[Reference](https://mp.weixin.qq.com/s?__biz=MzIwMzY1OTU1NQ==&mid=2247486105&idx=1&sn=d672c62e72167c6cc1971e88adada862&chksm=96cd4ad5a1bac3c31634359f25200e43014326080986636a2994d14015ea82fe38f4139db27e&mpshare=1&scene=1&srcid=&key=1873ed4ed1cb893e0ce9d9c3efe2fe412885c9e5416931c4e99779074b5afec6c2a75789badcb9b7c45c7e234dbd3b59e66e3902b76c281fdbc788f1a692c77c9f582c23a93e56ae16c18bfbf0eb8fe0&ascene=1&uin=Mjc3ODQ1MTk0MA%3D%3D&devicetype=Windows+10&version=62060739&lang=zh_CN&pass_ticket=U1Rycznd0MxcqdQDoQj52v1tXH53Lbffgbif%2FRFGmW%2F1SWy%2F7LEwVLTizUTqU4gb)

# Part 1 Concept

## Base

逻辑结构：数据之间的逻辑关系

储存结构：数据元素及其关系在储存器中的表示

顺序储存结构：逻辑上相邻的节点在物理位置上相邻的储存单元里

规模：求解问题的输入量 （n的大小）

渐进时间复杂度：n趋近无穷大，时间复杂度的数量阶

limT(n)/n`3=lim(2n`3+3n`2)/n`3=2; T(n)和n`3同阶。记作T(n)=O(n`3)=O(f(n))<=C。f(n),即渐进时间复杂度，常称为时间复杂度。f(n)一般是频度最大的语句频度

区别空串(长度为0)和空白串

‘\0’作为串值的终结。

为了提高储存密度：

Typedef struct node{

char data[NodeSize];

struct node \* next;

}

<string.h> 里定义的字符串

Typedef struct{

char \* ch;

int length;

}Hstring

char \* p="abcdef";

printf("%d\n",strlen(p));

## 文件

**逻辑结构和储存结构**：

数据结构的存储结构是和相应的数据在内存中的物理地址之间的关系有关。而逻辑结构只是描述**数据**之间的关系（三大逻辑结构的一种）。

举例说，线性表（元素之间的逻辑关系是线性的）可以是顺序存储的方式，即所有元素相邻存放，在物理地址上是连续的（存储结构）；而对于链式存储的线性表，他的所有元素之间不一定是线性相连的，可能是第一个结点（元素）的地址为0x123,而第二个元素又出现在物理地址0x100上。也就是说逻辑结构是线性的但是存储结构不一定就是线性的了。

**文件的组织方式（存储结构）：**

在外存的组织结构：顺序组织，索引组织。。。

**顺序文件：**

逻辑结构和物理结构顺序一致

**索引文件：**

**索引顺序文件：**

ISAM（Indexed Sequential Access Method）：索引顺序存取

静态索引结构，

VSAM(Virtual Storage Access Method):虚拟储存存取方法

也是顺序索引结构，采用B+树的动态索引结构

**散列文件：**

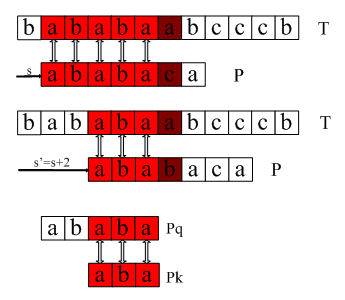
**多关键字文件：**

# Part 2 数据结构

## 串

子串定位又模式匹配Pattern Matching。子串称模式串，主串称目标串；

朴素的串匹配：



## 顺序表

Sequential list

1. #define ListSize 100
2. Typedef **int** DataType;
3. Typedef **struct**{
4. DataType data[ListSize];
5. Int length;
6. }Seqlist

插入：

指在表第i（1<=i<=n+1）位置，插入x，使n长度变成n+1

void insertList(Seqlist \* L,DataType x,int i)

{

if(i<1||i>L->length+1) //可以在length+1上插

Error(“position error”);

