

# 十月百度，阿里巴巴，迅雷搜狗最新面试七十题（第 201-270 题）

## 引言

当即早已进入 10 月份，十一过后，招聘，笔试，面试，求职渐趋火热。而在这一系列过程背后浮出的各大 IT 公司的笔试/面试题则蕴含着诸多思想与设计，细细把玩，思考一番亦能有不少收获。

上个月，本博客着重整理[九月腾讯](#)，[创新工场](#)，[淘宝](#)等公司最新面试十三题，此次重点整理百度，阿里巴巴，迅雷和搜索等公司最新的面试题。同上篇一样，答案望诸君共同讨论之，个人亦在慢慢思考解答。多谢。

## 最新面试十一题

1. 十月百度：一个数组保存了  $N$  个结构，每个结构保存了一个坐标，结构间的坐标都不相同，请问如何找到指定坐标的结构（除了遍历整个数组，是否有更好的办法）？（要么预先排序，二分查找。要么哈希。hash 的话，坐标  $(x, y)$  你可以当做一个 2 位数，写一个哈希函数，把  $(x, y)$  直接转成 “ $(x, y)$ ” 作为 key，默认用 string 比较。或如 Edward Lee 所说，将坐标  $(x, y)$  作为 Hash 中的 key。例如  $(m, n)$ ，通过  $(m, n)$  和  $(n, m)$  两次查找看是否在 HashMap 中。也可以在保存时就规定  $(x, y)$ ， $x < y$ ，在插入之前做个判断。）
2. 百度最新面试题：现在有 1 千万个随机数，随机数的范围在 1 到 1 亿之间。现在要求写出一种算法，将 1 到 1 亿之间没有在随机数中的数求出来。（编程珠玑上有此类似的一题，如果有足够的内存的话可以用位图法，即开一个 1 亿位的 bitset，内存为  $100m/8 = 12.5m$ ，然后如果一个数有出现，对应的 bitset 上标记为 1，最后统计 bitset 上为 0 的即可。）
3. Alibaba 笔试题：给定一段产品的英文描述，包含  $M$  个英文字母，每个英文单词以空格分隔，无其他标点符号；再给定  $N$  个英文单词关键字，请说明思路并编程实现方法

String extractSummary(String description, String[] key words)

目标是找出此产品描述中包含  $N$  个关键字（每个关键词至少出现一次）的长度最短的子串，作为产品简介输出。（不限编程语言）20 分。（扫描过程始终保持一个  $[left, right]$  的 range，初始化确保  $[left, right]$  的 range 里包含所有关键字则停止。然后每次迭代：

- 1，试图右移动 left，停止条件为再移动将导致无法包含所有关键字。
- 2，比较当前 range's length 和 best length，更新最优值。
- 3，右移 right，停止条件为使任意一个关键字的计数+1。
- 4，重复迭代。

编程之美有[最短摘要生成](#)的问题，与此问题类似，读者可作参考。）

4. 搜狗：有  $N$  个正实数(注意是实数，大小升序排列)  $x_1, x_2 \dots x_N$ ，另有一个实数  $M$ 。需要选出若干个  $x$ ，使这几个  $x$  的和与  $M$  最接近。请描述实现算法，并指出算法复杂度（参考：[第五章、寻找满足条件的两个或多个数](#)）。
5. 迅雷：给你 10 台机器，每个机器 2 个 cpu，2g 内存，现在已知在 10 亿条记录的数据库里执行一次查询需要 5 秒，问用什么方法能让 90% 的查询能在 100 毫秒以内返回结果。（@geochway：将 10 亿条记录排序，然后分到 10 个机器中，分的时候是一个记录一个记录的轮流分，确保每个机器记录大小分布差不多，每一次查询时，同时提交给 10 台机器，同时查询，

因为记录已排序, 可以采用二分法查询。

如果无法排序, 只能顺序查询, 那就要看记录本身的概率分布, 否则不可能实现。

一个机器 2 个 CPU 未必能起到作用, 要看这两个 CPU 能否并行存取内存, 取决于系统架构。)

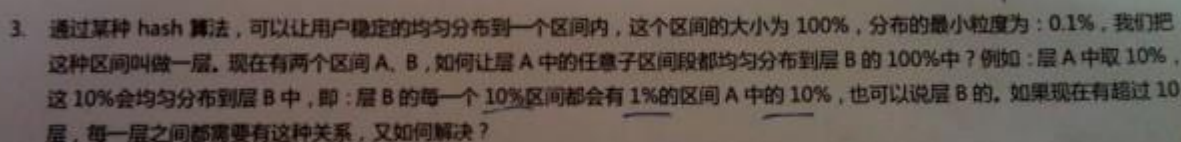
6. 给定一个函数 rand() 能产生 0 到 n-1 之间的等概率随机数, 问如何产生 0 到 m-1 之间等概率的随机数?
7. 腾讯: 五笔的编码范围是 a ~ y 的 25 个字母, 从 1 位到 4 位的编码, 如果我们把五笔的编码按字典序排序, 形成一个数组如下:

a, aa, aaa, aaaa, aaab, aaac, ... ..., b, ba, baa, baaa, baab, baac ... ..., yyyw, yyyx, yyyy

其中 a 的 Index 为 0, aa 的 Index 为 1, aaa 的 Index 为 2, 以此类推。

- 1) 编写一个函数, 输入是任意一个编码, 比如 baca, 输出这个编码对应的 Index;
  - 2) 编写一个函数, 输入是任意一个 Index, 比如 12345, 输出这个 Index 对应的编码。
8. **2011.10.09 百度笔试题** (下述第 8-12 题): linux/unix 远程登陆都用到了 ssh 服务, 当网络出现错误时服务会中断, linux/unix 端的程序会停止。为什么会这样? 说下 ssh 的原理, 解释中断的原理。
  9. 一个最小堆, 也是完全二叉树, 用按层遍历数组表示。
    1. 求节点 a[n] 的子节点的访问方式
    2. 插入一节点的程序 void add\_element(int \*a, int size, int val);
    3. 删除最小节点的程序。
  10. a) 求一个全排列函数: 如 p([1,2,3]), 输出: [123],[132],[213],[231],[321],[323]。  
b) 求一个组合函数: 如 p([1,2,3]), 输出: [1],[2],[3],[1,2],[2,3],[1,3],[1,2,3]。

这两问可以用伪代码 (全排列请参考这里的第 67 题: [微软、Google 等公司非常好的面试题及解答\[第 61-70 题\]](#))。



3. 通过某种 hash 算法, 可以让用户稳定的均匀分布到一个区间内, 这个区间的大小为 100%, 分布的最小粒度为: 0.1%。我们把这种区间叫做一层。现在有两个区间 A、B, 如何让层 A 中的任意子区间都均匀分布到层 B 的 100% 中? 例如: 层 A 中取 10%, 这 10% 会均匀分布到层 B 中, 即: 层 B 的每一个 10% 区间都会有 1% 的区间 A 中的 10%, 也可以说层 B 的。如果现在有超过 10 层, 每一层之间都需要有这种关系, 又如何解决?

