

## 一、必填项

必填项前五项



前面五项是这个程序运行最基本的基础，填写不对就无法运行（下面五项其实也是，不过一般按默认值运行也不会直接挂掉）。

填写规则是这样的：

可以只填写 cad 中显示的部分连续的字符串，例如：cad 中的【层底深度 (m)】可以填成【层底深度】或者【底深】或者【底深度(m)】等，要注意的是填写的字符串不能在别的项也存在，比如【层底深度(m)】和【层底标高(m)】中都包含的字符串【层底】，就不能填到程序中。

但是，由一个一个文字对象 (Text) “纵向排列”而成的字符串假如多于两列，比如【地质时代及成因】，在程序中读取出来其实会是【地及质成时因代】(文本对象先按 y 坐标降序排列，当 y 相等，再按 x 升序排列，再按得到的顺序将文字内容相连)

A	B	C	添加条件(A)    删除条件(D)    复制条件(C)    选项(O)... <input checked="" type="checkbox"/> 数据包含标题(H)		
x坐标	y坐标	文字内容	列	排序依据	次序
2250.836974	1333.257321	‘地’	主要关键字	y坐标	降序
2254.982429	1330.900233	‘及’	次要关键字	x坐标	升序
2250.836974	1328.751517	‘质’			
2254.982474	1326.603112	‘成’			
2250.836974	1324.232046	‘时’			
2254.982474	1322.092176	‘因’			
2250.836974	1319.613538	‘代’			

介绍一下 Text (文字) 和 MText (多行文字)

\*在 cad 中一般用到两类文本对象，分别是 Text (文字) 和 MText (多行文字)，



右键特性查看对象类型

1. Text (文字): 每个字符串分别是独立的对象，选中后左下角可见一个蓝色点 (对象插入点 InsertionPoint，下面用到的 xy 坐标也是指这个点的坐标)

这种情况可以只填其中的一个比较特殊的字，比如【因】字，它在其他地方都不会出现，

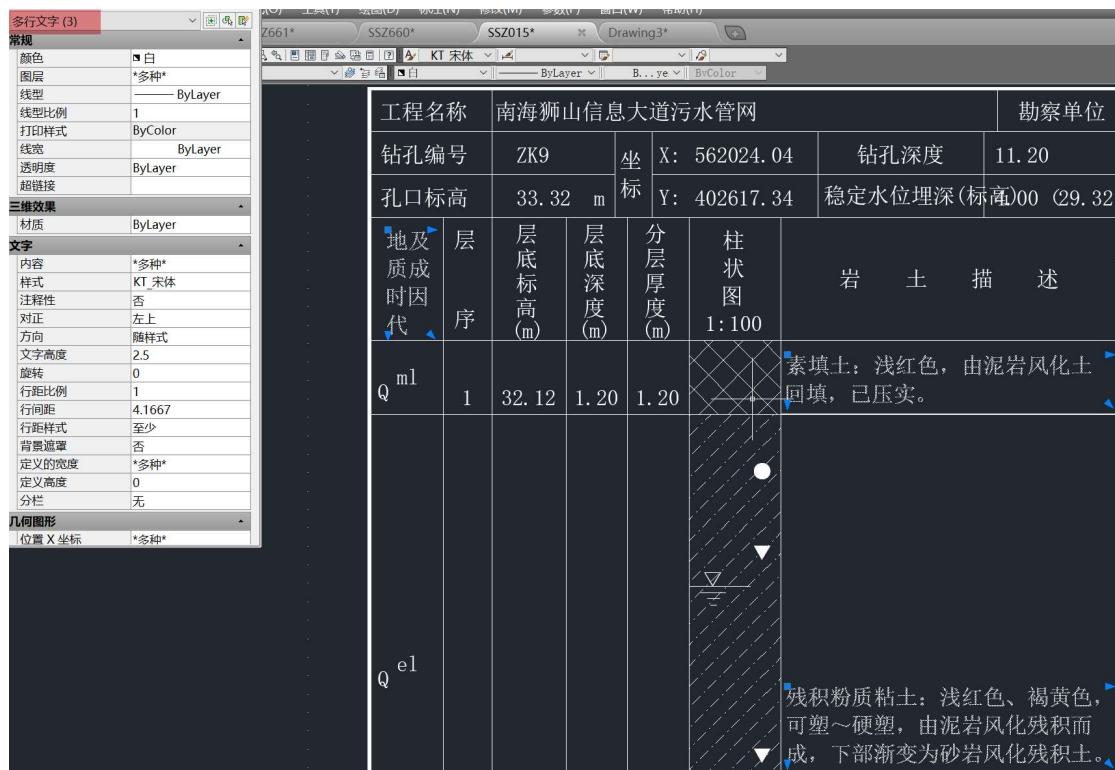
必填项字段名	CAD中的名称
柱状图标标题	钻孔柱状图
层底深度	层底深度
时代成因	因
剖面层号	层序
岩土描述	岩土描述
两侧宽度增加值	0
底部高度增加值	0