If(L->length>=ListSize)

Error(“overflow”);

For(j=L->length-1;j>=i-1;j--)

{

L->data[j+1]=L->data[j];

}

L->data[i-1]=x;

L->length++;

表中位置i比向量下标大1，

## Stack

**顺序栈：**

#define StackSize 100

typedef struct{

DataType data[StackSize];

int top;

} SeqStack;

Int StackEmpty(SeqStack \* s)

{

Return s->top==-1;

}

void push(SeqStack \* s,DataType x)

{

s->data[++s->top]=x;

}

**链栈**

struct StackNode{

DataType data;

struct StackNode \* next;

}

Typedef struct{

StackNode \* top;

}LinkStack;

指向指针的指针。LinkStack \*s;

initStack :s->top=NULL;

int StackEmpty

{

return s->top==NULL;

};

入栈

void push(LinkStack \* s,x)

{

StackNode \* p=(StackNode \*)malloc(sizeof(StackNode));

p->data=x;

p->next=s->top;

s->top=p;

}

## Queue

区别于线性表list

双链表

Typedef struct dlistnode{

DataType data;

Struct dlistnode \* prior,\* next;

}DlistNode;

**队列：**

struct Queue

{

char data[QueueSize];

int front,rear;

};

**循环队列：**

顺时针旋转，rear逻辑上在front之前,m[0]上有元素，rear提前有个空位置时front=(rear-len+m)%m。

rear没有时（题目中会说**rear指向尾元素的位置**），（rear-len+m+1）%m

Rear，front的长度都<QuenueSize，当转一圈后，从num为1；

typedef struct

{

char data[QueueSize];

int count;

int rear,front;

}CirQueue;

void initCirQueue(CirQueue \*cirQueue)

{

cirQueue->count=0;

cirQueue->front=0;

cirQueue->rear=0;

}

int isCirQueueEmputy(CirQueue \*cirQueue)

{

return cirQueue->count==0;

}

int isCirQueueFull(CirQueue \*cirQueue)

{

return cirQueue->count==QueueSize;

}

void EnQueue(CirQueue \*cirQueue,char x)

{

if(isCirQueueFull(cirQueue))

{

printf("cirQueue is Full");

exit(1);

}

cirQueue->data[cirQueue->rear]=x; //队尾进入

int num= (++cirQueue->rear)%QueueSize;//如果队列长10，

cirQueue->rear=num;

cirQueue->count++;

}

char DeQueue(CirQueue \*cirQueue)

{

if(isCirQueueEmputy(cirQueue))

{

printf("cirQueue is Emputy");

exit(1);

}

char str=cirQueue->data[cirQueue->front];

cirQueue->front=(++cirQueue->front)%QueueSize;都是++，顺序循环

cirQueue->count--;

return str;

}

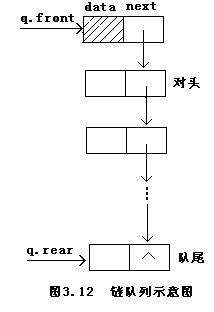
**循环队列**

Typedef struct quenuenode{

DataType data;

Struct quenuenode \* next;

}QuenuNode;



Typedef struct{

QueueNode \* front;

QueueNode \* rear;

}LinkQueue;

循环单链表：

rear->next->next来表示front

双链表：

Typedef struct dlistnode{

DataType data;

Struct dlistnode \* prior,\*next;

}DlistNode;

## Graph

 connected component

 connected component of an [undirected graph](http://en.wikipedia.org/wiki/Undirected_graph) is a [subgraph](http://en.wikipedia.org/wiki/Subgraph) in which any two vertices are [connected](http://en.wikipedia.org/wiki/Connected_graph) to each other by [paths](http://en.wikipedia.org/wiki/Path_(graph_theory)), and to which no more vertices or edges (from the larger graph) can be added while preserving its connectivity. That is, it is a maximal connected subgraph.

