**百度2012实习生校园招聘笔试题**

阅读次数: 42次 发布时间: 2012-05-08 10:39:44发布人: pengzl

来源: 网络转载

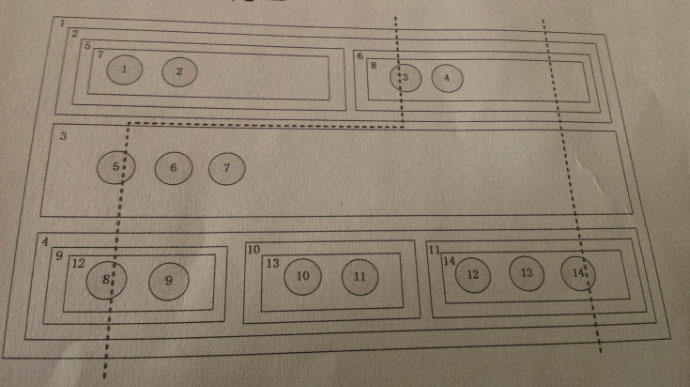
1、给一个单词a，如果通过交换单词中字母的顺序可以得到另外的单词b，那么b是a的兄弟单词，比如的单词army和mary互为兄弟单词。  
现在要给出一种解决方案，对于用户输入的单词，根据给定的字典找出输入单词有哪些兄弟单词。请具体说明数据结构和查询流程，要求时间和空间效率尽可能地高。  
字典树的典型应用  
2、系统中维护了若干数据项，我们对数据项的分类可以分为三级，首先我们按照一级分类方法将数据项分为A、B、C......若干类别，每个一级分类方法产生的类别又可以按照二级分类方法分为a、b、c......若干子类别，同样，二级分类方法产生的类别又可以按照是三级分类方法分为i、ii、iii......若干子类别，每个三级分类方法产生的子类别中的数据项从1开始编号。我们需要对每个数据项输出日志，日志的形式是key\_value对，写入日志的时候，用户提供三级类别名称、数据项编号和日志的key，共五个key值，例如，write\_log（A,a,i,1,key1），获取日志的时候，用户提供三级类别名称、数据项编号，共四个key值，返回对应的所有的key\_value对，例如get\_log（A,a,i,1,key1），  
请描述一种数据结构来存储这些日志，并计算出写入日志和读出日志的时间复杂度。  
  
3、C和C++中如何动态分配和释放内存？他们的区别是什么？  
malloc/free和new/delete的区别  
4、数组al[0,mid-1]和al[mid,num-1]是各自有序的，对数组al[0,num-1]的两个子有序段进行merge，得到al[0,num-1]整体有序。要求空间复杂度为O(1)。注：al[i]元素是支持'<'运算符的。  
5、线程和进程的区别及联系？如何理解“线程安全”问题？  
答案：进程和线程都是由操作系统所体会的程序运行的基本单元，系统利用该基本单元实现系统对应用的并发性。  
1、简而言之,一个程序至少有一个进程,一个进程至少有一个线程.  
2、线程的划分尺度小于进程，使得多线程程序的并发性高。  
3、另外，进程在执行过程中拥有独立的内存单元，而多个线程共享内存，从而极大地提高了程序的运行效率。  
4、线程在执行过程中与进程还是有区别的。每个独立的线程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。  
5、从逻辑角度来看，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以及资源分配。这就是进程和线程的重要区别。

算法与程序设计一、  
网页爬虫在抓取网页时，从指定的URL站点入口开始爬取这个站点上的所有URL

link，抓取到下一级link对应的页面后，同样对页面上的link进行抓取从而完成深度遍历。为简化问题，我们假设每个页面上至多只有一个link，如从www.baidu.com/a.html链接到  
www.baidu.com/b.html再链到www.baidu.com/x.html，当爬虫抓取到某个页面时，有可能再链回www.baidu.com/b.html，也有可能爬取到一个不带任何link的终极页面。当抓取到相同的URL或不包含任何link的终极页面时即完成爬取。爬虫在抓取到这些页面后建立一个单向链表，用来记录抓取到的页面，如：a.html->b.html->x.html...->NULL。  
问：对于爬虫分别从www.baidu.com/x1.html和www.baidu.com/x2.html两个入口开始获得两个单向链表，得到这两个单向链表后，如何判断他们是否抓取到了相同的URL？（假设页面URL上百亿，存储资源有限，无法用hash方法判断是否包含相同的URL）  
请先描述相应的算法，再给出相应的代码实现。（只需给出判断方法代码，无需爬虫代码）  
两个单向链表的相交问题。  
算法与程序设计二、  
4、有一种结构如下图所示，它由层的嵌套组成，一个父层中只能包含垂直方向上或者是水平方向上并列的层，例如，层1可以包含2、3、4三个垂直方向上的层，层2可以包含5和6两个水平方向的层，在空层中可以包含数据节点，所谓的空层是指不包含子层的层，每个空层可以包含若干个数据节点，也可以一个都不包含。  
在这种结构上面，我们从垂直方向上划一条线，我们约定每一个子层中我们只能经过一个数据节点，在这种情况下，每条线可以经过多个数据节点，也可以不经过任何数据节点，例如，线1经过了3、5、8三个数据节点，线2只经过了14个数据节点。  
（1）给出函数，实现判断两个数据节点，是否可能同时被线划中，给出具体的代码。

（2）给出函数，输出所有一条线可以划中的数据节点序列，

可以给出伪代码实现。

[](http://www.cn-java.com/www1/batch.download.php?aid=3130)  
 思路：（1）判断两个数所属的同一层次的相同矩形框的下一层次矩形框是水平排列的还是垂直排列的，垂直排列在能在一条线上，水平排列则不能。  
（2）用回溯算法求出所有在一条直线上的字符串，用两字符串是否在同一直线上进行剪枝操作。  
系统设计题  
1、相信大家都使用过百度搜索框的suggestion功能，百度搜索框中的suggestion提示功能如何实现？请给出实现思路和主要的数据结构、算法。有什么优化思路可以使得时间和空间效率最高？  
应用字典树来求前缀和TOP

K对热词进行统计排序  
2、两个200G大小的文件A和B，AB文件里内容均为无序的一行一个正整数字（不超过2^63），请设计方案，输出两个文件中均出现过的数字，使用一台内存不超过16G、磁盘充足的机器。  
方案中指明使用java编程时使用到的关键工具类，以及为什么?