，或者实在不好区分，就按上面提到的排序

规则选取几个字填写，比如这里可以填【地及】【成时因】等等



2. MText (多行文字): 经常出现在岩土描述中, 单击它四角都会有蓝色的标记



对于多行文字中“纵向排列”的字符串, 比如上图的【地质时代及成因】, 应该双击它并直接复制内容粘贴到程序输入框

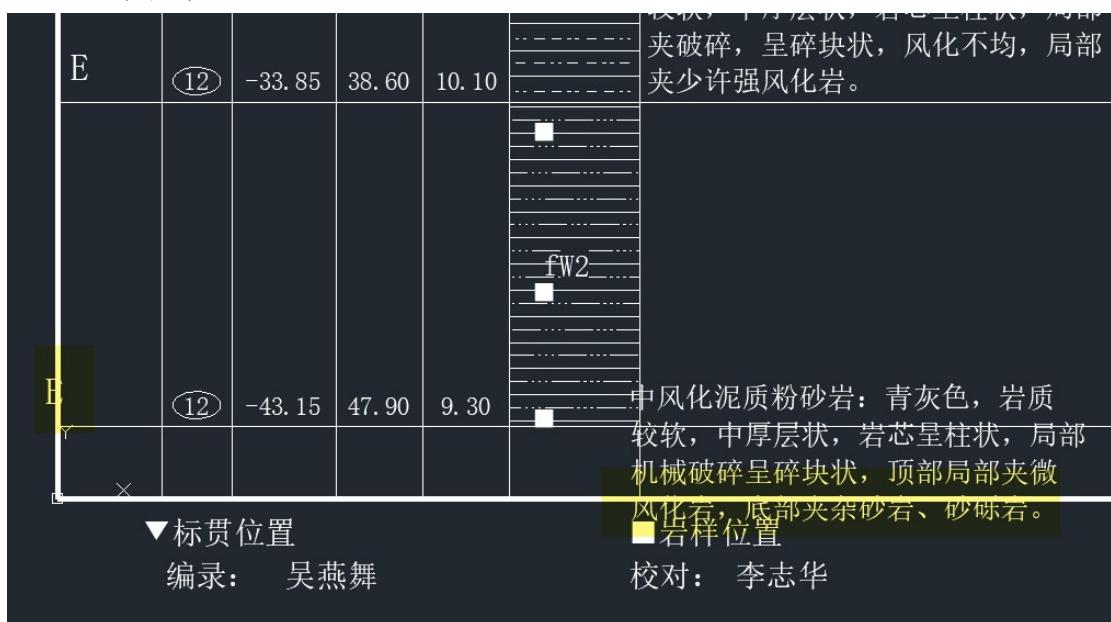


粘贴后可以看到顺序其实也是乱的

必填项后五项：

两侧宽度增加值	0
底部高度增加值	0
插入点或中心点	中心点
岩土描述排序方式	A
土工标贯形式	分数

\*两侧宽度和底部高度增加值：以前测试的柱状图数据出现过由于字数太多部分文字超出最外面的框的情况，所以这里的两个增加值其实是人工给这个柱状图范围往外(左右和下方)增加一定宽度，使得这个范围可以包含所有文字对象以便在后续的搜索中不会漏掉超出框外的文字。如果遇到这种情况，增加值填多大可以移动鼠标看着 CAD 左下角坐标栏算个大概值