- 11.
12. 有这样一种编码: 如,  $N=134$ ,  $M=f(N)=143$ ,  $N=020$ ,  $M=fun(N)=101$ , 其中 N 和 M 的位数一样, N, M 可以均以 0 开头, N, M 的各位数之和要相等, 即  $1+3+4=1+4+3$ , 且 M 是大于 N 中最小的一个, 现在求这样的序列 S, N 为一个定值, 其中  $S(0)=N$ ,  $S(1)=fun(N)$ ,  $S(2)=fun(S(1))$ 。
13. 有 1000 万条 URL, 每条 URL 50 字节, 只包含主机前缀, 要求实现 URL 提示系统:
  - (1) 要求实时更新匹配用户输入的地址, 每输出一个字符, 输出最新匹配 URL
  - (2) 每次只匹配主机前缀, 例如对 [www.abaidu.com](#) 和 [www.baidu.com](#), 用户输入 [www.b](#) 时只提示 [www.baidu.com](#)
  - (3) 每次提供 10 条匹配的 URL
  - (4) 以用户需求为主。
14. 海量记录, 记录形式如下: TERMID URLNOCOUNT urlno1 urlno2 ..., urlnon  
怎么考虑资源和时间这两个因素, 实现快速查询任意两个记录的交集, 并集等, 设计相关的数据结构和算法。
15. 百度最新笔试题 (感谢 xiongyangwan 提供的题目): 利用互斥量和条件变量设计一个消息队列, 具有以下功能:
  - 1 创建消息队列 (消息中所含的元素)

- 2 消息队列中插入消息
- 3 取出一个消息（阻塞方式）
- 4 取出第一消息（非阻塞方式）

16. 百度移动终端研发笔试：系统设计题（40 分）

对已排好序的数组 A，一般来说可用二分查找可以很快找到。现有一特殊数组 A[]，它是循环递增的，如 A[]={ 17 19 20 25 1 4 7 9}，试在这样的数组中找一元素 x，看看是否存在。

请写出你的算法，必要时可写伪代码，并分析其空间、时间复杂度。

17. #include<stdio.h>

```
#include <string.h>
```

```
void main()
```

```
{
```

```
int a[2000];
```

```
char *p = (char *)a;
```

```
int i;
```

```
for( i = 0; i < 2000; i++)
```

```
    a[i] = -i -1;
```

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

```
}
```

写出输出结果

（onlyice: i = FFFFFFF0H 的时候，才有'\0'出现，就是最后一个字节，C 风格字符串读到'\0'就终止了。FFFFFF0H 是 -256，就是 i 的值为 255 时 a[i] = FFFFFFF0H ）....

18. 腾讯 10.09 测试笔试题：有  $N + 2$  个数， $N$  个数出现了偶数次，2 个数出现了奇数次（这两个数不相等），问用  $O(1)$  的空间复杂度，找出这两个数，不需要知道具体位置，只需要知道这两个值。（@Rojay: xor 一次，得到 2 个奇数次的数之和 x。第二步，以 x（展开成二进制）中有 1 的某位（假设第 i 位为 1）作为划分，第二次只 xor 第 i 位为 1 的那些数，得到 y。然后  $x \text{ xor } y$  以及 y 便是那两个数。）

19. @well: 一个整数数组，有 n 个整数，如何找其中 m 个数的和等于另外 n-m 个数的和？（与上面第 4 题类似，参考：[第五章、寻找满足条件的两个或多个数](#)）。

20. 阿里云笔试题：一个 HTTP 服务器处理一次请求需要 500 毫秒，请问这个服务器如何每秒处理 100 个请求。

## 21. 今天 10.10 阿里云笔试@土豆：1、三次握手；

TCP 连接是通过三次握手进行初始化的。三次握手的目的是同步连接双方的序列号和确认号并交换 TCP 窗口大小信息。以下步骤概述了通常情况下客户端计算机联系服务器计算机的过程：

1. 客户端向服务器发送一个SYN置位的TCP报文，其中包含连接的初始序列号x和一个窗口大小（表示客户端上用来存储从服务器发送来的传入段的缓冲区的大小）。
2. 服务器收到客户端发送过来的SYN报文后，向客户端发送一个SYN和ACK都置位的TCP报文，其中包含它选择的初始序列号y、对客户端的序列号的确认x+1和一个窗口大小（表示服务器上用来存储从客户端发送来的传入段的缓冲区的大小）。
3. 客户端接收到服务器端返回的SYN+ACK报文后，向服务器端返回一个确认号y+1和序号x+1的ACK报文，一个标准的TCP连接完成。

TCP 使用类似的握手过程来结束连接。这可确保两个主机均能完成传输并确保所有的数据均得以接收

TCP Client	Flags	TCP Server
1 Send SYN (seq=x)	----SYN-->	SYN Received
2 SYN/ACK Received	<---SYN/ACK----	Send SYN (seq=y), ACK (x+1)
3 Send ACK (y+1)	----ACK-->	ACK Received, Connection Established
x: ISN (Initial Sequence Number) of the Client		
y: ISN of the Server		

第一次是客户端发起连接；第二次表示服务器收到了客户端的请求；第三次表示客户端收到了服务器的反馈。这之后双方均确认了连接的有效性，如果第三次服务器未收到，假设一个C向S发送了SYN后无故消失了，那么S在发出SYN+ACK应答报文后是无法收到C的ACK报文的（第三次握手无法完成），这种情况下S一般会重试（再次发送SYN+ACK给客户端）并等待一段时间后丢弃这个未完成的连接，这段时间的长度我们称为SYN Timeout，一般来说这个时间是分钟的数量级（大约为30秒-2分钟）；

2、

死锁的条件。（**互斥条件**（Mutual exclusion）：1、资源不能被共享，只能由一个进程使用。2、请求与保持条件（Hold and wait）：已经得到资源的进程可以再次申请新的资源。3、非剥夺条件（No pre-emption）：已经分配的资源不能从相应的进程中被强制地剥夺。4、循环等待条件（Circular wait）：系统中若干进程组成环路，该环路中每个进程都在等待相邻进程正占用的资源。**处理死锁的策略**：1. 忽略该问题。例如鸵鸟算法，该算法可以应用在极少发生死锁的情况下。为什么叫鸵鸟算法呢，因为传说中鸵鸟看到危险就把头埋在地底下，可能鸵鸟觉得看不到危险也就没危险了吧。跟掩耳盗铃有点像。2. 检测死锁并且恢复。3. 仔细地对资源进行动态分配，以避免死锁。4. 通过破除死锁四个必要条件之一，来防止死锁产生。）