1.undirected graph

2.in graph,any two vertices are connected

3.connected component is maximal connect subgraph(to which no more vertices or edges)

**strongly connected component**

A [directed graph](http://en.wikipedia.org/wiki/Directed_graph) is called *strongly connected* if there is a [path](http://en.wikipedia.org/wiki/Path_(graph_theory)) from each [vertex](http://en.wikipedia.org/wiki/Vertex_(graph_theory)) in the graph to every other vertex. In particular, this means paths in each direction; a path from *a* to *b* and also a path from *b* to *a*.(两个点双向都能达到就是强连通的)

The strongly connected components of a directed graph *G* are its maximal strongly connected [subgraphs](http://en.wikipedia.org/wiki/Subgraph). If each strongly connected component is contracted to a single vertex, the resulting graph is a directed acyclic graph, the condensation of *G*. A directed graph is acyclic if and only if it has no (nontrivial) strongly connected subgraphs (because a cycle is strongly connected, and every strongly connected graph contains at least one cycle).

1.directd graph

2.maximal strongly connected,

3.a single vertex or cycle

源(source):当一个强连通分量没有一条从其他连通分量指向这个连通分量中的任何一个顶点时(就是没有入度)，我们就可以称之为一个源；比如说上图的A就是一个源。

汇(sink):当一个强连通分量没有一条从这个连通分量指向其他连通分量中的顶点时(就是没有出度)，我们就可以称之为一个汇；比如说上图的{G,H,I,J,K}就是一个汇。

## 图

**图**

G=(V,E),v是Vertex，E是Edge。

<Vj,Vi>。有向边也称弧（Arc），始点是弧尾（Tail），终点弧头（Head）；

**有向图（Digraph）顶点集和边集的表示：**

V（G1）={v1,v2,v3}

E(G1)={<v1,v2>,<v2,v1>,<v2,v3>}; //无向图不用写重复的

无向图（unDigraph）的度（Degree）：

**D(v)关联该顶点边的数目**

有向图是D(v)=ID(v)+OD(v);

**路径**：只要能从一个顶点到另个。如果除了开始和结尾有重复，则称是**简单路径，**若开始结尾相同就称**简单回路或简单环**

**无向图：**

Vj到Vi有路径称Vj和Vi连通，若V（G）任意连通即有路径称G为**连通图**（Connecte Graph）

极大连通子图是**连通分量**（Connected Component）。极大指走过节点多。即全部走未走过的节点

**有向图：**

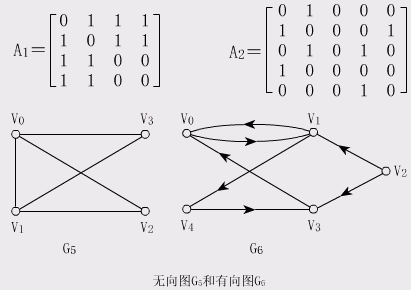
有根图：存在顶点v能到所有顶点，称图为有根图，v称为图的**根**

连通图有向时是**强连通图**

**图的储存结构**

1. **邻接矩阵表示法**

 　设G=(V，E)是具有n个顶点的图，则G的邻接矩阵是具有如下性质的n阶方阵：

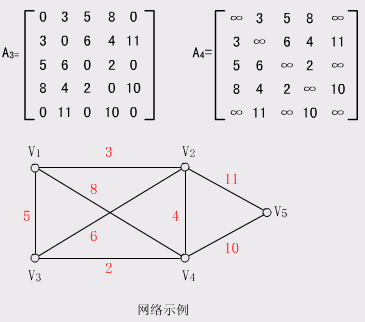


**总结：**

1、有几个顶点就有几行几列。做深度遍历时同邻接表在顶点(Vertex)的跳动，它是在不同行之间的跳动。

3．网络的邻接矩阵

若G是网络，则邻接矩阵可定义为：  
        
    
  
  其中：  
      wij表示边上的权值；  
      ∞表示一个计算机允许的、大于所有边上权值的数。  
【例】下面带权图的两种邻接矩阵分别为A3和A4。