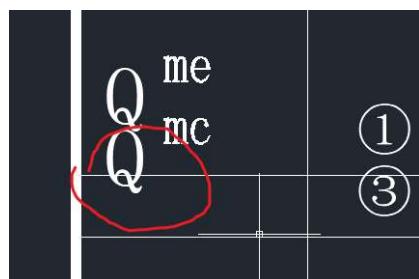
\*插入点和中心点是用来做坐标比较的，这个程序的基本原理就是读取出 cad 文件中各点线对象的坐标，继而得到柱状图中的一个一个由竖线和横线组成的“格子”的坐标范围，结合在程序输入框中填写的各种内容得到目标结果应该落在的“格子”的坐标范围，以这个范围去筛选读取出的文字对象坐标，最后把符合坐标范围的一个或多个文字对象的内容按一定规则进行拼接得出结果。

插入点就是上面提到的点击 cad 对象左下角的那个点的坐标，所有 cad 对象都会有的一个属性（应该是吧），这个值是可以直接读取出来的；中心点是我为了应对特殊情况计算出来的，自定义为最小外接矩形（这个属性 BoundingBox 的四个顶点坐标也可以直接读取出来）的中心，也就是矩形对角线中点

地 质 时 代	及 成 因	层 序	层 底 标 高 (m)
		③	-0.25

0<sup>me</sup>  
~~Q<sup>mc</sup>~~

为什么要用这个中心点的坐标呢，这大概是因为不同柱状图生成程序的设计者有不同的喜好或者要求，用之前数据测试的时候遇到有些人喜欢把文字对象放到格子的正中间，而有些则是按插入点放置对象，比如下图



这个 Q 应该是属于下面比较窄的一层，事实上它的插入点（左下角）也的确落在下面那个小格子里，但是这个 Q 的大部分其实已经处在上面的格子中了，这种情况如果程序中选择了【中心点】的选项，那么结果会把这个 Q 与上面那层的其他文字对象相连，如果选择【插入点】，则会把它识别为下层的时代成因。当然，像上图这个下面的时代成因 Qmc，实际上 Q 和 mc 分别是两个文字对象，而右上角的 mc 无论是插入点或是中心点的坐标都落在上面的格子中了，这种情况即使选择【插入点】的坐标方式，下层也只会读出一个 Q，而 mc 会被连到上层的 Qme 中，这种情况在层厚太小的层中可能会出现，但是一般最后将结果复制到 excel 中浏览会很容易发现，上层的结果大概是

`[Q @ me @ mc]` (从不同文字对象中读取出的字符间以'@'连接)，下层则只有一个 `[Q]`。

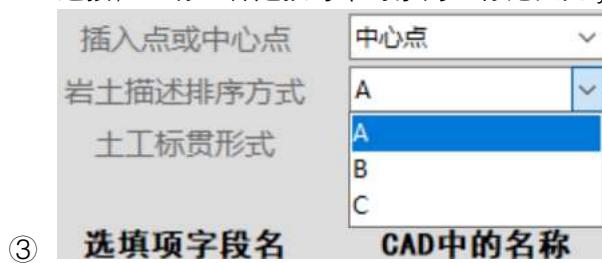
岩土描述排序方式：这里写了三种方法，分别为 A、B、C，各自具体排序方式我也忘了，要看代码。

介绍下对岩土描述进行排序的原因，不同模板的柱状图会用不同的方式组织岩土描述的，具体也是上面提过的两种：Text（文字）和 MText（多行文字）

先说多行文字，如果制作者用的是多行文字的对象来表达描述，那么读取过程就十分简单了，遇到过的描述都是横向排列，也就是从左往右，从上往下的方式，多行文字直接读取也不会出现竖向排列那种乱序的情况，这时候直接读取多行文字对象的内容就可以得到完美的结果。

然而，有相当一部分柱状图是用 Text（文字）对象来表达岩土描述的，这个时候最好的情况是每句话只有一行，并且一句话就是一个文字对象，这样读取基本与多行文字无异，但是，更为复杂的情况是大多数：