## 22. 微软 2011 最新面试题（以下三题，第 22、23、24 题皆摘自微软亚洲研究院的邹欣老师博客）：浏览过本人的[程序员编程艺术系列](#)的文章，一定对其中的这个问题颇有印象：[第七章、求连续子数组的最大和](#)。

-32, -10, 33, -23, 32, -12, 41, -12, 12

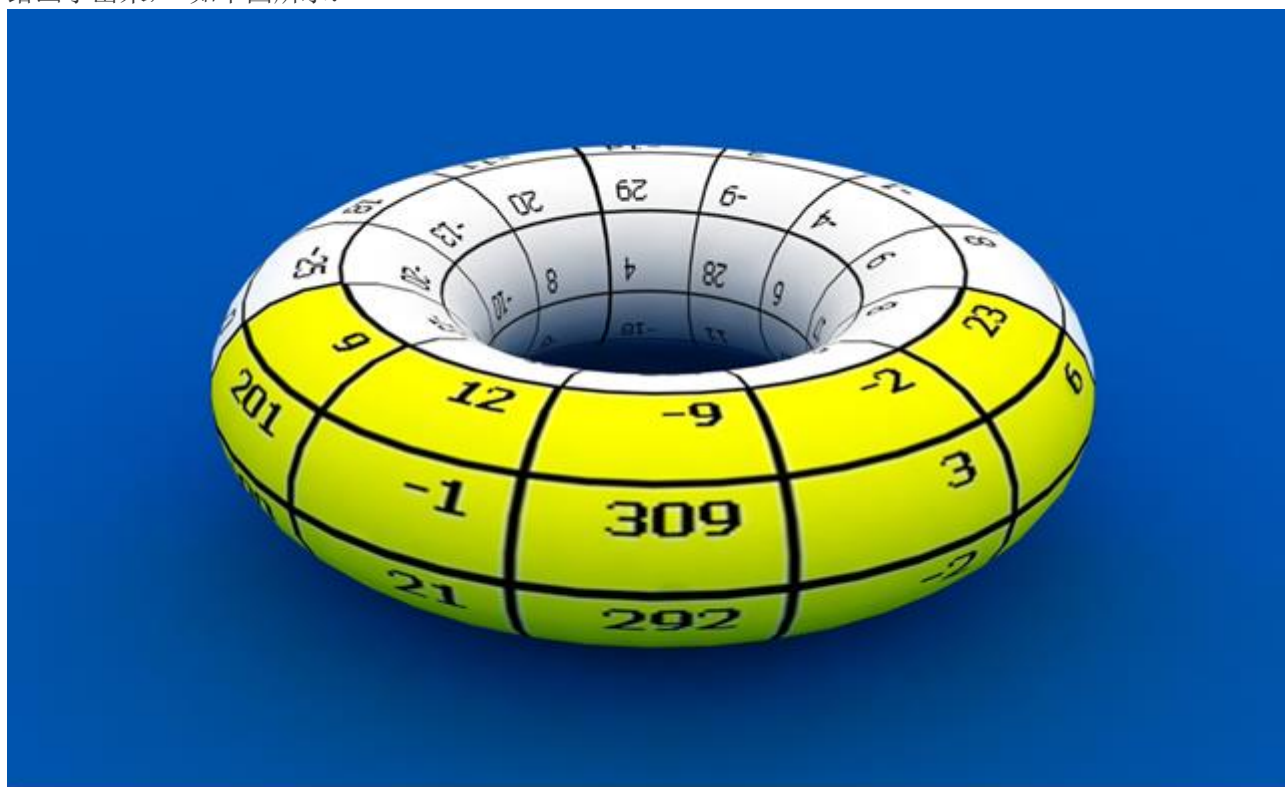
求数组最大子数组的和最初来源于编程之美，[我在编程艺术系列中提供了多种解答方式](#)，然而这个问题若扩展到二维数组呢？



8	-10	-3	26	-11	-1	-6	12	17	6	28	4
20	-13	-20	-13	-15	-254	5	8	9	-4	-9	29
-11	18	-25	9	12	-9	-2	23	8	-1	3	-14
-16	-7	0	201	-1	309	3	6	-18	11	24	-8
-1	-7	11	100	21	292	-2	2	-18	-8	-10	9
26	-11	-19	-18	20	-981	2	-14	12	-14	1	27
9	-20	5	28	-15	26	-20	-8	-16	30	3	20
-6	-7	-5	-9	-16	-15	5	-16	22	-17	11	-18

再者，若数

组首尾相连，像一个轮胎一样，又怎么办呢？聪明的同学还是给出了漂亮的答案，并且用 SilverLight/WPF 给画了出来，如下图所示：

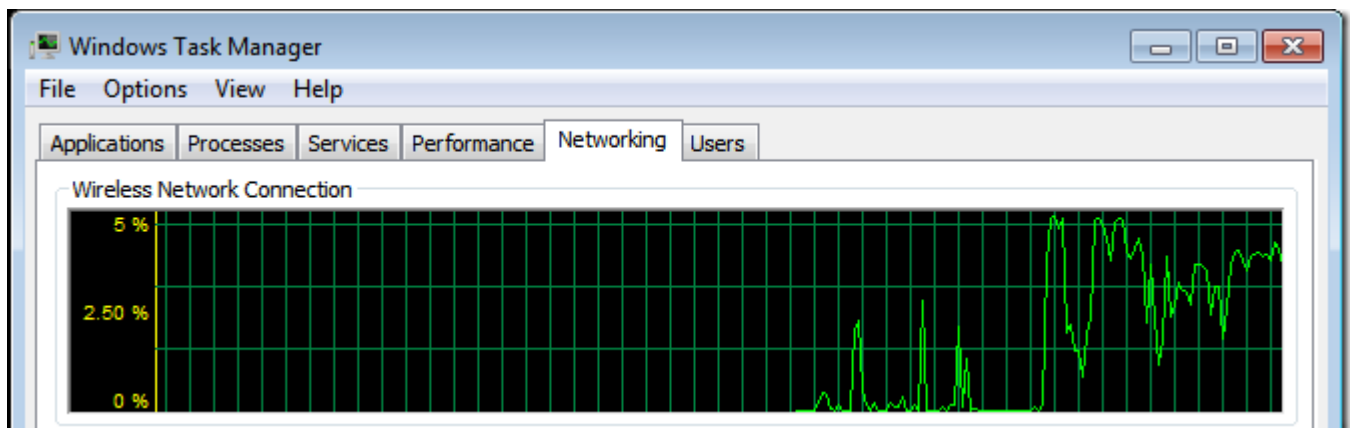


好，设想现在我们有一张纸带，两面都写满了像如上第一幅图那样的数字，我们把纸带的一端扭转，和另一端接起来，构成一个莫比乌斯环（Möbius Strip, 如将一个长方形纸条 ABCD 的一端 AB 固定，另一端 DC 扭转半周后，把 AB 和 CD 粘合在一起，得到的曲面就是麦比乌斯圈，也称莫比乌斯带。），如下图所示：

