自动生成测试测试用例

# 问题：给一个函数的源代码怎么自动生成多个满足这个函数的输入参数组。

**意义：**测试最重要的就四个问题：（1）输入的自动生成；（2）自动判断（高效判断）对应输出是否正确；（3）知道有bug后如何定位（locate）；（4）自动（辅助）修复bug （auto fixed）。**我们要解决的是（1）输入的自动生成。**

**方法：**机器学习(自然语言处理、文本挖掘)+Judge Online Problems 数据集

**例如：**

[*http://poj.org/problem?id=1011*](http://poj.org/problem?id=1011)

**“Description**

George took sticks of the same length and cut them randomly until all parts became at most 50 units long. Now he wants to return sticks to the original state, but he forgot how many sticks he had originally and how long they were originally. Please help him and design a program which computes the smallest possible original length of those sticks. All lengths expressed in units are integers greater than zero.

**Input**

The input contains blocks of 2 lines. The first line contains the number of sticks parts after cutting, there are at most 64 sticks. The second line contains the lengths of those parts separated by the space. The last line of the file contains zero.

**Output**

The output should contains the smallest possible length of original sticks, one per line.

**Sample Input**

9

5 2 1 5 2 1 5 2 1

4

1 2 3 4

0

**Sample Output**

6

5

”

可以看出POJ的题目可以分为Description、Input、Output、Sample Input、Sample Output五部分。其中“Description”是对题目背景的描述，“Input”是对输入的描述，“Output”是对输出的描述，“Sample Input”是一个样本输入，“Sample Output”是一个样本输出。

**步骤：**

算法第一步：人为提出输出的几个特性（feature）比如，基本类型：int、double、string，组合类型：数组，固定大小的数组，范围：非负数，小于多少。

算法第二步：根据“Input“和“Sample Input”给出每道题feature的取值。

算法第三步：使用Lucence 索引每道题的”input”段文字。

算法第四步：选择机器学习算法给出“Input”索引同对应feature值之间的关系。

**关键技术（子问题）：**

1. 爬虫 我们需要POJ上至少1千道题目*。*这个工作工作由云谦同学完成并把每道安各个部分存入数据库中以备后期使用。
2. 要代码 没有正确代码我们没办法自动判断我们学到的输入是否正确。建议由婷婷同学负责。
3. 索引和机器学习算法 具体的算法我后期会给出几个后选项。这个工具我建议由小龙同学来完成。
4. Judge Online 环境搭建。 这个由建议由我和婷婷同学来弄*。*
5. 相关文献阅读和方式搜集。做这个事情对外体现之一是文章，希望我们能在几个月完成一篇文章（如何过程中由其它有趣的问题，我们也可以多写几篇） 建议由雪莲和秦晓完成。

上面列的子问题都可以并行进行，所以希望大家能分开完成自己的问题。