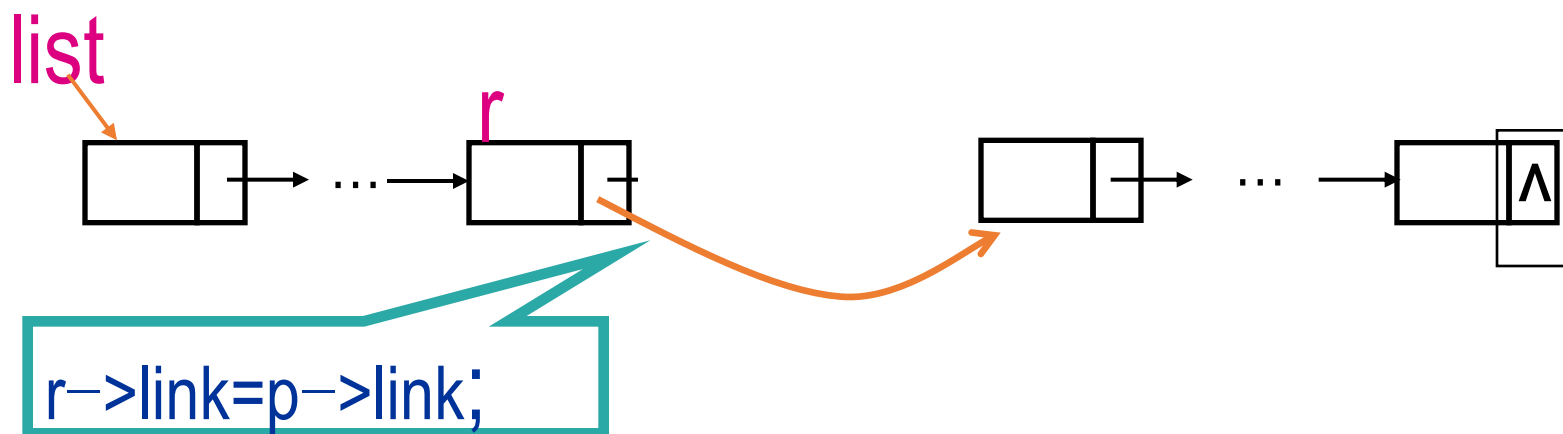
2. 应使用如下方式调用insertNode函数：  
list = insertNode(list, item);

时间复杂度O(n)

## 6. 从非空线性链表中删除p指向的链结点, 设p的直接前驱结点由r指出



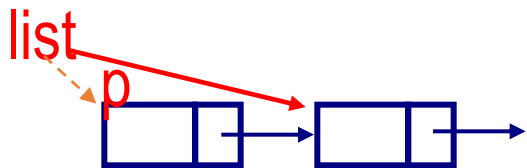
情况1：删除链表的第一个结点



情况2：删除链表中非第一个结点

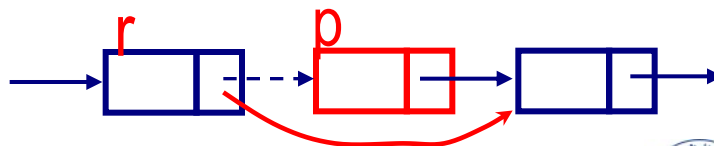
# 算法

```
Nodeptr deleteNode1( Nodeptr list, Nodeptr r, Nodeptr p )
{
    if(p==list)
        list=p->link;      /* 删除链表的第一个链结点*/
    else
        r->link=p->link; /* 删除p指的链结点*/
    free(p);              /* 释放被删除的结点空间*/
    return list
}
```



(情况1)