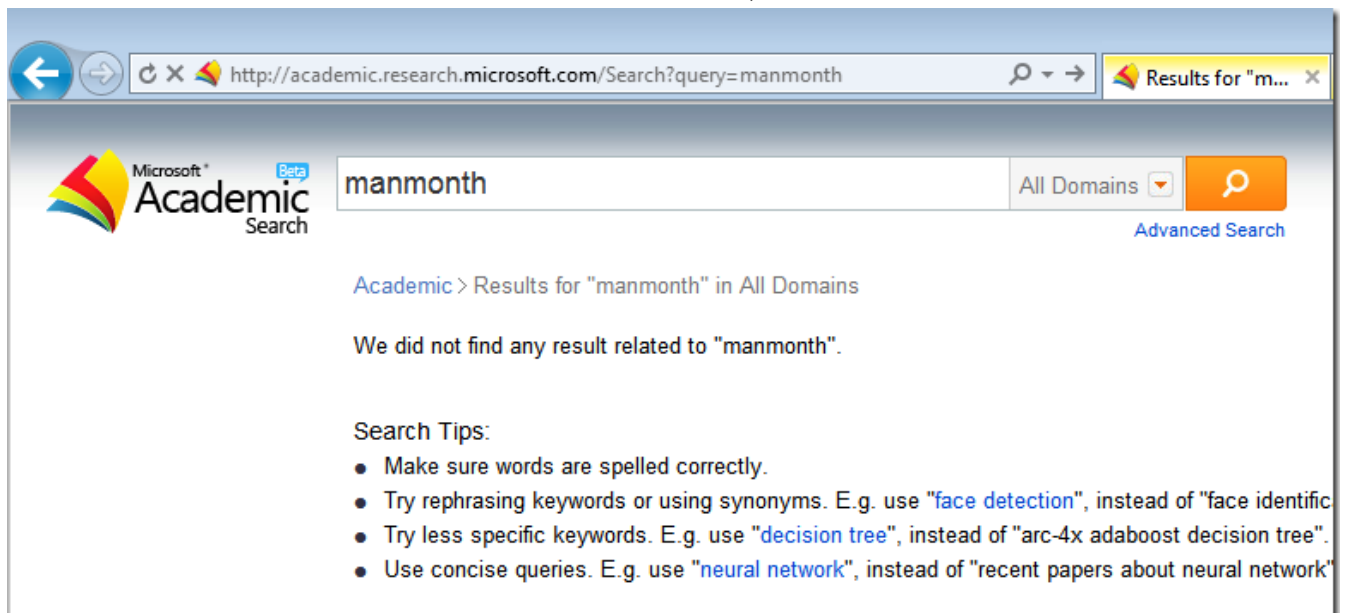


如上，尽管这个纸带扭了一下，但是上面还是有数组，还是有最大子数组的和，对么？在求最大子数组的和之前，我们用什么样的数据结构来表示这些数字呢？你可以用 Java, C, C#，或其他语言的数据结构来描述这个莫比乌斯环上的数组。数据结构搞好了，算法自然就有了。（@风大哥：莫比乌斯带，用环形数组或者链表可以表示。环型数组的话，1-N，到 N 特殊处理一下，连到 1 就是环型数组了，一个纸带上正反两面各有 N 个数， $A_1 \dots A_n, B_1 \dots B_n$ ，那么就可以构造一个新的数组： $A_1-A_n-B_1-B_n$ 。访问到  $B_n$  下一位就是  $A_1$ ，就是环形的数组了。从某个位置 k 开始，用 i, j 向一个方向遍历，直到 i 到达 k 位置，或者  $i=j$ ，被追上，用数组需要一点技巧，就是 j 再次过 k 需要打个标志，以便计算终止条件和输出。当然，如果用链表就更简单了。把链表首尾相接即可，即  $A_n$  执行  $B_1, B_n$  指向  $A_1$  即可。）

23. 《编程之美》的第一题是让 Windows 任务管理器的 CPU 使用率曲线画出一个正弦波。我一直在想，能不能把 CPU 使用率边上的网络使用率也如法炮制一下呢？比如，也来一个正弦曲线？



24. 如果你没看过，也至少听说<人月神话> (The Mythical Man-month) 这本在软件工程领域很有影响的书。当你在微软学术搜索中输入“manmonth”这个词的时候，你会意外地碰到下面这个错误：



经过几次试验之后，你发现必须要输入“man-month”才能得到希望的结果。这不就是只差一个‘-’符号么？为什么这个搜索引擎不能做得聪明一些，给一些提示 (Query Suggestion)? 或者自动把用户想搜的结果展现出来 (Query Alteration)? 我们在输入比较长的英文单词的时候，也难免会敲错一两个字母，网站应该帮助用户，而不是冷冰冰地拒绝用户啊。

微软的学术搜索 (Microsoft Academic Search) 索引了超过 3 千万的文献, 2 千万的人名, 怎么能以比较小的代价, 对经常出现的输入错误提供提示? 或直接显示相关结果, 避免用户反复尝试输入的烦恼?

你可能会说, 这很难吧, 但是另一家搜索引擎似乎轻易地解决了这个问题 (谷歌, 读者可以一试)。所以, 还是有办法的。

这个题目要求你:

- 1) 试验不同的输入, 反推出目前微软的学术搜索是如何实现搜索建议 (Query Suggestion) 的。
- 2) 提出自己的改进建议, 并论证这个解决方案在千万级数据规模上能达到“足够好”的时间 (speed) 和空间 (memory usage) 效率。
- 3) 估计这事需要几个 人·月 (man-month) 才能做完? (备注: 顺便给邹欣老师传个话, 如果应届毕业生可以能做好上述全部三个题目, 便可直接找他。

<http://www.cnblogs.com/xinz/archive/2011/10/10/2205232.html>)。

25. 今天 10.10 阿里云部分笔试题目:

- 1、一个树被序列化为数组, 如何反序列化。
- 2、如何将 100 百万有序数据最快插入到 STL 的 map 里。
- 3、有两个线程 a、b 分别往一条队列 push 和 pop 数据, 在没有锁和信号量的情况下如何避免冲突访问。
- 4、写一个函数, 功能是从字符串 s 中查找出子串 t, 并将 t 从 s 中删除。

