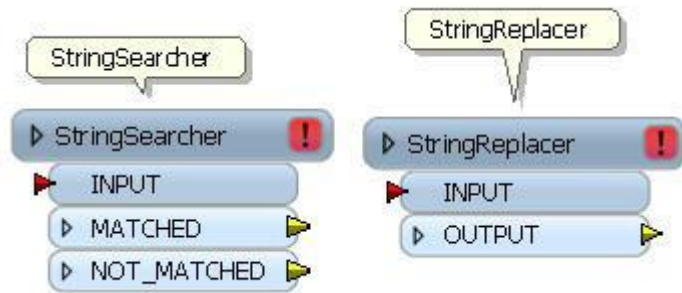


FME 中最为常用的字符串处理函数有二个：StringSearcher、StringReplacer

StringSearcher 为查询匹配，StringReplacer 是可以对查询到的属性进行替换



属性内容匹配提取

1. 查询属性值里是否包含数字: [0-9]

不限制匹配的值是否包括中文\英文\符号



解释：只要值包含数字，就把第一个数字做为匹配结果

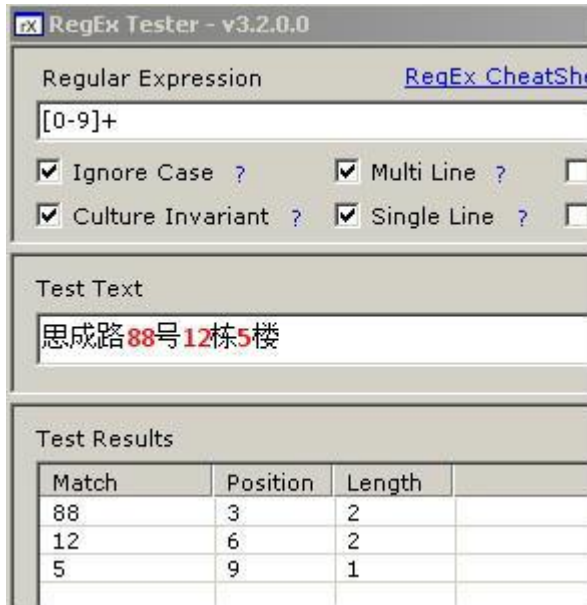
匹配成功例子：123 思成路5号（其中123匹配结果为1，思成路5号匹配结果为5）

匹配失败例子：思成路一号

2. 查询属性值里的数字串: [0-9]+

解释：只要出现的数字，就匹配出来。FME的表达式这里面有个问题，就是匹配结果只能是第一个数字串，第二个以后的无法匹配出结果。

比如：思成路88号12栋5楼，匹配出来的结果是88，后面的12和5无法显示出来。其他工具匹配出来的结果是有的。如截图：



3. 查询属性值里只有数字: `^[0-9]+$`

解释: 可参考 FME 的帮助

`^` 是指匹配开头的值,值为后面的[0-9]区间

`[0-9]`是指匹配的内容是纯数字

`+` 是指一个或多个 0-9 区间的值

`$` 是指结尾匹配值只能是 0-9 区间的值

匹配成功例子: 1 123 4455

匹配失败例子: 思成路 2 号

4. 奇偶数匹配提取: `([13579]|[1-9]\d*[13579])$`

解释:

匹配单字节的 1 3 5 7 9 或者 1-9 区间为第一位,中间数字无限制,最后一位只能是 1 3 5 7 9 的数值。

5. 指定字符串的提取: `(([0-9a-zA-Z-]+)(栋|幢|座))\S*`

在处理地址串中,经常要截取其中的一串字符。

解释:

`[0-9a-zA-Z-]` 以数字或英文或符号“-”为开始位置

`+` 一个或多个[0-9a-zA-Z-]包含的值

`(栋|幢|座)` 上一个匹配值的后面只能是(栋|幢|座)其中一个中文

`\S*` 任意非空白字符

匹配成功例子:

思成路 888 号 1 栋 5 楼 401 房

思成路 888 号 21 栋 5 楼 401 房

思成路 888 号 1-B 栋 5 楼 401 房

思成路 888 号 B1 栋 5 楼 401 房

思成路 888 号 21 幢 5 楼 401 房

思成路 888 号 1B 栋 5 楼 401 房

思成路 888 号 B1 栋 5 楼 401 房

匹配失败例子:

思成路 888 号 B 北栋东梯

未能匹配是因为多了个北字未在匹配规则里, 如何把这种类型处理好, 看下一个表达式。

6. `(([0-9a-zA-Z-]+)\S?(栋|幢|座))\S*`

这样就可以把思成路 888 号 B 北栋东梯匹配成功

解释:

