1. 开发背景

设计院每个月需要出图，有一些图纸图号及名称中有连续序号，需要拆分成各分册，类似下图所示。需要把黄色部分转换成绿色的部分。每个月图号及名称往往有上百条，手工处理费时费力，急需自动化处理的软件。



1. 初步梳理需求分析

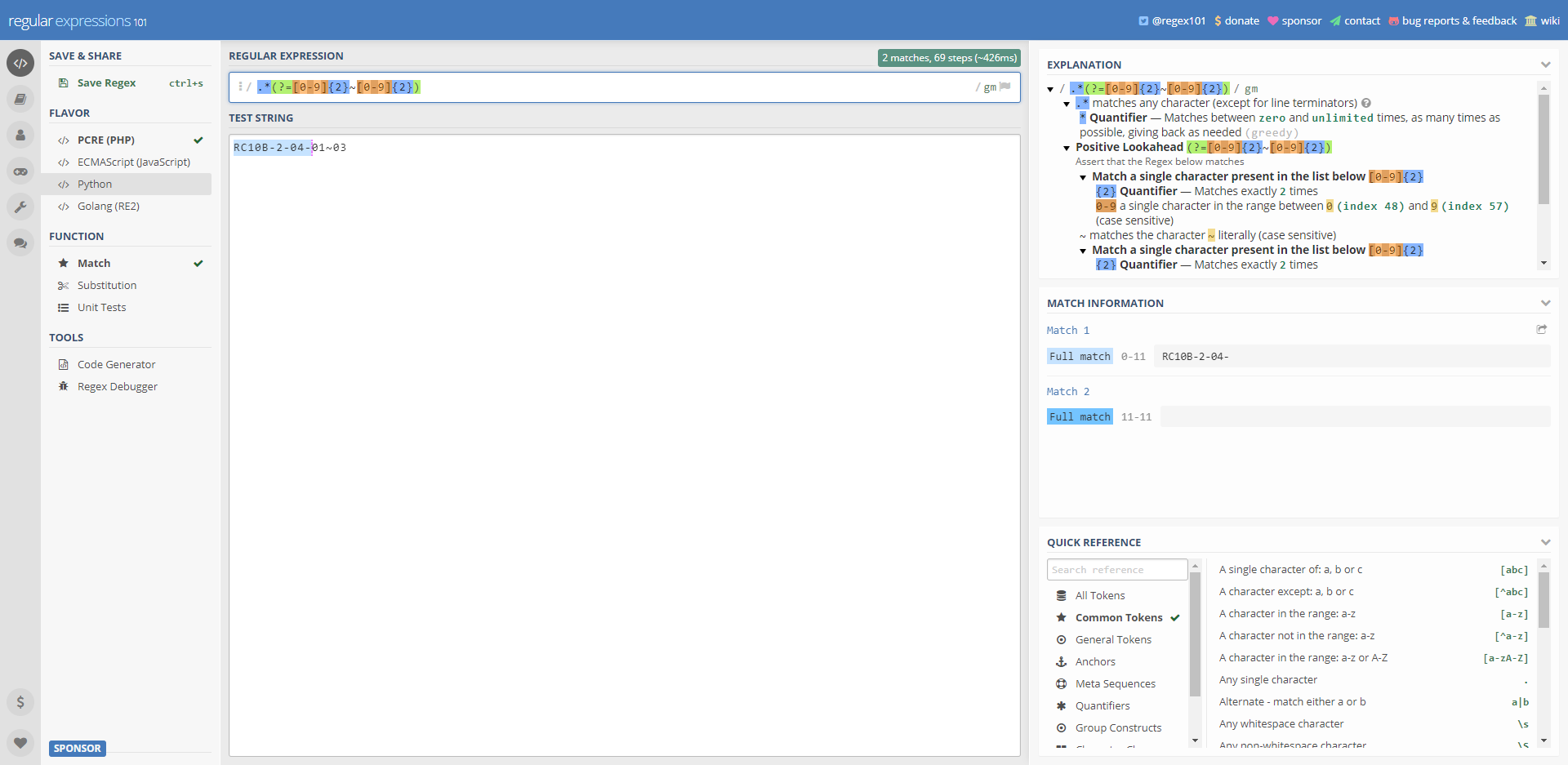
图纸包括图号和名称，图号和名称是一一对应的。

图号若有序号，序号占6位，格式为-xx~xx，xx为2为数字，范围为01-99。

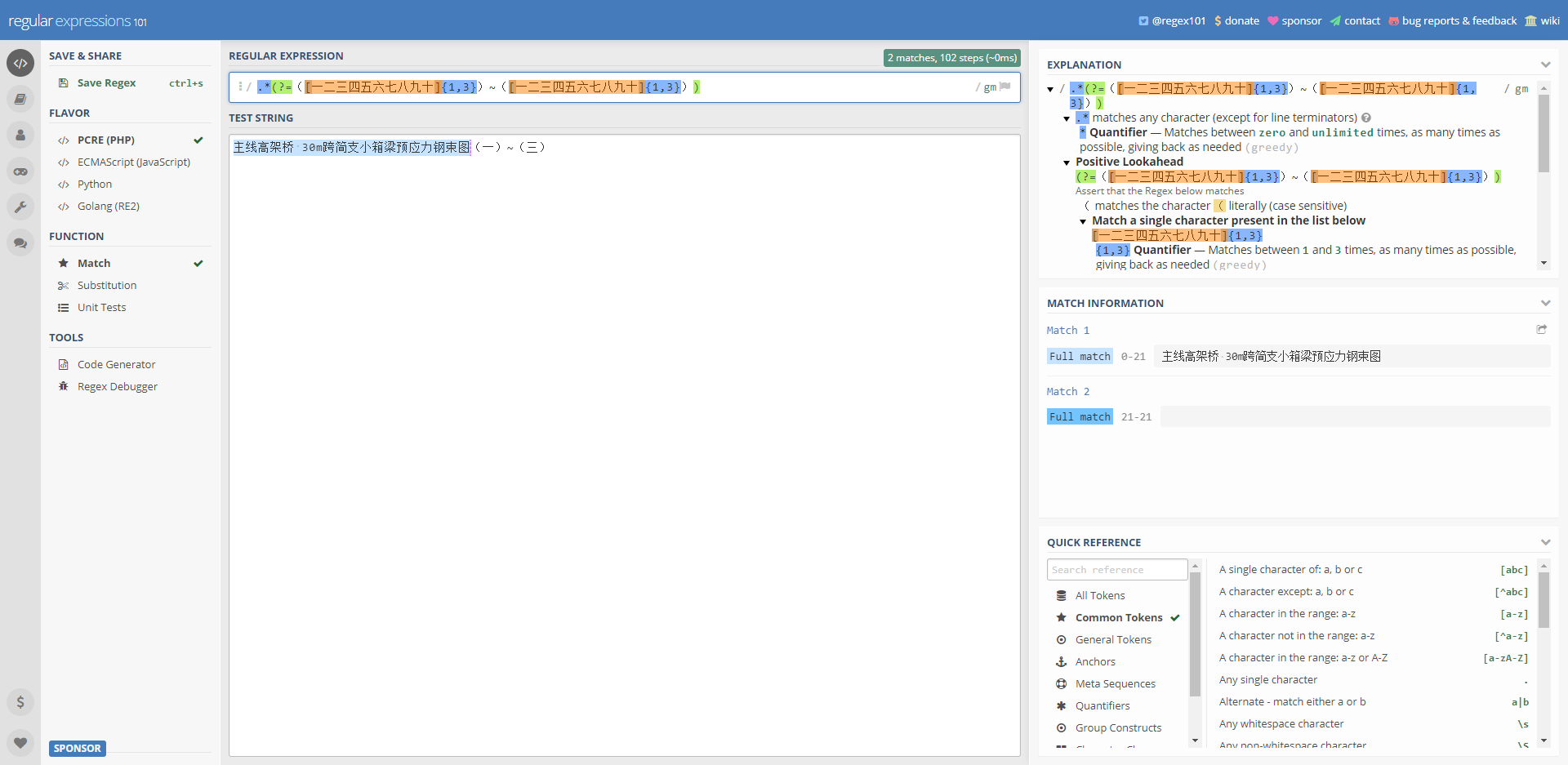
1. 图纸若只有1册，则没有序号，没有序号的，处理方法为保持不变。例：“RC10B-2-06-1-01”及“主线高架桥 30m跨简支小箱梁C型、D型中梁钢筋图”原封不动转换为“RC10B-2-06-1-01”及“主线高架桥 30m跨简支小箱梁C型、D型中梁钢筋图”。
2. 图纸超过1册，有连续序号的，最小数字为1，最大数字为图号末尾2位数。例：“RC10B-2-04-01~03”及“主线高架桥 30m跨简支小箱梁预应力钢束图（一）~（三）”程序需要分别拆成3册，即“RC10B-2-04-01”、“RC10B-2-04-02”、“RC10B-2-04-03”及“主线高架桥 30m跨简支小箱梁预应力钢束图（一）”、“主线高架桥 30m跨简支小箱梁预应力钢束图（二）”、“主线高架桥 30m跨简支小箱梁预应力钢束图（三）”。
3. 进一步梳理需求分析

为便于程序分析，将图号及名称中序号部分分离，剩下部分称为“主体字符串”。

3.1利用正则表达式.\*(?=[0-9]{2}~[0-9]{2})匹配图号来识别图纸是否只有1册，匹配数量>=1说明不止1册，匹配不到说明只有1册。如果只有1册，原封不动保留，若超过1册，截取后2位作为数量，截取第1位至倒数第7位作为“主体字符串”。可以在在网站<https://regex101.com/>中测试匹配结果，如下图所示：