26. 将长度为 m 和 n 的两个升序数组复制到长度为 m+n 的数组里, 升序排列。

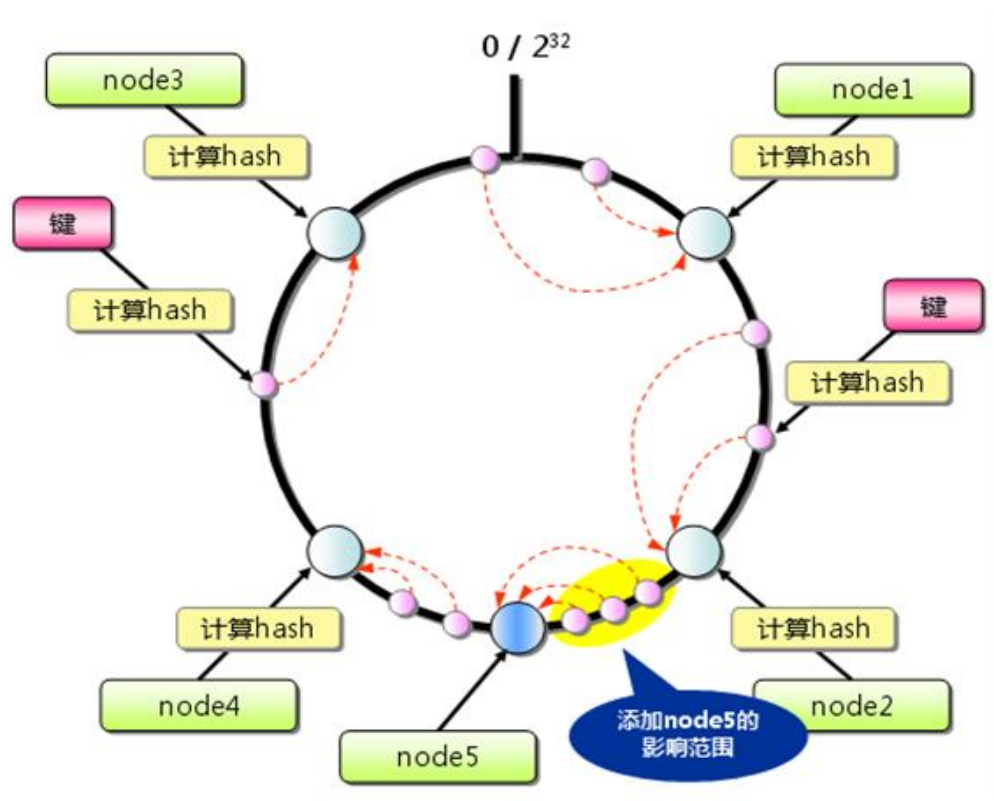
27. tencent2012 笔试题附加题

问题描述: 例如手机朋友网有 n 个服务器, 为了方便用户的访问会在服务器上缓存数据, 因此用户每次访问的时候最好能保持同一台服务器。

已有的做法是根据 `ServerIPIndex[QQNUM%n]` 得到请求的服务器, 这种方法很方便将用户分到不同的服务器上去。但是如果一台服务器死掉了, 那么 n 就变为了 n-1, 那么 `ServerIPIndex[QQNUM%n]` 与 `ServerIPIndex[QQNUM%(n-1)]` 基本上都不一样了, 所以大多数用户的请求都会转到其他服务器, 这样会发生大量访问错误。

问: 如何改进或者换一种方法, 使得:

- (1) 一台服务器死掉后, 不会造成大面积的访问错误,
- (2) 原有的访问基本还是停留在同一台服务器上;
- (3) 尽量考虑负载均衡。(思路: 往分布式一致哈希算法方面考虑。关于此算法, 可参见此文:



28. 腾讯面试题: A.txt 和 B.txt 两个文件, A.txt 有 1 亿个 QQ 号, B.txt 100W 个 QQ 号, 用代码实现交、并、差。
29. 说出下面的运行结果

```
#include <iostream>
using namespace std;
```

```
class A
{
public:
    virtual void Fun(int number = 10)
    {
        std::cout << "A::Fun with number " << number<<endl;
    }
};
```

```
class B: public A
{
public:
    virtual void Fun(int number = 20)
    {
        std::cout << "B::Fun with number " << number<<endl;
    }
};
```



```

    }
};

int main()
{
    B b;
    A &a = b;
    a.Fun();
    return 0;
} //虚函数动态绑定=>B, 非 A, 缺省实参是编译时候确定的=>10, 非 20 。

```

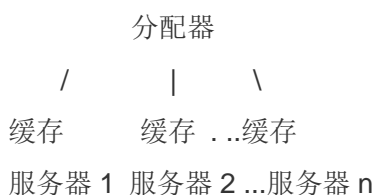
30. 今晚阿里云笔试：有 101 根电线 每根的一头在楼底 另一端在楼顶 有一个灯泡 一个电池 无数根很短的电线 怎么样在楼上一次在楼下去一次将电线的对应关系弄清楚。

31. 金山笔试题：

- 1、C ++为什么经常将析构函数声明为虚函数？
- 2、inline 和#define 的如何定义 MAX，区别是什么。
- 3、const 的用法，如何解除 const 限制。
- 4、智能指针的作用和设计原理。
- 5、STL 中 vector 如何自己设计，关键设计点，函数声明，自定义删除重复元素的函数。
- 6、如何用一条 SQL 语句，删除表中某字段重复的记录。

32. 淘宝：

在现代 web 服务系统的设计中，为了减轻源站的压力，通常采用分布式缓存技术，其原理如下图所示，前端的分配器将针对不同内容的用户请求分配给不同的缓存服务器向用户提供服务。



1) 请问如何设置分配策略，可以保证充分利用每个缓存服务器的存储空间（每个内容只在一个缓存服务器有副本）

2) 当部分缓存服务器故障，或是因为系统扩容，导致缓存服务器的数量动态减少或增加时，你的分配策略是否可以保证较小的缓存文件重分配的开销，如果不能，如何改进？

3) 当各个缓存服务器的存储空间存在差异时（如有 4 个缓存服务器，存储空间比为 4: 9: 15: 7），如何改进你的策略，按照如上的比例将内容调度到缓存服务器？（思路：往 memcached 或者一致性 hash 算法方面考虑，但具体情况，具体分析。）

33. 腾讯：50 个台阶，一次可一阶或两阶，共有几种走法（老掉牙的题了，详见微软面试 100 题 2010 版。

```

long long Fibonacci_Solution1(unsigned int n)
{

```

```

int result[2] = {0, 1};
if(n < 2)
    return result[n];

return Fibonacci_Solution1(n - 1) + Fibonacci_Solution1(n - 2);
})。

```

34. 有两个 float 型的数，一个为 fmax,另一个为 fmin,还有一个整数 n,如果 (fmax - fmin)/n ,不能整除，怎么改变 fmax,fmin,使改变后可以整除 n 。

35. 2011.10.11 最新百度电面：

1、动态链接库与静态链接库的区别（**静态链接库**是 .lib 格式的文件，一般在工程的设置界面加入工程中，程序编译时会把 lib 文件的代码加入你的程序中因此会增加代码大小，你的程序一运行 lib 代码强制被装入你程序的运行空间，不能手动移除 lib 代码。