时间复杂度 $O(1)$

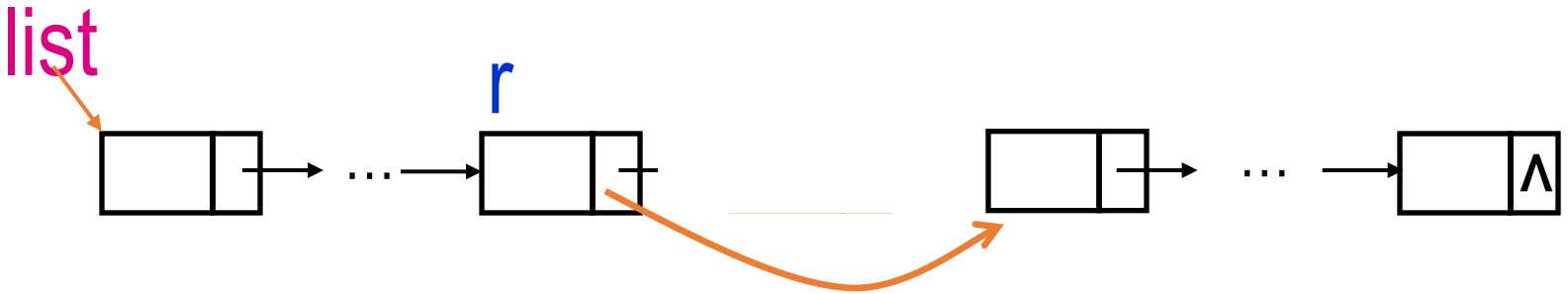
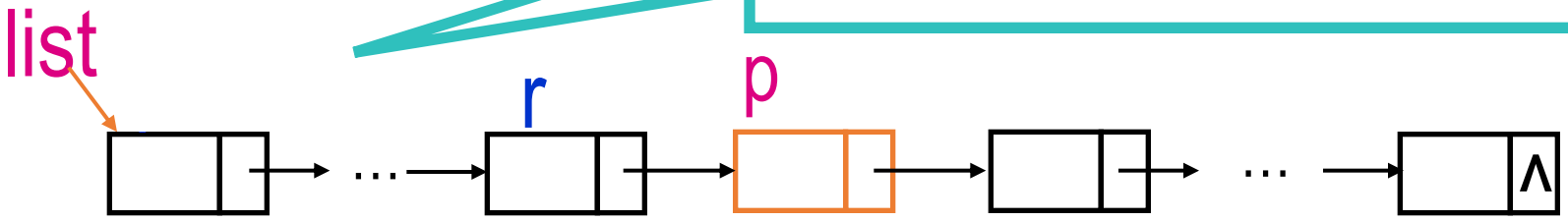


(情况2)

# 7. 从非空线性链表中删除p指向的链结点

设r是p的直接前驱结点指针

```
for(r=list; r->link!=p; r=r->link)  
;
```



# 算法

```
Nodeptr deleteNode2( Nodeptr list, Nodeptr p )
```

```
{  
    Nodeptr r;  
    if(p==list){  
        list=list->link;  
        free(p);  
    }  
    else{  
        for(r=list; r->link!=p && r->link!=NULL; r=r->link)  
            ;  
        if(r->link!=NULL){  
            r->link=p->link;  
            free(p);  
        }  
    }  
    return list;  
}
```