3.2利用正则表达式.\*(?=（[一二三四五六七八九十]{1,3}）~（[一二三四五六七八九十]{1,3}）)匹配名称提取图纸名称中的“主体字符串”，如下图所示：



有连续序号的，将图纸的图号及名称中主体部分及序号分割开，根据序号最后两位得到每份图纸的总册数，然后使用for循环自动生成。

1. 编程思路

4.1 读取excel文件中的图号和名称，分别保存为sourceStringList1及sourceStringList2，数据结构为List<string>。

4.2 利用正则表达式.\*(?=[0-9]{2}~[0-9]{2})匹配图号来判断图纸是否只有1册，若只有1册，图号及名称原封不动保留，若不只1册，截取图号最后两位作为该份图纸数量（该变量为maxNumber），截取第1位至倒数第7位作为“主体字符串”，利用for循环生成拆分后各图纸分册的图号。

4.3 图纸若不只1册，利用正则表达式.\*(?=（[一二三四五六七八九十]{1,3}）~（[一二三四五六七八九十]{1,3}）)提取图纸名称中的“主体字符串”，再利用匹配图号过程中得知的最大数量（maxNumber），利用for循环生成拆分后各图纸分册的名称，其中数字需要转换成中文，参考了<https://www.cnblogs.com/zuqing/p/4567384.html>的代码，参考的代码发现一个bug，转换10-19的代码结果不正确，已在本人程序中进行了修复。

完整代码详见：<https://github.com/zjkl19/AutoSeparateSequence/tree/master/AutoSeparateSequence>，其中读写excel使用到了EPPlus插件，详见https://github.com/JanKallman/EPPlus