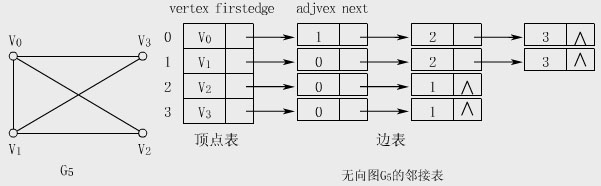


1. **邻接表表示**

把所有邻接于vi的顶点链接称链，称vi的**邻接表**。

无向图的邻接表为**边表**。

有向图是**出边表**。有向图只链接从某顶点的出边，逆邻接表是**入边表**



**邻接矩阵（Adjacency Matrix）的逻辑结构**

#define MaxVertexNum 100;

typedef char VertexType; //顶点类型

typedef int EdgeType; //权值类型

typedef struct{

VertexType vexs[MaxVertexNum];

EdeType edges[MaxVertexNum][ MaxVertexNum];

int n,e;

}

**邻接表的逻辑结构：**

typedef struct node{

int adijvex;

struct node \* next;

}EdgeNode; //边表节点

typedef struct vnode{

VertexType vertex;

EdgeNode \* firstedge;

}VertexNode; //顶点节点

typedef VertexNode AdjList[MaxVerexNum];

typedef struct{

AdjList adjlist;

int n,e; //n当前顶点数，e 边数

}ALGraph;

**图的遍历**

Typedef enum{TRUE,FALSE} Boolean;

Boolean visited[MaxVertexNum];

void DFSTtraverse(ALGraph \* G)

{

int i;

for(i=0;i<G->n;i++)

Visited[i]=FALSE;

for(i=0;i<G->n;i++)

If(!visited[i])

DFS(G,i);

}

void DFS(ALGraph \* G,int i)

{

EdgeNode \* p;

print(“vist vertex:%c”,G->adjlist[i].vertex);

visted[i]=TRUE;

p=G->adjlist[i].firstedge;

while(p)

{

if(!visted[p->adjvex])

DFS(G,p->adjvex);

p=p->next;

}

}

void DFSM(Mgraph \* G, int i)

{

int j;

printf(“visit vertex:%c”,G->vexs[i]);

visited[i]=TRUE;

for(j=0;j<G->n;j++)

if(G->edges[i][j]==1&&!visited[j])

DFSM(G,j)

}

**广度遍历(Breadth-First Traversal)**

void BFS(ALGraph \* G,int k)

{

int i;

CirQuenu Q;

EdgeNode \* p;

initQueue($=&Q);

printf(“visit vertex:%c”,G->adjlist[k].vertex);

visted[k]=TRUE;

EnQueue(&Q,k);

while(!Queue(&Q)){

i=DeQueue(&Q);

p=G->adjlist[i].firstedge;

while(p)

{

if(!visited[p->adjvex]){

printf(“visit vertex:%c”,G->adjlist[k].vertex);

visited[p->adjvex]=true;

EnQueue(&Q,p->adjvex);

}

p=p->next;

}

}

}

**生成树和最小生成树**

无回路的连通图，没确定根，也称**自由树**。

连通图的子图如果是包含G的所有顶点的树，称**生成树**(Spanning Tree)

实际生成树是有N个顶点和有n-1条边的极小连通子图。极小指边数最小。

深度遍历（DFS）：从v开始，搜索**第一个**邻接点w，从w开始重新DFS。若有未访问过的，**随便**选一个开始DFS。

深度优先生成树：

广度优先生成树：

**最小生成树**

Prim算法：

1. 从0开始，用虚线相连确定权最小用实线相连，依次到实线那头继续。。

Kruska算法：

## 多维数组和广义表

计算机内存是一维的。一般不对数组插入，采用顺序存储。

1. 行优先顺序：a11,a12…,a

LOC(aij)=LOC(a11)+[(i-1)\*n+j-1]\*d

1. 列优先:a11,a21..an1,a12,a22…an2….

**矩阵的压缩存储**

**对称矩阵（**对角线上有数据的**）**

N阶方阵 Aij=Aji;一般只存一半；压缩存储时

对角线上是aii.

