

北京大学信息科学技术学院 2006-2007 学年

第一学期本科生期末考试试卷

考试科目：数据结构与算法实习 考试时间：2007 年 1 月 12 日

_____专业_____级_____班 考场_____室

姓名_____ 学号_____ 教员 张铭

题号	一	二	三	总分
得分				

诚实考试宣言：

我真诚地保证严格遵循考场纪律，不携带试卷出考场，也不会把考题以任何方式泄露出去。保证人：_____

一、简答题

- （15 分）对于一条四只脚呈长方形的椅子，试建立一个数学模型，论证一下能否在不平的地面上放稳。要求：
 - 对于建立的模型一定要做出合理的假设；
 - 做出假设后要把问题用严格的数学语言表述出来；
 - 对于自己所作的结论给出严格的数学证明。
- （10 分）典型的人-机交互风格有 5 种：
 - 直接操作；
 - 菜单选择；
 - 表单填充；
 - 命令行语言；
 - 自然语言；请列出其中 3 种风格各自的优缺点以及典型应用场景。
- （10 分）等价类划分是一种典型的黑盒测试方法，用这一方法设计测试用例完全不考虑程序的内部结构，只根据对程序的需求和说明，即需求规格说明书。请根据下面给出的规格说明，利用等价类划分的方法，给出足够的测试用例。
（只需要给出等价类即可）

“一个程序读入 3 个整数，把这三个数值看作一个三角形的 3 条边的长度值。这个程序要打印出信息，说明这个三角形是不等边的、是等腰的、还是等边的。”

4. （10 分）活动选择问题：

$S = \{1, 2, \dots, n\}$ 为 n 项活动的集合， s_i 和 f_i 分别是活动 i 的开始时间和结束时间，

活动 i 和活动 j 相容当且仅当 $s_i \geq f_j$ 或者 $s_j \geq f_i$

活动集如下：

I	1	2	3	4	5	6	7	8	9	10	11
s_i	0	3	1	3	5	5	6	8	8	2	12
f_i	6	5	4	8	7	9	10	11	12	13	14

答题要求：请写出排序后的活动集表，然后写出两两相容的最大活动集合。

5. （10 分）搜索引擎要处理海量的数据，在处理这些数据时，为了提高性能，系统往往要采取一些特别的措施。请你设想一下，为了提高系统的网页抓取或者索引或处理查询的性能，搜索引擎可以采取的措施。简要答出 3 点即可。

二、算法填空题：请补充 算法中的空缺，使得算法 完整并正确。

6. （8 分）“背包问题”：有一个背包，能盛放的物品总重量为 S ，设有 N 件物品，其重量分别为 w_1, w_2, \dots, w_n ，希望从 N 件物品中选择若干件物品，所选物品的重量之和恰能放入该背包，即所选物品的重量之和等于 S 。

```
#include<stdio.h>
```

```
int w[8] = {0, 1, 4, 3, 4, 5, 2, 7};
```

```
int knap(int S, int n) {
```

```
    if (S == 0) return 1;
```

```
    if (S < 0 || (S > 0 && n < 1)) return 0;
```

```
    if ( //? 1 ) {
```

```
        printf("%4d", w[n]); return 1;
```

```
    }
```

```
    return //? 2 ;
```

```
}
```

```
void main() {
```

```
    int N=7, S=15;
```

```
    if (knap(S, N) ) printf( "OK!\n" );
```

```
    else printf( "N0!\n" );
```

```
}
```

7. （12 分）人类基因可以用 4 种核苷酸转移酶 A、C、G、T 表示。生物学家对鉴

别基因以及确定它们的功能很感兴趣，因为这能用于诊断人类疾病进而研制针对这些疾病的药。人的基因可以通过一些生物试验以及在计算机的帮助下确定。一旦获得了基因序列，下一步的工作就是确定其功能。

一次数据库搜索会返回一个和查询基因序列相似的基因序列列表，生物学家假设基因序列相似代表着功能相似，所以我们需要对两个基因序列进行比较，决定它们的相似度。

给定两个基因序列 AGTGATG 和 GTTAG，它们的相似度是多少？一个度量相似度的方法就是对数据进行对齐，在一个校正中，如果需要的话，空格会被插入到适当的位置中，让两个基因序列长度相同。