/\*当删除链表第一个结点\*/

/\*释放被删除结点的空间\*/

/\*移向下一个链结点\*/

时间复杂度 $O(n)$

寻找p结点的直接前驱 r





8

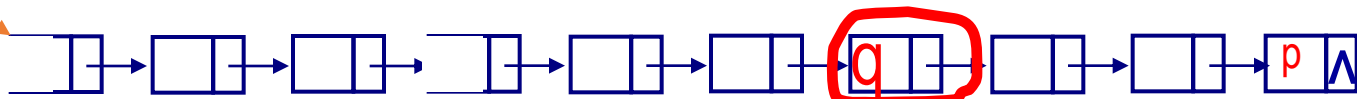
## 2009年硕士研究生入学考试计算机专业基础综合全国联考试题

请写一算法，该算法用尽可能高的时间效率找到由 list 所指的线性链表的倒数第 k 个结点。若找到这样的结点，算法给出该结点的地址，否则，给出NULL。

限制

1. 算法中不得求出链表长度;
2. 不允许使用除指针变量和控制变量以外的其他辅助空间。

list



```
p=p->link;
```

```
q=q->link;
```

倒数第k=4个结点

## 步骤

1. 设置一个指针变量  $p$ ，初始时指向链表的第1个结点；
2. 然后令  $p$  后移指向链表的第  $k$  个结点；
3. 再设置另一个指针变量  $q$ ，初始时指向链表的第1个结点；
4. 利用一个循环让  $p$  与  $q$  同步沿链表向后移动；当  $p$  指向链表最后一个结点时， $q$  指向链表的倒数第  $k$  个结点。



```
Nodeptr searchNode(Nodeptr list,int k)
```

```
{  
    Nodeptr p,q;  
    int i;  
    if(list!=NULL && k>0){  
  
        p=list;  
        for(i=1;i<k;i++){ /* 循环结束时, p指向链表的第k个结点 */  
            p=p->link;  
            if(p==NULL){  
                printf("链表中不存在倒数第k个结点! ")  
                return NULL;  
            }  
        }  
        q=list;  
        while(p->link!=NULL){  
            p=p->link;  
            q=q->link;  
        } /* p指向链表最后那个结点, q指向倒数第k个结点 */  
    }  
    return q; /* 给出链表倒数第k个结点(q指向的那个结点)的地址 */  
}
```



## 步骤

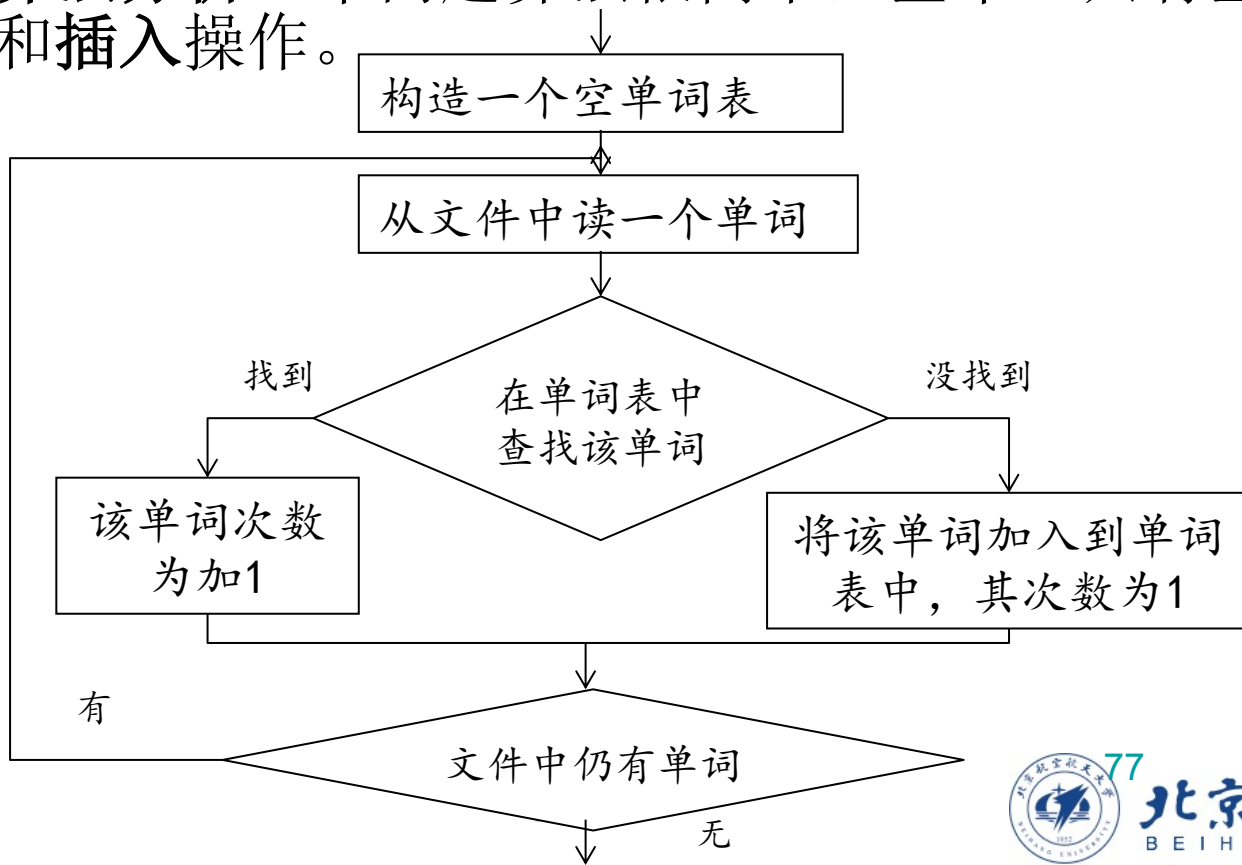
1. 设置一个指针变量 $p$ ，初始时指向链表的第1个结点；
- ② 然后令 $p$ 后移指向链表的第 $k$ 个结点： 执行 $k-1$ 次
3. 再设置另一个指针变量 $q$ ，初始时指向链表的第1个结点；
- ④ 利用一个循环让 $p$ 与 $q$ 同步沿链表向后移动； 当 $p$ 指向链表最后那个结点时， $q$ 指向链表的倒数第 $k$ 个结点。 执行 $n-k$ 次