**三角矩阵（**对角线上有数据的**）**

上三角矩阵：下三角（不包括主角线）均为常数；

**对角矩阵：**

**稀疏矩阵：**

1. 三元组表
2. 带行表的三元组表

**广义表（Lists）**

对线性和树的推广。

LS=（a1,a2,….an）; a1是表头，其余元素组成的表LS称表尾

A=(x,L)=(x,(a,b)) 或(x,L(a,b))

**空表 E=（）。不能运算。tail（（））=( );**

**tail( (( )) )=head( (( )) )=( )。(( ),( ))不对？**

1. 子表大写
2. 深度指单方向括号的层数。A是2
3. 线性表都是原子项。与树对应纯表。
4. head(LS) 原子和字表都行

广义表的图形表示：

**A( ) B(e) C(a,(b,c,d)) D(A(),B(e),C(a,(b,c,d)))**

**E((a,(a,b),((a,b),c)))**



纯表: 与树对应

再入表：允许节点再入

递归表：允许节点递归

广义表是线性表和树的推广

# Part 3 Tree

常见的数组、链表、栈和队列都是线性结构，在存储大量数据时访问速度比较慢，而树(tree)则是一种非线性结构，使得访问时间复杂度降低到O(logn)。

术语

度：节点孩子数量。一棵树的度指最大孩子树。

阶：m阶B-tree指每个节点最多有m个子节点。

深度(h)=高度

 森林(Forest)：由多棵不相交的树的集合。

## 3.1 Binary Tree

即二叉树。二叉树中每个元素都恰好有两棵子树（其中一个或两个可能为空）。

**二叉树常见的性质：**

1. 深度为K至多有 2`k-1
2. n0=n2+1

证明：

(1)n=n0+n1+n2 (n0,n1,n2指度为2，n指数量) 从自身角度

(2)n=n1+2\*n2+1 算孩子的总数

### 完全二叉树

只有最下两层阶段的度数可以<2，且集中在最左边

从根节点编号1，2，3.。。。。。。

i>1,双亲编号i/2;

### 满二叉树

完全二叉树的特例，每一个非叶子节点都存在左右子树，并且二叉树中所有的叶子节点都在同一层中。

### 存储结构

顺序存储结构

链式存储结构：

1. **typedef** **struct** BinTNode{
2. **char** data;
3. BinTNode \* lchild,\* rchild;
4. }BinTNode;
5. **typedef** BinTNode \* BinTree;

**二叉树，森林，树的转换：**

（1）都去掉双亲到右孩子的线

二叉树—>右孩子到双亲

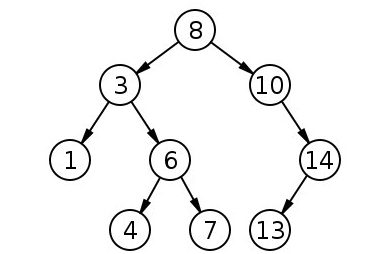
森林-->连接兄弟。双亲到右孩子

顺时针旋转45·

**二叉树的遍历：**

### Binary Sort Tree

即二叉排序树，又称“二叉查找树”、“二叉搜索树”。 左孩子<双亲<右孩子，中序遍历得到有序递增序列。通常采用二叉链表作为存储结构。



插入和生成：

<http://61.142.209.149:88/sjjg/flashhtml/erchapaixushengcheng.htm>

每次插入的新的结点都是二叉排序树上新的叶子结点，在进行插入操作时，不必移动其它结点，只需改动某个结点的指针。搜索、插入、删除的时间复杂度等于树高，期望O(logn)，最坏O(n)。