例如：一个空格插入到 AGTGATG 中产生了 AGTGAT-G，三个空格插入到 GTTAG 中产生了 -GT--TAG（空格用 短横- 表示）。现在这两个序列等长：

AGTGAT-G
-GT--TAG

在这个对齐方式中，有四个基因是匹配的，每对匹配的字符都会根据右表分配一个得分：

	A	C	G	T	-
A	5	-1	-2	-1	-3
C	-1	5	-3	-2	-4
G	-2	-3	5	-2	-2
T	-1	-2	-2	5	-1
-	-3	-4	-2	-1	*

注意：空格对应空格的匹配是不允许出现的。上面一种对齐方式的得分为：(-3)+5+5+(-2)+(-3)+5+(-3)+5=9。

当然，其他的对齐方式也是可以的，比如：

AGTGATG
-GTTA-G

这种对齐方式的得分为：(-3)+5+5+(-2)+5+(-1)+5=14。所以这种对齐方式比前面那种更好一些。事实上，它是最好的一种排列方式。

请完善下列代码，使其能够给出两个基因序列的最大相似度。

```
char Weight[5][5] = {5,-1,-2,-1,-3,-1,5,-3,-2,-4,-2,-3,5,-2,-2,-1,-2,-2,5,-1,-3,-4,-2,-1,0};
int LenA, LenB, i, j;
int Sum;
bool Cur;
char F[256]; char Str[101]; char GeneA[101], GeneB[101];
short Sim[2][101];
inline short max( short ValA, short ValB )
{ return ValA > ValB ? ValA : ValB; }
void main() {
    F['A']=0; F['C']=1; F['G']=2; F['T']=3;
    scanf("%d", &Sum);
    while ( Sum-- > 0 ) {
        scanf("%d %s", &LenA, &Str);
        for ( i = 0; i < LenA; i++ )
            GeneA[ i+1 ] = F[ Str[ i ] ];
        scanf( "%d %s", &LenB, &Str);
        for ( i = 0; i < LenB; i++ )
```

```

        GeneB[ i+1 ] = F[ Str[ i ] ];
    Sim[0][0] = 0;
    for ( j = 1; j <= LenB; j++ )
        ① _____;

    for ( i = 1, Cur = 1; i <= LenA; ② _____ )
    {
        ③ _____;
        for ( j = 1; j <= LenB; j++ )
        {
            ④ _____;
        }
    }
    printf("%d\n",Sim[!Cur][LenB]);
}
}

```

三、数据结构与算法设计题

注意：除特别说明外，算法设计题都需要描述数据结构和算法思想，写出带必要注释的算法框架，分析算法空间和时间复杂度。

8. （15 分）进行一次竞赛，总时间 T 固定，有若干种可选择的题目，每种题目可选入的数量不限，每种题目有一个 t_i （解答此题所需的时间）和一个 s_i （解答此题所得的分数），现要选择若干题目，使得解这些题的总时间在 T 以内的前提下，所得的总分最大，求最大的得分。
9. （15 分）整数区间问题：一个整数区间 $[a, b]$ ， $a < b$ ，是一个以 a 开始以 b 结束的所有整数的区间。对于这样的整数区间的集合来说，我们希望找到一个最小的结果集合，使得整数区间集合中的任意一个整数区间，都至少有两个整数存在于我们的结果集合中。
10. （15 分）用 K 条直线分割一个 $M \times N$ 的矩形，将矩形分成许多小块，求三角形小块的数量。其中这 K 条直线只可能为横，竖和 45° 斜线三种情况。 M, N 均不超过 50，对应的 K 不超过 296。

- (1) 此题求解可以基于点采用枚举法。而枚举法的关键点之一是彼此独立，完全穷尽。请你基于上面的示意图，列举出针对每个点需要枚举的所有可能情形。
 - (2) 在第一步的基础上，请设计合理的数据结构，可以使用伪码、C 或者 C++ 语言，描述需要枚举的情形中的任一种。要求：采用良好的设计风格，包括变量命名、注释的规范性都属于考核范畴。
 - (3) 请分析基于点枚举的算法复杂度。本题不要求写出具体的算法。