若用 $n$ 表示链表中结点的个数, 对于任意 $k$  ( $1 \leq k \leq n$ )

$O(n)$

## 问题2.1：词频统计 – 链表

- 问题：编写程序统计一个文件中每个单词的出现次数（词频统计），并按字典序输出每个单词及出现次数。
- 算法分析：本问题算法很简单，基本上只有查找和插入操作。



## 问题2.1：词频统计 – 链表

由于本问题有如下特点：

1. 问题规模不知（即需要统计的单词数量未知）
2. 单词表需要频繁的执行插入操作

因此，采用顺序表（数组）来构造单词表面临如下问题：

1. 单词表长度太小，容易满，太大，空间浪费
2. 插入操作效率低（经常需要移动大量数据）

而链表具有动态申请结点（空间利用率高），插入和删除结点操作不需要移动结点，插入和删除算法效率高！但单词查找效率低（不能用折半等高效查找算法）

## 问题2.1: 词频统计 — 链表 (代码实现)

```
#include <stdio.h>
#include <string.h>
#include <stdlib.h>
#define MAXWORD 32
struct node {
    char word[MAXWORD];
    int count;
    struct node *link;
}; //单词表结构
struct node *Wordlist = NULL; //单词表头指针
int getWord(FILE *fp, char *w);
int searchWord(char *w);
int insertWord(struct node *p, char *w);
```

**Wordlist**是一个全局变量!

```
int main()
{
    char filename[32], word[MAXWORD];
    FILE *bfp;
    struct node *p;

    scanf("%s", filename);
    if((bfp = fopen(filename, "r")) == NULL){ //打开一个文件
        fprintf(stderr, "%s can't open!\n", filename);
        return -1;
    }
    while( getWord(bfp, word) != EOF) //从文件中读入一个单词
        if(searchWord(word) == -1) { //在单词表中查找插入单词
            fprintf(stderr, "Memory is full!\n");
            return -1;
        }
    for(p=Wordlist; p != NULL; p=p->link) //遍历输出单词表
        printf("%s %d\n", p->word, p->count);
    return 0;
}
```

## 问题2.1: 词频统计 - 链表 (代码实现)

/\*在链表中p结点后插入包含给定单词的结点, 同时置次数为1\*/

```
int insertWord(struct node *p, char *w)
{
    struct node *q;

    q = (struct node *)malloc(sizeof(struct node));
    if(q == NULL) return -1; //没有内存空间
    strcpy(q->word, w);
    q->count = 1;
    q->link = NULL;
    if(Wordlist == NULL) //空链表
        Wordlist = q;
    else if (p == NULL){ //插入到头结点前
        q->link = Wordlist;
        Wordlist = q;
    }
    else {
        q->link = p->link;
        p->link = q;
    }
    return 0;
}
```