虽然二叉排序树的最坏效率是O(n)，但它支持动态查找，且有很多改进版的二叉排序树可以使树高为O(logn)，如AVL、红黑树等。

## 3.2 B树

### B-tree

即Balanced Tree，目前大部分数据库系统及文件系统都采用B-Tree或其变种B+Tree作为索引结构。

度的大小取决于磁盘大小

B-tree主要用在文件系统以及部分数据库索引，比如MongoDB

一棵度为m的B-树称为m阶B-树。有的B-树所以辅助信息都存在叶节点，内部节点都存关键字和指向孩子的指针，使内部节点的度数尽可能的大

定义：

一个m阶B-树数据域形式

(j,Po,K1,P1,K2….Kj,Pj) keys(Po)<K1<keys(P1)<K2<…….Kj<keys(Pj)

1.叶节点位于同一层，即为树的高度

2.非叶子关键字[m/2]-1<=j<=m-1 子树[m/2]<=zishu<=m。根至多m-1关键字，m棵子树

3.1001的B-树结构

#define Max 1000 //关键字max=m-1

#define Min 2000 //[m/2]-1

Typedef int KeyType

Typedef struct node{

int keynum;

KeyType key[Max+1]; //key[0]不用

struct node \* parent;

struct node \* son[Max+1];

}BtreeNode;

typedef BtreeNode \* BTree;

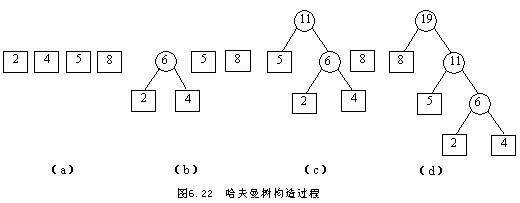
4.查找：

5.插入和生成：

## 3.3 最优二叉树

也叫哈夫曼树。在权为w1，w2…..的n个**叶子**所构成的二叉树带权路径长度最小的二叉树

1. n个节点，每个都是孤立的节点。经过n-1次合并产生n-1个节点共2n-1个



遍历：

# Part 4 Search

## 线性表

### 顺序查找

二分法查找（Binary Search）折半查找：

1. //R[1……..n]
2. **int** BinarySearch(SeqList R,Key K){
3. **int** low=1,high=n,mid;
4. **while**(low<=high){
5. mid=(low+high)/2;
6. **if**(R[mid].key==K)
7. **return** mid;
8. **if**(R[mid].key>K)
9. high=mid-1;
10. **else**
11. low=mid+1;
12. }
13. **return** 0;
14. }

### 分块查找

每块不一定有序，但分块有序。将[1..n]分为n块，前b-1快都有[n/b];抽取每块最大关键字组成ID[1..b];由于块内无序，只能用顺序查找。

## Tree

非线性结构。

### Binary Sort Tree

中序遍历得到有序递增序列

### B-Tree

一棵度为m的B-树称为m阶B-树。有的B-树所以辅助信息都存在叶节点，内部节点都存关键字和指向孩子的指针，使内部节点的度数尽可能的大

定义：

一个m阶B-树数据域形式

(j,Po,K1,P1,K2….Kj,Pj) keys(Po)<K1<keys(P1)<K2<…….Kj<keys(Pj)

1.叶节点位于同一层，即为树的高度

2.非叶子关键字[m/2]-1<=j<=m-1 子树[m/2]<=zishu<=m。根至多m-1关键字，m棵子树

3.1001的B-树结构

#define Max 1000 //关键字max=m-1

#define Min 2000 //[m/2]-1

Typedef int KeyType

Typedef struct node{

int keynum;

KeyType key[Max+1]; //key[0]不用

struct node \* parent;

struct node \* son[Max+1];

}BtreeNode;

typedef BtreeNode \* BTree;

4.查找：

5.插入和生成：

**散列表的概念**

# Part 5 Sort

## 排序

排序后记录次序不变则方法**稳定**的

排序过程都放在内存不涉及内外存交换，则称**内排序。**适应记录少的小文件，下面的都是

排序算法所需辅助空间不依赖于问题的规模n，辅助空间是O（1），称**就地排序**。大多数算法的开销是记录的比较和记录的移动。

**各种内部排序方法的选择和比较**

## 选择排序

**直接选择排序：**

每次查找从所有无序区中找到最小的。第i趟排序需n-i次比较，总比较次数：

n(n-1)/2=O(n`2)