**动态链接库**是程序运行时动态装入内存的模块，格式\*.dll，在程序运行时可以随意加载和移除，节省内存空间。

在大型的软件项目中一般要实现很多功能，如果把所有单独的功能写成一个一个 lib 文件的话，程序运行的时候要占用很大的内存空间，导致运行缓慢；但是如果将功能写成 dll 文件，就可以在用到该功能的时候调用功能对应的 dll 文件，不用这个功能时将 dll 文件移除内存，这样可以**节省内存空间**。）

2、指针与引用的区别（**相同点**：1. 都是地址的概念；  
指针指向一块内存，它的内容是所指内存的地址；引用是某块内存的别名。

**区别：**

1. 指针是一个实体，而引用仅是个别名；
2. 引用使用时无需解引用(\*)，指针需要解引用；
3. 引用只能在定义时被初始化一次，之后不可变；指针可变；
4. 引用没有 const，指针有 const；
5. 引用不能为空，指针可以为空；
6. “sizeof 引用”得到的是所指向的变量(对象)的大小，而“sizeof 指针”得到的是指针本身(所指向的变量或对象的地址)的大小；
7. 指针和引用的自增(++)运算意义不一样；
8. 从内存分配上看：程序为指针变量分配内存区域，而引用不需要分配内存区域。）

3、进程与线程的区别（①**从概念上**：

进程：一个程序对一个数据集的动态执行过程，是分配资源的基本单位。

线程：一个进程内的基本调度单位。

线程的划分尺度小于进程，一个进程包含一个或者更多的线程。

②**从执行过程中来看**：

进程：拥有独立的内存单元，而多个线程共享内存，从而提高了应用程序的运行效率。

线程：每一个独立的线程，都有一个程序运行的入口、顺序执行序列、和程序的出口。但是线程不能够独立的执行，必须依存在应用程序中，由应用程序提供多个线程执行控制。

③**从逻辑角度来看**：（重要区别）

多线程的意义在于一个应用程序中，有多个执行部分可以同时执行。但是，操作系统并没有将多个线程看做多个独立的应用，来实现进程的调度和管理及资源分配。）

4、函数调用入栈出栈的过程

4、c++对象模型与虚表

5、海量数据处理，以及如何解决 Hash 冲突等问题

6、系统设计，概率算法

36. 今天腾讯面试：

一个大小为  $N$  的数组，里面是  $N$  个整数，怎样去除重复，

要求时间复杂度为  $O(n)$ ，空间复杂度为  $O(1)$ （此题答案请见@作者 hawksoft：

<http://blog.csdn.net/hawksoft/article/details/6867493>）。

37. 一个长度为 10000 的字符串，写一个算法，找出最长的重复子串，如 abczzacba,结果是 bc（思路：后缀树/数组的典型应用，@well：就是求后缀数组的 height[] 的最大值）。

38. 今晚 10.11 大华笔试题：建立一个 data structure 表示没有括号的表达式，而且找出所有等价（equivalent）的表达式

比如：

$3 \times 5 == 5 \times 3$

$2 + 3 == 3 + 2$

39. 今晚 10.11 百度二面：判断一个数的所有因数的个数是偶数还是奇数（只需要你判断因数的个数是偶数个还是奇数个，那么可以这么做@滨湖&&土豆：那只在计算质因数的过程中统计一下当前质因数出现的次数，如果出现奇数次则结果为偶，然后可以立即返回；如果每个质因数的次数都是偶数，那么结果为奇。如果该数是平方数 结果就为奇 否则就为偶了）。

40. 比如 A 认识 B，B 认识 C，但是 A 不认识 C，那么称 C 是 A 的二度好友。找出某个人的所有十度好友。数据量为 10 万（BFS，同时记录已遍历过的顶点，遍历时遇到的已遍历过的顶点不插入队列。此是今晚 10.11 人人笔试题目，但它在上个月便早已出现在本人博客中，即此文第 23 题第 2 小题：[九月腾讯，创新工场，淘宝等公司最新面试十三题](#)）。

41. map 在什么情况下会发生死锁；stl 中的 map 是怎么实现的？（有要参加淘宝面试的朋友注意，淘宝喜欢问 STL 方面的问题）

42. 昨日笔试：有四个人，他们每次一起出去玩的时候，用同时剪刀包袱锤的方式决定谁请客。设计一种方法，使得他们只需出一次，就可以决定请客的人，并且每个人请客的几率相同，均为 25%。

43. Given two sets of  $n$  numbers  $a_1, a_2, \dots, a_n$  and  $b_1, b_2, \dots, b_n$ , find, in polynomial time, a permutation  $\Pi$  such that  $\sum_i |a_i - b_{\Pi(i)}|$  is minimized? Prove your algorithm works.

有两个数组，在多项式时间里找到使 两数组元素 的差 的绝对值 的和 最小 的一种置换。

并证明算法的有效性。注意，关键是证明。（此题个人去年整理过类似的一题，详见微软面试 100 题 2010 版第 32 题：[http://blog.csdn.net/v\\_JULY\\_v/archive/2011/01/10/6126444.aspx](http://blog.csdn.net/v_JULY_v/archive/2011/01/10/6126444.aspx)）

44. 对已排好序的数组 A，一般来说可用二分查找 可以很快找到。

现有一特殊数组 A[]，它是循环递增的，如  $A[] = \{ 17 \ 19 \ 20 \ 25 \ 1 \ 4 \ 7 \ 9 \}$ ,

试在这样的数组中找一元素  $x$ ，看看是否存在。

请写出你的算法，必要时可写伪代码，并分析其空间 时间复杂度。

45. 网易：题意很简单，写一个程序，打印出以下的序列。

(a),(b),(c),(d),(e).....(z)

(a,b),(a,c),(a,d),(a,e).....(a,z),(b,c),(b,d).....(b,z),(c,d).....(y,z)

(a,b,c),(a,b,d)....(a,b,z),(a,c,d)....(x,y,z)

....