在本程序中：  
查找算法复杂度为 $O(n)$   
插入算法复杂度为 $O(1)$

/\*在链表中查找一单词, 若找到, 则次数加1; 否则将该单词插入到有序表中相应位置, 同时次数置1\*/

```
int searchWord(char *w)
{
    struct node *p, *q=NULL; //q为p的前序结点指针
    for(p=Wordlist; p != NULL; q=p, p=p->link){
        if(strcmp(w, p->word) < 0)
            break;
        else if(strcmp(w, p->word) == 0){
            p->count++;
            return 0;
        }
    }
    return insertWord(q, w);
}
```



## 问题2.1: 词频统计 - 链表

思考

- 采用链表方式构造单词表具有如下特点:
  - 优点
    - 由于采用动态申请结点, 能够适应不同规模的问题, 空间利用率高
    - 算法简单, 插入操作效率高
  - 不足
    - 由于采用顺序查找, 单词查找效率低

还有更好的单词表的构造及查询方法吗?



## 问题2.2: 多项式相加 (链表实现) \*

【问题描述】编写一个程序实现任意 (最高指数为任意正整数) 两个一元多项式相加。

【输入形式】从标准输入中读入两行以空格分隔的整数，每一行代表一个多项式，且该多项式中各项的系数均为0或正整数。对于多项式  $a^n x^n + a^{n-1} x^{n-1} + \dots + a^1 x^1 + a^0 x^0$  的输入方法如下： $a^n$   $n$   $a^{n-1}$   $n-1$  ...  $a^1$   $1$   $a^0$   $0$  即相邻两个整数分别表示表达式中一项的系数和指数。在输入中只出现系数不为0的项。

【输出形式】将运算结果输出到屏幕。将系数不为0的项按指数从高到低的顺序输出，每次输出其系数和指数，均以一个空格分隔。最后要求换行。

【样例输入】

```
54 8 2 6 7 3 25 1 78 0
43 7 4 2 8 1
```

【样例输出】

```
54 8 43 7 2 6 7 3 4 2 33 1 78 0
```

【样例说明】输入的两行分别代表如下表达式：

$$54x^8 + 2x^6 + 7x^3 + 25x + 78$$

$$43x^7 + 4x^2 + 8x$$

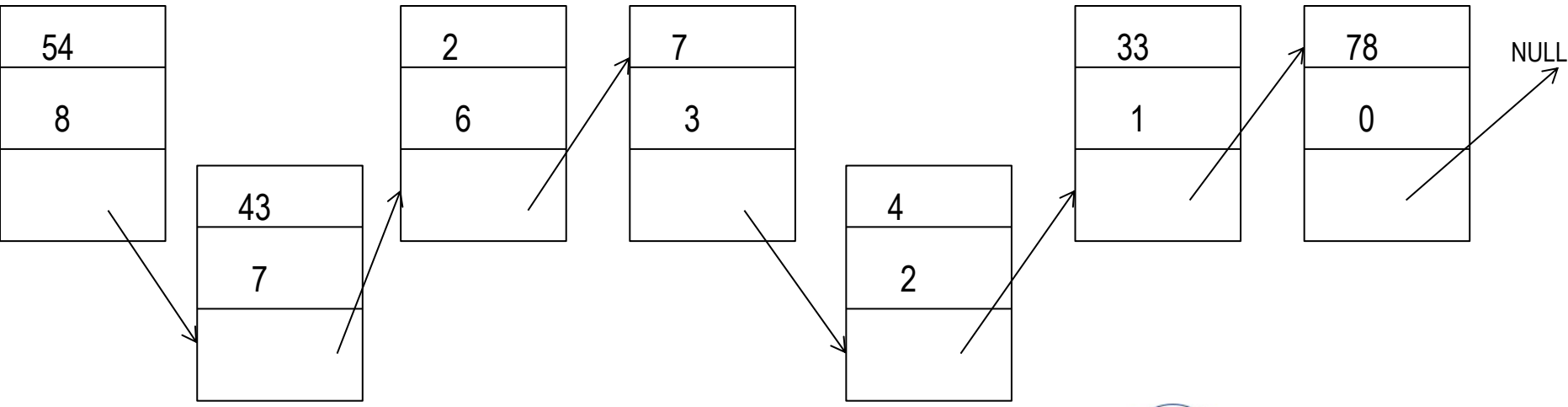
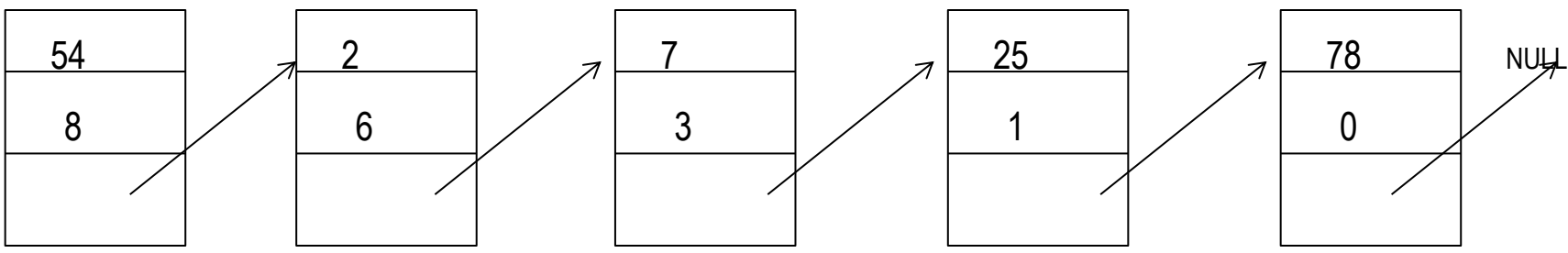
其和为