**堆排序**

满足条件：Ki<=K2i且Ki<=K2i+1(小根堆)，讨论二叉堆Binary Heap

void heapSort(SeqList R)

{

int I;

BuildHeap(R);

for(i=n;i>1;i--){

R[0]=R[1];R[1]=R[i];R[i]=R[0]; //R[1]最大。讨论大根堆

Heapify(R,1,i-1);

}

}

void BuidHeap(SeqList R) //heapify是针对一个的

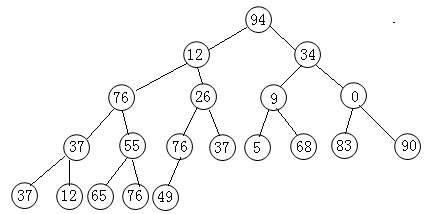
{

int i;

for(i=n/2;i>0;i--) // n/2求它父母节点

heapify(R,i,n); .

}



## 插入排序

**插入排序**

**直接插入排序：**

void insertSort(int \* a,int n)

{

int i,j;

for(i=2;i<n;i++) //6 3 5

{

a[0]=a[i]; //搬走后留下空位置

j=i-1;

while(a[0]<a[j])

{

a[j+1]=a[j]; //搬走后留下空位置

j--;

}

a[j+1]=a[0];

}

}

课本：

void insertSort(SeqList R)

{

int i,j;

for(i=2;i<=n;i++)

if(R[i].key<R[i-1].key){

R[0]=R[i];j=i-1;

do{

R[j+1]=R[j];

j--;

}while(R[0].key<R[j].key);

R[j+1]=R[0];

}

}

**希尔排序：**

分组的的插入排序

void shellSort(Seqlist R)

{

int increment=1;

do{

increment=increment/3+1;

shellPass(R,increment);

}while(increment>1)

}

void shellPass(Seqlist R,int d)

{

int i,j;

for(i=d+1;i<=n;i++)

{

a[0]=a[i]; //搬走后留下空位置

j=i-d;

while(a[0]<a[j]&&j>0)

{

a[j+d]=a[j]; //搬走后留下空位置

j=j-d;

}

a[j+d]=a[0];

}

}

## 交换排序

**冒泡排序：**

void bubbleSort(SeqList R)

{

int i,j;

Boolean exchange;

for(i=1;i<n;i++) //最多做n-1趟排序

{

exchange=false;

for(j=n-1;j>=i;j--)

{

}

}

}

最好情况：比较次数：n-1； 移动：0

**快速排序：**

## 归并排序

最简单的方式两个有序子文件放在R[low…m],R[m+1…high]合并到R1，复制回R[low,high].如

Void merge(SeqList R,int low,int m,int high)

{

Int i=low,j=m+1;p=0;

RecType \* R1;

while(i<=m&&j<=high>

R1[p++]=(R[i].key<=R[j].key)?R[i++]:R[j++]//R指向R1。

}

注意：申请的空间较大，加判断。

归并排序的两种实现：

自底向上和自顶向下。即为**二路归并排序**。把R[1..n]看成长度是1的顺序子文件顺序两两相邻归并。在某趟归并，子文件长度length(最后一个字文件可能小于)，归并前有[n/length]个子文件，R[1..length],R[length+1…..2length],…(R[n/length]-1)\*length+1…n]

## 分配排序

**箱排序（Bin Sort）也称桶排序（Bucket Sort）：**

箱个数取决关键字取值范围，R[0…n-1],分配时间O（n），收集时间O（m）或O（m+n）

1箱子容量不可预知，2保持稳定故箱子采用**链队列**

注意：适用于关键字取值范围较小。

箱排序的变种：桶排序，假设输入n个关键字是分布在[0,1）上，找某个合适的数映射到此区间。区间是一个桶。指针B[0…n-1]表示n个桶，记录0<=R[i]<=n-1通过n\*(R[i].key)

**基数排序：**