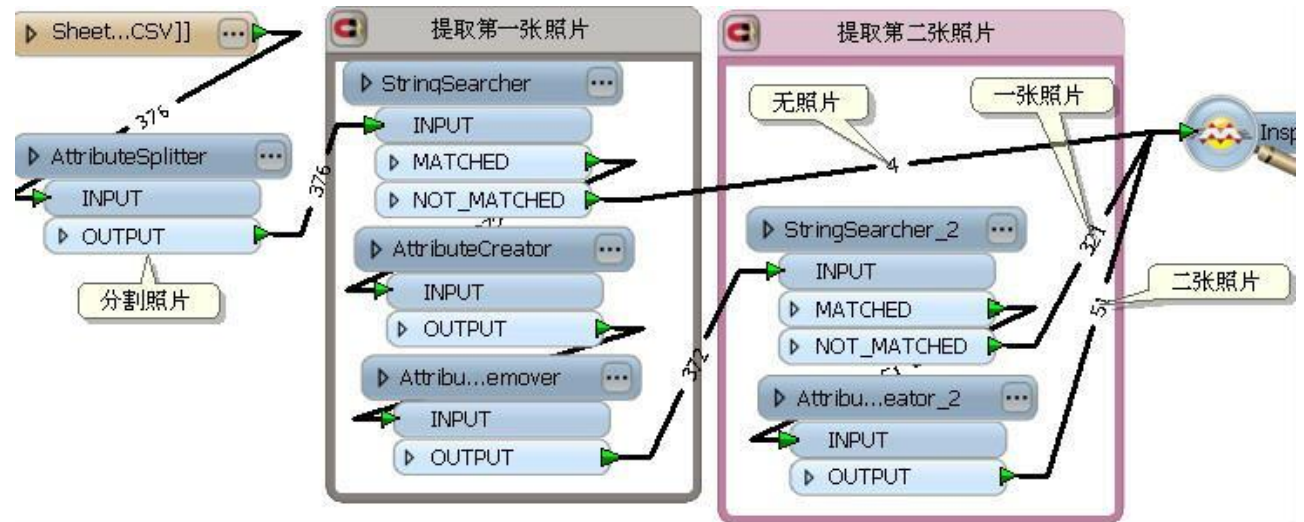
为什么 `\S?` 不写为 `\S*` 如果写成 `\S*`, 前面的 888 号也会匹配到结果中

7. 指定字符串提取 (2): `[0-9a-zA-Z-]+.jpg` 或 `[^~]+.jpg`

如下所示, 要把 `1t72wjb1019d199.jpg` 和 `1t72wjb1019d199a.jpg` 提取出来。

```
<?xml version="1.0" encoding="utf-8"?><multimedia><pictures><picture name="1t72wjb1019d199.jpg"><desc><![CDATA[门址照片]]></desc></picture><picture name="1t72wjb1019d199a.jpg"><desc><![CDATA[门址照片]]></desc></picture></pictures><audios></audios></multimedia>
```

这里要用到另一个函数: `AttributeSplitter`, 先对二个 JPG 之间进行分割, 分割符可以设置为 "[门址照片]", 然后再对分割后的 list 分别提取。



运行后的结果:

属性名称	属性值
_list{0}	<?xml version="1.0" encoding="utf-8"?><mi
_list{1}]]></desc></picture><picture name="lt72w
_list{2}]]></desc></picture></pictures><audios/;
_matched_characters	lt72wjb1019d199a.jpg
_matched_parts{0}	lt72wjb1019d199a.jpg
_matched_parts{1}	.jpg
addrType	自编
bh	lt72wjb1019d199
bigType	场所
csv_line_number	39
csv_number_of_fields	16
csv_type	csv_none
dw	琳琅时装店
dyh	
dzmx	
floor	1
fme_basename	Sheet2
fme_dataset	F:\中堂数据\FME20130118\excel\Sheet2.cs
fme_feature_type	Sheet2
fme_geometry	fme_undefined
fme_type	fme_no_geom
imagePath	<?xml version="1.0" encoding="utf-8"?><mi
jlx	共联商业街
link	F:\程序\共联\1022\朱良勇\合并images\
lp	
middleType	商贸场所
mp	中70号
natureofhouse	平房
notes	
pcsmc	
PHOTO1	lt72wjb1019d199.jpg
PHOTO2	lt72wjb1019d199a.jpg
roomno	
smallType	商铺

5. 提取汉字: [\u4e00-\u9fa5]

思成路888号21栋5楼A区401房

FME 中的正则表达式常用字符:

- | 分隔“分支”（或选项）
- * 0个或多个其之前的匹配的的顺序序列
- + 一个或多个其之前的匹配的的顺序序列
- ? 0个或一个其之前的匹配
- .
- 匹配任何单个字符
- ^ 匹配开头的值
- \$ 匹配结尾的值
- [] 包含某个字符选择的设置
- () 包含某个子表达式——无论匹配的是_{ }列表属性里的哪个子表达式
- A 能够列入可匹配的任意字符
- \d 匹配一个数字字符。等价于 [0-9]
- \D 匹配一个非数字字符。等价于 [^0-9]
- \S 匹配任何非空白字符
- \s 匹配任何空白字符，包括空格、制表符、换页符等等