## 1、概述：

本次标注的“命名实体”可以细分为：

（1）表示区位的词组，如：A区、5层、孔底。

（2）表示建筑构件的词汇或者词组，如：楼梯、10#塔吊、上箍筋、钢筋网、人工挖孔桩、高支模、钢筋连接、钢筋丝牙。

（3）表示建筑材料的词汇，如：混凝土、钢筋。

（4）建筑构件加“位置后缀”形成的词组，如：10#塔吊周围、悬挑梁底部、高支模区域、楼梯间、施工缝部位。

## 2、标注规则：

（1）采用‘B’‘I’‘O’三种标签标注出语料库中的“命名实体”。‘B’代表命名实体的开头，‘I’代表命名实体的其他部位，‘O’代表标注的单词不属于“命名实体”。下面为标注的例子：



（2）在标注语料中会经常出现如下情况：多个“命名实体”连续排列的情况，如：“A区5层悬挑梁底部箍筋”，此时需要将该字符串有效拆分为多个“命名实体”。拆分的原则是：

①以表“位置”后缀词作为拆分标志。如“A区5层悬挑梁底部箍筋”中，表位置的后缀词为“区”“层”“底部”，那么该词组应拆分为“A区”、“5层”“悬挑梁底部”“箍筋”。

②对于没有表位置的后缀词的情况。如“悬挑梁箍筋”虽然可以分为“悬挑梁”和“箍筋”，但是仍然将其作为一个命名实体处理。

## 3、标注技巧

由于需要标记的“O”很多，所以在标注时不标注“O”只标注“B”“I”即可

## 4、备注

若标注过程中，有拿捏不准之处，烦请沟通。