- ① 一段描述不止一行，这样需要把不同行连接起来，这个也是比较简单，基本不会出问题
- ② 岩土描述里的逐个字是拆开的，也就是每个字（有些是几个字）是一个 Text 对象，在按当前分层对应格子的范围搜索文字对象时会筛选出数十个符合坐标范围的文字对象，但是它们都是独立存在的，这个时候就需要对这些文字对象进行排序再连接成一段完整的话，在程序中根据测试时遇到的情况写了三种方式，如果结果顺序很奇怪，可以切换其他两种试试（记得其中有一种是按内部编号的顺序连接，还有一种是按每个对象的坐标范围先 y 降序再 x 升序得到的排序）



土工标贯形式：分数和单行，只能处理这两种，再复杂的整不了了，对应选择

	$\frac{\text{LBZK1-1}}{1.20-1.40}$
$10(9.2)$	

这是分数

6.10-6.30	ZK48-1	$fr=2700$
-----------	--------	-----------

这是单行

## 二、钻孔信息字段 (CAD)

\*带个“CAD”是想说明填写的内容是 CAD 中存在的字符串。

这个区域获取的内容对程序运行没有影响，填不填都无所谓，填错了也没关系，按需要选择填写即可

数据获取来源是表中最上面的一些基本信息

三个方向“右”“上”“下”是指目标字符串位于所填字段**相邻**格子的特定方向，举个例子，

【ZK9】是需要获取的信息，它位于【钻孔编号】的【右】侧相邻格子，所以

钻孔编号 右 这样填就行

又比如

工程名称	兴业路排水管道工程					勘察单位	广东佛山地质工程勘察院		
钻孔编号	孔口标高	孔 径	静止水位	孔 深	开孔日期	2014/3/16	编 录	制 图	审 核
ZK48	6.68 m	91 mm	1.75 m	15.80 m	终孔日期	2014/3/16	邓志雄	邓志雄	李志华

这里的【ZK48】位于【钻孔编号】下方的格子，所以方向选择【下】

说一种特殊的情况：

首层入土情况小结表									
	坐	X: 562024.04							
m	标	Y: 402617.34			稳定				
层	分	中	柱						

这个【坐标】的右边有两个格子【X:123456】和【Y:123456】，这个时候有两种情况要注意：首先点击【坐标】这个对象，①如果【坐】和【标】是独立的两个对象，那么程序中不能填成【坐标】，只能填【坐】或者【标】（与必填项不同，在必填项里是可以直接填合并后的字符串，因为那是后来做的，这个部分是最开始写的，写的时候没考虑把同一个格子内分开的字符连起来这个操作，后来写久了代码也乱了就一直没加上）。②如果【坐标】本来就是一个整体当然就填【坐标】好了……

填好【坐】后，运行程序会读取到类似这样的结果：【X: \$ Y: \$ 562024.04 \$ 402617.34】也就是把右边上下两个格子的内容都读取到了并且拼接在了一起，这是因为在这个部分，寻找目标文字对象的规则是这样的：

程序读取到输入的【坐】这个关键字，然后从当前处理范围（当前柱状图外框内的区域）中寻找包含“坐”的文字对象，找到后会得到它的插入点或者中心点坐标，再根据这个坐标x,y值，在线坐标集合中的找到距离该点最近的上下两条横线和左右两条竖线（下图黄框），从而得到“坐”字所在格子的四个角点坐标，这时，就可以知道，“坐”字所在格子的范围( $x0, x999), (y0, y999)$ 坐标极值，如果要向【右】寻找目标文字对象，则会在线集合中筛选出x坐标大于 $x999$ 的竖线，然后选择这些竖线中x值最小的那条，也就是“坐”字所在格子右侧相邻格子的右边竖线（下图绿线），再加上 $(y0, y999)$ 的y坐标范围，就可以得到“坐”字所在格的右侧格子坐标范围（下图【X:56xxxx】和【Y:40xxxx】两个格子），用这个范围从文字对象集合中筛选出包含两个坐标值的文字对象，最后将内容相连，就会得到结果【X: \$ Y: \$ 562024.04 \$ 402617.34】。

首层入土情况小结表									
	坐	X: 562024.04							
m	标	Y: 402617.34			稳定				
层	分	中	柱						

Diagram illustrating the coordinate extraction process:

- The "坐" character is highlighted in the first column.
- A yellow box labeled "x999" and "y999" indicates the bounding box of the "坐" character cell.
- A red arrow points from the "坐" cell to the right boundary of the column, which is highlighted by a green vertical line labeled "x0" and "y0".