(a,b,c,d,.....x,y,z)（思路：全排列问题）

46. `int global = 0;`

```
// thread 1
```

```
for(int i = 0; i < 10; ++i)
```

```
    global -= 1;
```

```
// thread 2
```

```
for(int i = 0; i < 10; ++i)
```

```
    global += 1;
```

之后 `global` 的可能的值是多少（多种可能）？

47. 今天 10.13 新浪笔试：

1、用隐喻说明 `class` 和 `object` 的区别，要求有新意。

2、DDL，DML，DCL 的含义，和距离

3、TCP 建立连接的三次握手

4、设计人民币面值，要求种类最好，表示 1——1000 的所有数，平均纸币张数最少

5、UML

48. 一个数组。里面的数据两两相同，只有两个数据不同，要求找出这两个数据。要求时间复杂度  $O(N)$  空间复杂度  $O(1)$ 。

49. 两个数相乘，小数点后位数没有限制，请写一个高精度算法。

50. 面试基础题：

1、静态方法里面为什么不能声明静态变量？

2、如果让你设计一个类，什么时候把变量声明为静态类型？

3、抽象类和接口的具体区别是什么？

51. 谷歌昨晚 10.13 算法笔试三题：

1. 一个环形公路，上面有  $N$  个站点， $A_1, \dots, A_N$ ，其中  $A_i$  和  $A_{i+1}$  之间的距离为  $D_i$ ， $A_N$  和  $A_1$  之间的距离为  $D_0$ 。

高效的求第  $i$  和第  $j$  个站点之间的距离，空间复杂度不超过  $O(N)$

它给出了部分代码如下：

```
#define N 25
```

```
double D[N]
```

```
....
void Preprocess()
{
    //Write your code1;
}
double Distance(int i, int j)
{
    //Write your code2;
}
```

2. 一个字符串，压缩其中的连续空格为 1 个后，对其中的每个字串逆序打印出来。比如"abc efg hij"打印为"cba gfe jih"。

3. 将一个较大的钱，不超过 1000000(10^6)的人民币，兑换成数量不限的 100、50、10、5、2、1 的组合，请问共有多少种组合呢？（其它选择题考的是有关：操作系统、树、概率题、最大生成树有关的题，另外听老梦说，谷歌不给人霸笔的机会。）。

#### 52. 谷歌在线笔试题：

输入两个整数 A 和 B，输出所有 A 和 B 之间满足指定条件的数的个数。指定条件：假设 C=8675 在 A 跟 B 之间，若  $(8+6+7+5) / 4 > 7$ ，则计一个，否则不计。

要求时间复杂度： $\log(A)+\log(B)$ 。

已知二叉树的前序遍历为：- + a \* b - c d / e f  
后序遍历为：a b c d - \* + e f / -  
求其中序遍历？

#### 53.

#### 54. 十五道百度、腾讯面试基础测试题 @fengchaokobe:

- 1、写一个 C 的函数，输入整数 N，输出整数 M，M 满足：M 是 2 的 n 次方，且是不大于 N 中最大的 2 的 n 次方。例如，输入 4,5,6,7，都是输出 4。
- 2、C++ 中虚拟函数的实现机制。
- 3、写出选择排序的代码及快速排序的算法。
- 4、你认为什么排序算法最好？
- 5、tcp/ip 的那几层协议，IP 是否是可靠的？为什么？
- 6、进程和线程的区别和联系，什么情况下用多线程，什么时候用多进程？
- 7、指针数组和数组指针的区别。
- 8、查找单链表的中间结点。
- 9、最近在实验室课题研究或工作中遇到的技术难点，怎么解决的？
- 10、sizeof 和 strlen 的区别。
- 11、malloc-free 和 new-delete 的区别
- 12、大数据量中找中位数。
- 13、堆和栈的区别。



14、描述函数调用的整个过程。

15、在一个二维平面上有三个不在一条直线上的点。请问能够作出几条与这些点距离相同的线？

55. 搜狐的一道笔试题：

```
char *s="mysohu";
```

```
s[0]=0; //..
```

```
printf("%s",s);
```

输出是什么啊？

搜狐的一道大题：

数组非常长，如何找到第一个只出现一次的数字，说明算法复杂度。（与个人之前整理的微软面试 100 题中，第 17 题：在一个字符串中找到第一个只出现一次的字符。类似，读者可参考。）

56. 百度笔试 3. 假设有一台迷你计算机，1KB 的内存，1MHZ 的 cpu，已知该计算机执行的程序可出现确定性终止(非死循环)，问如何求得这台计算机上程序运行的最长时间，可以做出任何大胆的假设。

57. 微软 10.15 笔试：对于一个数组{1,2,3}它的子数组有{1,2}，{1,3}{2,3}，{1,2,3}，元素之间可以不是连续的，对于数组{5,9,1,7,2,6,3,8,10,4}，升序子序列有多少个？或者换一种表达为：数组 int

```
a[]={5,9,1,7,2,6,3,8,10,4}。求其所有递增子数组(元素相对位置不变)的个数， 例如：{5, 9}，{5, 7, 8, 10}，{1, 2, 6, 8}。
```

58. 今日腾讯南京笔试题：

M\*M 的方格矩阵，其中有一部分为障碍，八个方向均可以走，现假设矩阵上有 Q+1 节点，从(X0, Y0)出发到其他 Q 个节点的最短路径。

其中， $1 \leq M \leq 1000$ ， $1 \leq Q \leq 100$ 。

59. 另外一个笔试题：

一个字符串 S1：全是由不同的字母组成的字符串如：abcdefghijklmn

另一个字符串 S2：类似于 S1，但长度要比 S1 短。

问题是，设计一种算法，求 S2 中的字母是否均在 S1 中。（字符串包含问题，详见程序员编程艺术系列第二章：[http://blog.csdn.net/v\\_JULY\\_v/article/details/6347454](http://blog.csdn.net/v_JULY_v/article/details/6347454)）。

60. 检索一英语全文，顺序输出检测的单词和单词出现次数。

61. 今天 10.15 下午网易游戏笔试题：给一个有序数组 array[n]，和一个数字 m，判断 m 是否是这些数组里面的数的和。（类似于微软面试 100 题 2010 年版第 4 题，即相当于给定一棵树，然后给定一个数，要求把那些相加的和等于这个数的 所有节点打印出来）。

62. 一个淘宝的面试题

文件 A：

```
uid username
```

文件 B：

```
username password
```

文件 A 是按照 uid 有序排列的，要求有序输出合并后的 A,B 文件，格式为 uid username password（A B 两个文件都很大，内存装不下。）

63. 百度可能会问问 memcached（可下载此份文档看看：<http://tech.idv2.com/2008/08/17/memcached-pdf/>。

源码下载地址：<http://www.oschina.net/p/memcached>），apache 之类的。

64. 今上午 10.16 百度笔试：1.C++ STL 里面的 vector 的实现机制，

(1) 当调用 push\_back 成员函数时，怎么实现？（粗略的说@owen，内存足则直接 placement new 构造对象，否则扩充内存，转移对象，新对象 placement new 上去。具体的参见此文：

[http://blog.csdn.net/v\\_july\\_v/article/details/6681522](http://blog.csdn.net/v_july_v/article/details/6681522)）

(2) 当调用 clear 成员函数时，做什么操作，如果要释放内存该怎么做。（调用析构函数，内存不释放。clear 没有释放内存，只是将数组中的元素置为空了，释放内存需要 delete。）