$$54x^8 + 43x^7 + 2x^6 + 7x^3 + 4x^2 + 33x + 78$$



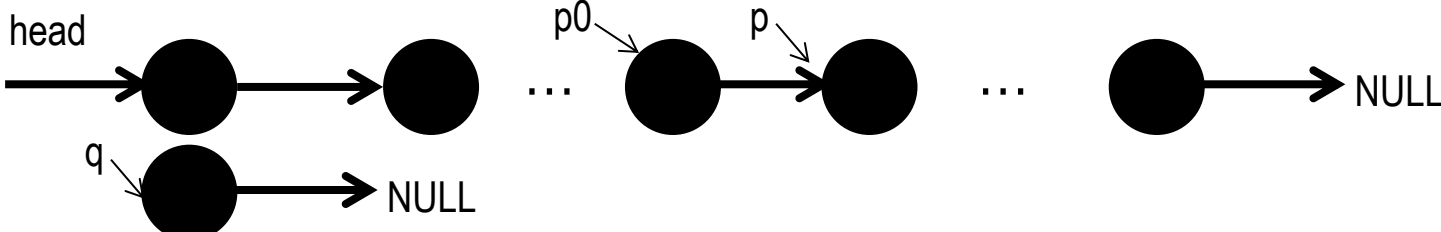
# 问题2.2: 算法设计\*

- 首先读入第一个多项式，并将其生成一个链表；
- 然后依次将第二个多项式每一项插入（或合并）第一个多项式中

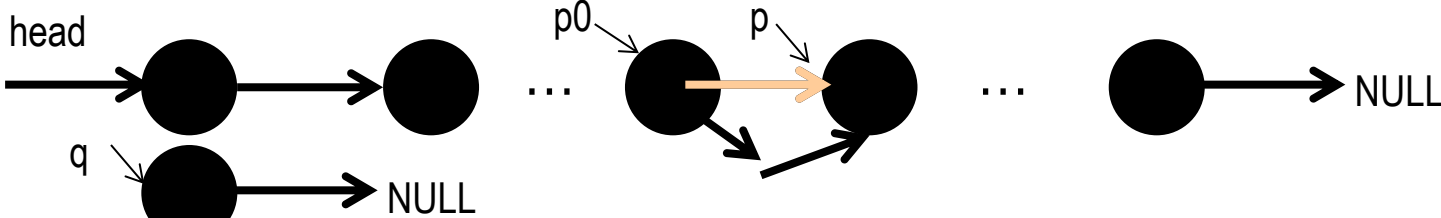


# 问题2.2: 算法设计\*

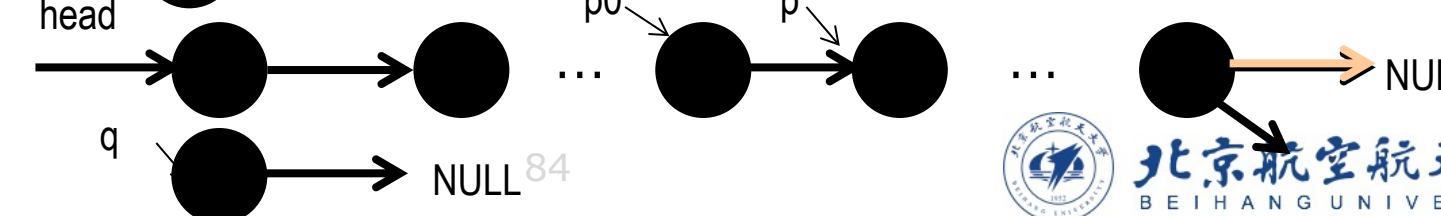
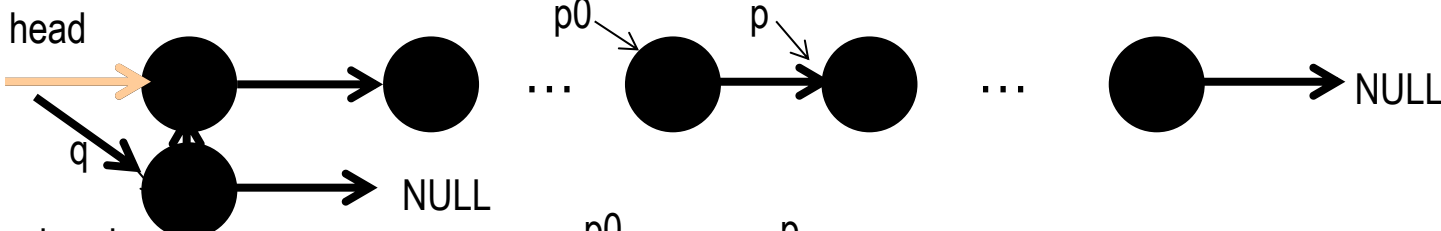
- 将第二个多项式的一个节点加到第一个多项式中有下面几种情况
  - 第一个多项式中存在相同指数的节点: 系数直接相加



- 第一个多项式中不存在相同指数的节点: 插入到某个节点前



- 第一个多项式中不存在相同指数的节点: 插入到头节点前, 或尾节点后



构造类型 - 数组和指针



## 问题2.2:

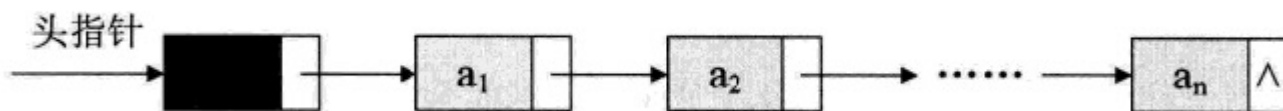
```
//c3_4c.c
#include <stdio.h>
#include <stdlib.h>
struct Node { //一个多项式节点结构
    int coe; //系数
    int pow; //幂
    struct Node *next;
};
int main()
{
    int a,n;
    char c;
    struct Node *head,*p,*q,*p0;
    head = p = NULL;
    do { //创建一个链表存放第一
        scanf("%d%d%c", &a, &n, &c);
        if( head == NULL)
            head = p = (struct Node *)malloc(sizeof(struct Node));
        else {
            p->next = (struct Node *)malloc(sizeof(struct Node));
            p = p->next;
        }
        p->coe = a; p->pow = n; p->next = NULL;
    } while ( c != '\n');
```

```
do { //将第二个多项式的每个项插入到第一个多项式链表中
    scanf("%d%d%c", &a, &n, &c); //生成第二个多项式的一个节点
    q = (struct Node *)malloc(sizeof(struct Node));
    q->coe = a; q->pow = n; q->next = NULL;
    for(p=head; p!=NULL; p0=p,p=p->next) {
        if(q->pow > p->pow) {
            if(p==head) { q->next = head; head = q; break; } //插入到头节点前
            else { q->next = p; p0->next = q; break; } //将q插入到p前
        }
        else if(q->pow == p->pow) { p->coe += q->coe;
            free(q);break; } //指数相等，系数相加
        }
        if(p== NULL) p0->next = q; //将q插入到尾节点后
    } while ( c != '\n');
    for(p=head; p!=NULL; p=p->next)
        printf("%d %d ", p->coe,p->pow);
    return 0;
}
```



