

关于后缀数组，基本上是绕不开罗穗骞的论文《后缀数组——处理字符串的有力工具》。以下大部分的题目，都是从论文上来的，题解也将不写得那么详细，因为论文上有。所以强烈建议先详细研读此论文的第一部分，然后再看第二部分的题解。

POJ 1743 【基础】

题目大意：

有 N ($1 \leq N \leq 20000$) 个音符的序列来表示一首乐曲，每个音符都是 $1 \cdots 88$ 范围内的整数，现在要找一个最长的重复主题。“主题”是整个音符序列的一个子串，它需要满足如下条件：

- 1、长度至少为 5 个音符。
- 2、在乐曲中重复出现。“重复”可能经过“转调”，“转调”的意思是主题序列中每个音符都被加上或减去了同一个整数值。
- 3、重复出现的同一主题不能有重叠部分。

输入：

有若干组测试数据。每一组测试数据第一行有一个整数 N ， $N=0$ 时测试数据结束。第二行有 N 个空格分隔的整数，表示乐曲的序列。

输出：

每一组测试数据输出一行，即最长的重复主题的长度。如果找不到，输出 0。

题解：

见论文。

POJ 3261 【基础】

题目大意：

给出 N 个整数 ($1 \leq N \leq 20000$)，每个整数的范围在 0 到 1000000。问可重复 K 次的不重叠子串的最长长度是多少。

输入：

第一行有两个整数 N 和 K ，接下来有 N 行，第 $i+1$ 行给出序列上第 i 个整数的值。

输出：

一行，一个整数，即可重复 K 次的可重叠子串的最长长度。

题解：

还是先二分分子串长度然后分组，不过这一次判断的是是否有一个组使得组内的后缀个数不小于 K，因此要求组内所有的 $\text{height}[i]$ 都大于等于当前枚举的子串长度。

另外有一点，POJ 的数据很弱，按道理 $\text{wd}[]$ 应该被开到 1000000 以上，但是实际上我做的时候开到 200 也够了（因为代码是直接从 POJ 1743 的改过来的，很偷懒吧）。正规的做法应该是把基数排序改成快排。

POJ 2774 【基础】

题目大意：

给出两个字符串 A、B，它们的长度都不超过 100000 个字符，求它们的最长公共子串。

输入：

两行，每一行有一个字符串。

输出：

一行，一个整数，即最长公共子串的长度。

题解：

见论文。

POJ 3693 【中等】

题目大意：

给出一个长度不超过 100000 个字符的字符串，求其重复次数最多的连续重复子串。

输入：

有多组测试数据，处理到文件结尾。每一组测试数据有一行，包含一个字符串。如果字符串是“#”，测试数据结束。

输出：

每一组测试数据输出一行，格式为 Case %测试数据组编号%: %重复子串%，其中测试数据组编号从 1 开始计算。

题解：

论文上提到的题目，但论文上写得并不详细。下面我在论文部分的基础上写一个详细一点的。

按论文所说，枚举循环节长度 $sectlen$ ，然后再枚举起点 $startpos=0$ ， $sectlen$ ， $2*sectlen$ ……求某一个起点的后缀和下一个起点的后缀的最长公共前缀长度 $maxfit$ ，那么前面那个起点应该包含 $1+maxfit/sectlen$ 个循环节，其中前面的 1 是两个起点中间的那一部分。考虑到枚举的起点和当前枚举的循环节长度所得到的循环节的真正起点很可能不在同一处，可以由 $sectlen-maxfit\%sectlen$ 求出向前位移量，那么位移以后两个起点就在循环节的头部了。当得到新的循环节长度大于现有保存的最大值时，更新最大值并清空候选组；如果等于当前最大值，那么把循环长度加入到候选组中。最后选答案的时候，由于需要字典序最小，只需要按 sa 数组从 1 开始来枚举就可以了。

P0J 3294 【中等】

题目大意：

给出 N 个 ($1 \leq N \leq 100$) 长度不超过 1000 个字符的字符串，求一最大长度的子串，要求这个子串出现在大于 $n/2$ 个字符串中。

输入：

有若干组测试数据。每一组测试数据第一行有一个整数 N ，接下来有 N 行，每一行有一个字符串。 $N=0$ 时测试数据结束。

输出：

对每一组测试数据，输出一行或若干行，每一行有一个符合要求的子串，若有多个则按字母表升序输出，若没有则输出“?”。两组测试数据的输出之间有一个空行分隔。

题解：

论文里讲的分组有点不清楚。分组应该是指在 $height$ 数组中所有连续的大于等于当前枚举的子串长度的元素各自分成一组。比如 $height[] = 2\ 3\ 3\ 1\ 4\ 2\ 3\ 3\ 0\ 1\ 1$ ，当前测试的子串长度为 3，那么分出来的组是 $3\ 3/4/3\ 3$ 。然后求每一个组内的元素，比如 $height[2]$ 和 $height[3]$ ，对应的以 $sa[2]$ 和 $sa[3]$ 为起始的后缀，是属于原来的哪些字符串中的。统计最后有哪些字符串中出现了当前组的成员，如果数量 $> n/2$ ，那么这个长度就是可以的。

P0J 3581 【中等】

题目大意：

有一个 N 个元素的整数序列 ($1 \leq N \leq 200000$)，要将这个序列切分成三节，每一节在原来的位置上逆序，得到一个新的序列。问怎么切分可以使新序列最小。

输入：

第一行有一个整数 N ，接下来有 N 个整数。

单组测试数据，不要处理多组。

输出：

N 行，依次输出新序列的每一个数字。

题解：

考虑到要逆序，所以读入的时候直接反过来读。由于三段的相对位置不变，所以第一次求出 sa 数组后，排名在前面且 $sa[i] > 1$ (因为另两段不能为空) 的第一个 i 为第一段。接下来就是处理如何将剩下的一截分成两段。由于分完以后，一个将直接接在另一个后面，所以不能直接再求一次 sa 数组，因为这时长度会在大小比较时限制后缀 (也就是最后放在中间的那一块)。这里举一个例子。比如 3 2 0 1 0 1，如果直接求 sa 数组然后取 $sa[1]$ ，那么将会得到 0 1 3 2 0 1，但答案应该是 0 1 0 1 3 2。错误的出现是因为 $0 1 < 0 1 0 1$ ，但 $0 1 3 2 > 0 1 0 1$ ，所以我们需要把剩下那一段原样复制一次在后面，然后再求一次 sa 数组，取第一个起始位置在前半段 (也就是原来的长度内) 的后缀作为中间的那一段。如果不清楚的，可以看一下代码。

POJ 3415 【难】

题目大意：

给定两个长度不超过 100000 个字符的字符串 A 、 B ，求长度不小于 K 的公共子串个数。

输入：

有多组测试数据。每一组测试数据第一行有一个整数 K ，接下来有两行，每一行有一个字符串，分别是 A 和 B 。 $K=0$ 时测试数据结束。

输出：

每一组测试数据输出一行，包含一个整数，即所求公共子串的个数。

题解：

完全没搞明白这题，连样例数据都没看明白为什么 aa 和 aa 有 5 个子串。附上一个网上的题解，仅供参考：

<http://blog.csdn.net/woshi250hua/article/details/7520955>