2. 函数 foo 找错，该函数的作用是将一个字符串中的 a-z 的字母的频数找出来

```
void foo(char a[100],int cnt[256])
```

```
{
memset(cnt,0, sizeof(cnt));
while (*a!='\0')
{
++cnt[*a];
++a;
}
for ( char c='a';c<='z';++c)
{
printf("%c:%d\n",c,cnt[c]);
}
}

int main()
{
char a[100]="百度 abc";
int cnt[256];
foo(a,cnt);
```

return 0;

```
linux-6v95:/home/owenliang/csdn/cAndCpp # cat main.cpp
#include <stdio.h>
#include <stdlib.h>
#include <string.h>

void foo(char a[100], int cnt[256])
{
    memset(cnt,0,sizeof(int)*256);
    unsigned char *p=(unsigned char*)a;
    while( (*p)!='\0' )
    {
        ++cnt[*p++];    //保证*p在0-255内
    }
    for(char c='a';c<='z';++c)
    {
        printf("(%c,%d)\n",c,cnt[c]);
    }
}

int main()
{
    char a[]="百度abc";
    int cnt[256];

    foo(a,cnt);

    return 0;
}
```

65. 腾讯长沙笔试：旅行商问题。

66. 今天完美 10.16 笔试题：2D 平面上有一个三角形 ABC，如何从这个三角形内部随机取一个点，且使得在三角形内部任何点被选取的概率相同。

67. 不用任何中间变量，实现 strlen 函数

68. 笔试：联合赋值问题：#include <stdio.h>

```
union A{
int i;
char x[2];
}a;

int main()
{
a.x[0]=10;
a.x[1]=1;
printf("%d\n",a.i);
return 0;
}

sizeof(a) = sizeof(int) = 4 byte
4 * 8 = 32 bit
```

```

a => 00000000 00000000 00000000 00000000
a.x[0]=10 => 00000000 00000000 00000000 00001010
a.x[1]=1 => 00000000 00000000 00000001 00001010
a.i = 1*256 + 1*8 + 1*2 = 256+10 = 266

```

69. 昨天做了中兴的面试题:

```

struct A{
    int a;
    char b;
    char c;
};

```

问 sizeof(A)是多大?

70. 你好:

今天 5 月 6 日百度笔试, 遇到一个题目, 没想到比较好的思路 在网上看了不太明朗, , 希望你帮我解答下

题目如下:

百度研发笔试题。设子数组  $A[0:k]$  和  $A[k+1:N-1]$  已排好序 ( $0 \leq k \leq N-1$ )。试设计一个合并这 2 个子数组为排好序的数组  $A[0:N-1]$  的算法。要求算法在最坏情况下所用的计算时间为  $O(N)$ , 只用到  $O(1)$  的辅助空间。

若论这道题的来源, 则是在高德纳的计算机程序设计艺术第三卷第五章排序中, 如下(第一张图是原题, 第二张图是书上附的答案):

18. [40] (M. A. Kronrod 给定仅含两个路段的  $N$  个记录的一个文件  
 $K_1 \leq \dots \leq K_M$  和  $K_{M+1} \leq \dots \leq K_N$   
 有可能在一个随机存取存储器中用  $O(N)$  个操作对这个文件排序吗? 而且不论  $M$  和  $N$  的大小如何, 只许使用少量固定的附加存储空间(本节描述的所有合并算法都使用与  $N$  成比例的额外存储空间)。

开诸于文件。)

18. 是的, 但它似乎是一项复杂的工件。要找出的头一个解作用下列巧妙的构造。[Doklady Akad. Nauk SSSR 186(1969), 1256~1258]: 设  $n \approx \sqrt{N}$ 。把这个文件分成为  $m+2$  个“段”  $Z_1 \dots Z_m Z_{m+1} Z_{m+2}$ , 其中  $Z_{m+2}$  包含  $(N \bmod n)$  个记录, 而每一个其它的段恰包含  $n$  个记录。把  $Z_{m+1}$  的诸记录同包含  $R_M$  的段进行交换; 现在这个文件有  $Z_1 \dots Z_m A$  的形式, 其中  $Z_1 \dots Z_m$  中的每一个恰巧包含  $n$  个排好序的记录, 而且这里  $A$  是包含  $s$  个记录的一个辅助区域, 其中的  $s$  在范围  $n \leq s < 2n$  中。

找出具有最小前导元素的段, 而且把整个该段同  $Z_1$  作交换; 如果有一个以上的段有最小前导元素, 则选择有最小尾元素的一个段。(这花费  $O(m+n)$  个操作。) 然后找出具有次小前导元素和尾元素的段, 而且把它同  $Z_2$  进行交换, 等等。最后, 通过  $O(m(m+n)) = O(N)$  个操作, 我们重新安排了  $m$  个段, 使得它们的前导元素是有序的。而且, 由于对于该文件原来的假定, 在  $Z_1 \dots Z_m$  中的每个键码现在都有少于  $n$  个反序。

我们利用下列技巧, 可以合并  $Z_1$  和  $Z_2$ : 把  $Z_1$  同  $A$  的头  $n$  个元素  $A'$  进行交换; 然后以通常方式合并  $Z_2$  和  $A'$ , 但当它们被输出时与  $Z_1 Z_2$  的元素相交换。例如, 如果  $n=3$ , 且  $x_1 < y_1 < x_2 < y_2 < x_3 < y_3$ , 则我们有

	段 1	段 2	辅助区域
初始值的内容:	$x_1 x_2 x_3$	$y_1 y_2 y_3$	$a_1 a_2 a_3$
交换 $Z_1$ :	$a_1 a_2 a_3$	$y_1 y_2 y_3$	$x_1 x_2 x_3$
交换 $x_1$ :	$x_1 a_2 a_3$	$y_1 y_2 y_3$	$a_1 x_2 x_3$

71. 百度实习生笔试题:

一个单词如果交换其所含字母顺序，得到的单词称为兄弟单词，例如 **mary** 和 **army** 是兄弟单词，即所含字母是一样的，只是字母顺序不同，用户输入一个单词，要求在一个字典中找出该单词的所有兄弟单词，并输出。给出相应的数据结构及算法。要求时间和空间复杂度尽可能低

目前思想：

```
struct {  
  
    char data;  
  
    int n  
  
};
```

根据数学定理：任何一个大于 1 的自然数  $N$ ，都可以唯一分解成有限个质数的乘积  $N=(P_1^{a1})*(P_2^{a2}).....(P_n^{an})$ ，这里  $P_1 < P_2 < ... < P_n$  是质数，且唯一。

例如

$a=2 \ b=3 \ c=5 \ d=7 \ e=11...$

$f(abcd)=2*3*5*7=210$

然后字典里找乘积 210 的位数相同的一定是这 5 个字母组合的单词就是兄弟单词

72. 更新至 2012.05.06 下午.....

更多面试题，参见[横空出世，席卷 Csdn--评微软等数据结构+算法面试 100 题](#)（在此文中，集结了本博客已经整理的 236 道面试题）。