有时在构造链表时会给链表设置一个**头指针**（非头结点指针，其操作list->link指向链表的头结点），如下图所示：

### 带头指针的非空链表



### 带头指针的空链表



请思考这样做有什么好处？

**设置头指针的最大好处是对链表结点的插入及删除操作统一了（不用考虑是否是头结点）。其数据域一般无意义（有时也可存放链表的长度）。**

## 2.3.4 链式存储结构的特点

无论位于链表何处，无论链表的长度如何，插入和删除操作的时间都是 $O(1)$ 。

### 1. 优点

- (1) 存储空间动态分配，可以根据实际需要使用。
- (2) 不需要地址连续的存储空间。
- (3) 插入/删除操作只须通过修改指针实现，不必移动数据元素，操作的时间效率高。

### 2. 缺点

- (1) 每个链结点需要设置指针域(占用存储空间小)。
- (2) 是一种非连续存储结构，查找、定位等操作要通过顺序遍历链表实现，时间效率较低。

时间为 $O(n)$



# 思考

线性表可以采用顺序存储结构，也可以采用链存储结构，在实际问题中，应该根据什么原则来选择其中最合适的一种存储结构？



# 顺序存储结构与链表结构的比较

## 存储分配方式

- 顺序存储用一段连续的存储单元依次存储线性表的数据元素
- 链表采用链式储存结构，用一组任意的存储单元存放线性表的元素

## 时间性能

- 查找
  - 顺序存储 $O(1)$
  - 链表 $O(n)$
- 插入和删除
  - 顺序存储需要平均移动表长一半的元素，时间为 $O(n)$
  - 链表在给出结点位置后，插入和删除时间仅为 $O(1)$

## 空间性能

- 顺序存储需要事先分配存储空间，分大了浪费，分小了易发生溢出
- 链表不需要事先分配存储空间，需要时分配结点，元素个数不受限制

## 结论：

1.若线性表需要频繁查找（通讯录），较少进行插入和删除操作时，宜采用顺序存储结构。若需要频繁插入和删除时（如单词表），宜采用链表结构。

2.当线性表中的元素个数变化较大或者根本不知道有多大时（如单词表），最好用链表结构。而如果事先知道线性表的大致长度，用顺序结构效率会高些。

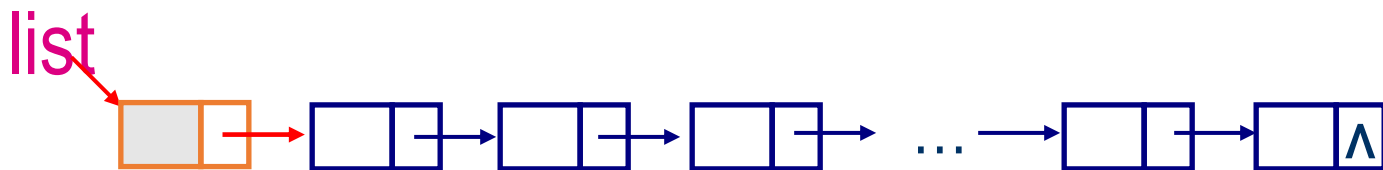


## 思考题

已知List为没有头结点的单链表中第一个结点的指针,每个结点数据域存放一个字符,该字符可能是英文字母或数字字符或其他字符,编写算法构造三个以带头结点的单向链表表示的线性表,使每个表中只含同一类字符。

(要求用尽量少的时间和尽量少的空间)

### 带头结点的线性链表



## 作业

设计一个算法删除一个单链表中元素值最大的节点。假设数据元素均为整数,且最大值节点唯一。



北京航空航天大学  
BEIHANG UNIVERSITY

Thanks ! & Any Questions ?