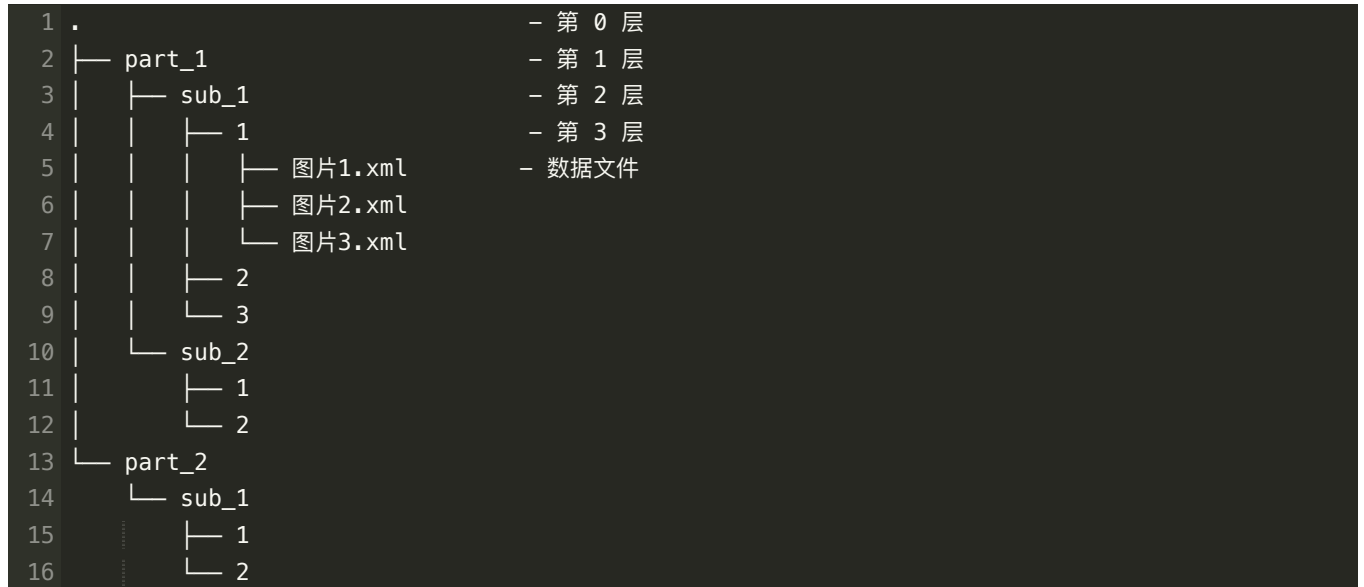


实习生笔试题

将代码跑出的答案(./intern文件夹)和代码一同打包后邮件发送到 recruit@stardust.ai.

请按照要求处理某路径下的文件, 文件结构见下图, 路径总共 3 层, 所有待处理文件在文件结构的第 3 层.



数据是 xml 文件, 如下:

```
1 <annotation>
2   <filename>图像 4251.png</filename>
3   <size>
4     <width>951</width>
5     <height>617</height>
6     <depth>3</depth>
7   </size>
8   <segmented>0</segmented>
9   <object>
10    <difficult>0</difficult>
11    <name>&*@HUST_text</name>
12    <bndbox>
13      <xmin>1</xmin>
14      <ymin>6</ymin>
15      <xmax>41</xmax>
16      <ymax>27</ymax>
17    </bndbox>
18  </object>
19  <object>...</object>
20 </annotation>
21 <!--
22 1. object 个数大于等于 0;
23 2. object name 字段中可能出现的值有:
24     ""
25     &*@HUST_text
```

```

26     &*@HUST_rotate
27     &*@HUST_special
28     &*@HUST_vertical
29     &*@HUST_shelter
30     内容错行
31     """"
32 3. bndbox xmin, ymin, xmax, ymax 描述了矩形的坐标情况,代表矩形四个顶点在坐标轴当中的范围。分别表示矩形在
    坐标系(以图片左上角为原点,向右作为 x 轴正轴,向下作为 y 轴正轴)中 x 最小值, y 最小值, y 最小值, x 最大值
    , y 最大值。
33 -->

```

根据上述要求对数据进行适配,要求如下:

1. 将 filename 和 size 中的数据合并为 info;

```

1 {
2     "name": "图像 4251.png",
3     "width": 951,
4     "height": 617,
5     "depth": 3
6 }

```

2. 将 object 数据转换为 annotation;

```

1 [
2     {
3         "difficult": 1,
4         "coordinate": {
5             "x": 28,
6             "y": 29
7         },
8         "width": 69,
9         "height": 21,
10        "type": "&*@HUST_text"
11    },
12    {...}
13 ]
14 # coordinate 为矩形左上角坐标, width 和 height 分别为矩形的宽和高
15 # 若 object 不存在, 则 annotation 为空列表

```

3. (可选)输入层数, 层数范围为 [0, 3], 将指定层数下的所有数据文件(需要递归处理子文件夹)中的数据经过前两个步骤的转换后生成 csv 文件保存在指定的文件夹层级下, 文件名为 result.csv.
指定生成数据格式如下:

```

1 info,annotation
2 {"name": "图像 4251.png", "width": 951, "height": 617, "depth": 3},[{"difficult": 1,
    "coordinate": {"x": 28, "y": 29}, "width": 69, "height": 21, "type": "&*@HUST_text"},
    {...}]

```

4. 若在读取数据文件的时候发生错误, 将错误文件的路径记录到给定文件夹的根目录下的 error.txt 中.

要求 3 为可选, 若不选择要求 3, 则将给定文件夹下所有数据文件按其他要求适配好后置于给定文件夹的根目录中.

最后写出自己所写程序中没有考虑到的可能会导致处理出现问题的地方, 并且你会怎么优化, 请写出你的思路.

