FME 在数据处理中的应用

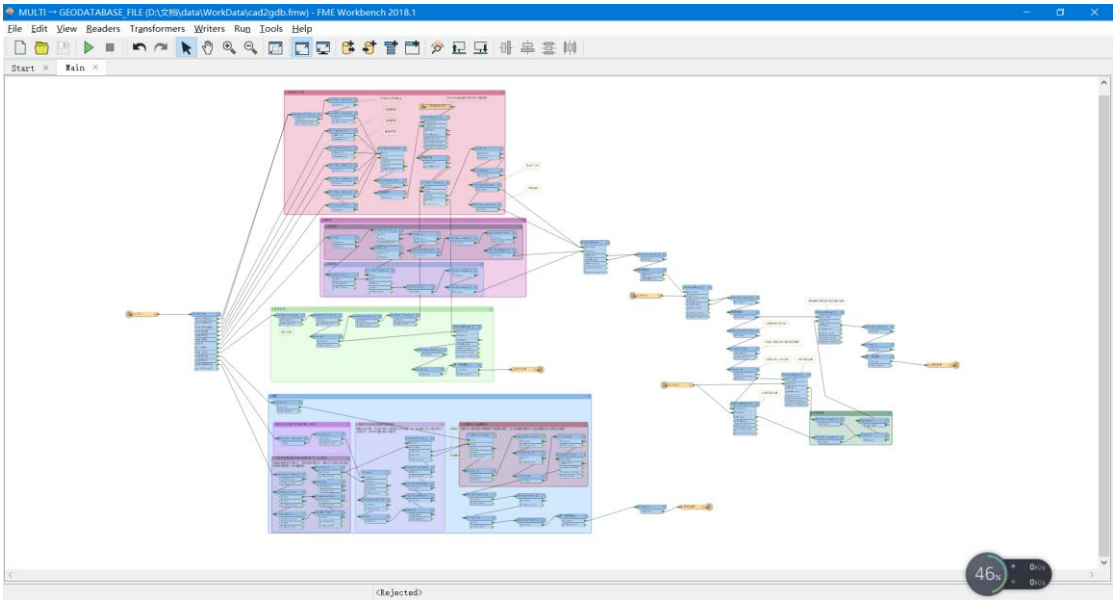
对各种数据格式的支持，再加上功能强大的转换器，友好的界面与先进的工作流设计，方便的第三方功能扩展，使得 FME 成为了数据处理与分析的强大工具。我也有很多的朋友在日常的工作中在使用 FME，用 FME 出图，用 FME 采集数据，用 FME 整理文件，甚至有人用 FME 做后台服务。似乎，能用 FME 做什么，完全取决于用户的想象力。而我将从自己应用 FME 的角度，介绍下 FME 在数据处理中的一些应用场景。

（一） 在空间数据处理中的应用

作为一个 GIS 行业从业者，当然免不了要处理各种空间数据。而 FME 在空间数据处理方面的能力，让这件事情变得没那么难。

1）CAD 数据的入库

CAD 数据的入库，大概是所有传统测绘领域企业头疼的事情。遥想自己当年在国内某测绘企业工作的时候，公司研发了专门的软件来进行数据入库，然，效率低下！后来我接触到了 FME，我基于 FME 做了 CAD 转 ArcGIS 地理数据库的工具，后经不断改进，到现在形成了如图一所示的模板：



图（一）

一个简单格式转换的模板，让当年的我加了薪。而随着对 FME 的学习与了解的深入，让我体会到了数据处理与分析的乐趣。

在不断改进中，模板中功能模块也越来越丰富，像空间分析：各种矢量的空间要素，或融合、或叠加、或裁剪、或分割，各种分析在 FME 中都能被支持；像属性的处理：在模板中，字段的增删改、有效信息的提取、字段间的数学运算、多源数据的属性融合，各种对属性的处理都能被实现；像自然语言的处理，各种文本的交换格式，像 json、XML 等格式都能在 FME 中方便的解析。

2）空间数据的处理与分析

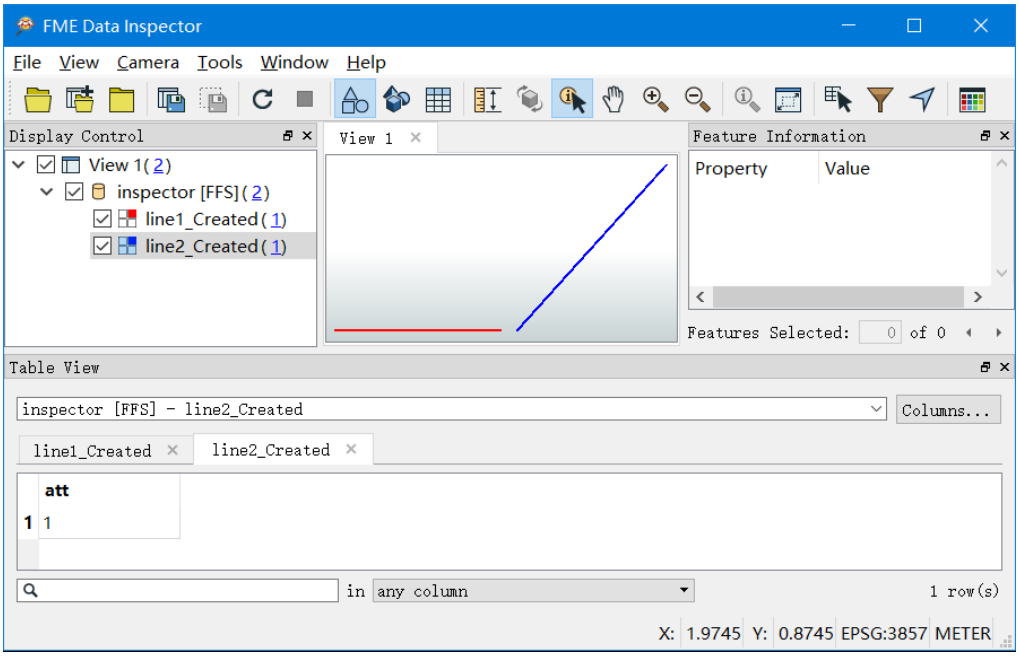
地理信息的核心是空间分析，透过空间分析，可以让空间数据发挥出自己的价值。然而现实是我们平常接触到的数据质量参差不齐，各个分析平台对数据都有一定的要求，这就导致了大多数时候拿到的数据不能用来做分析。所以在数据分析之前，一般都有很多的数据处理工作要做。

数据处理能做的很多，在这里我举一个 CAD-GIS 转换中的例子：

数据处理

连接断开的道路：

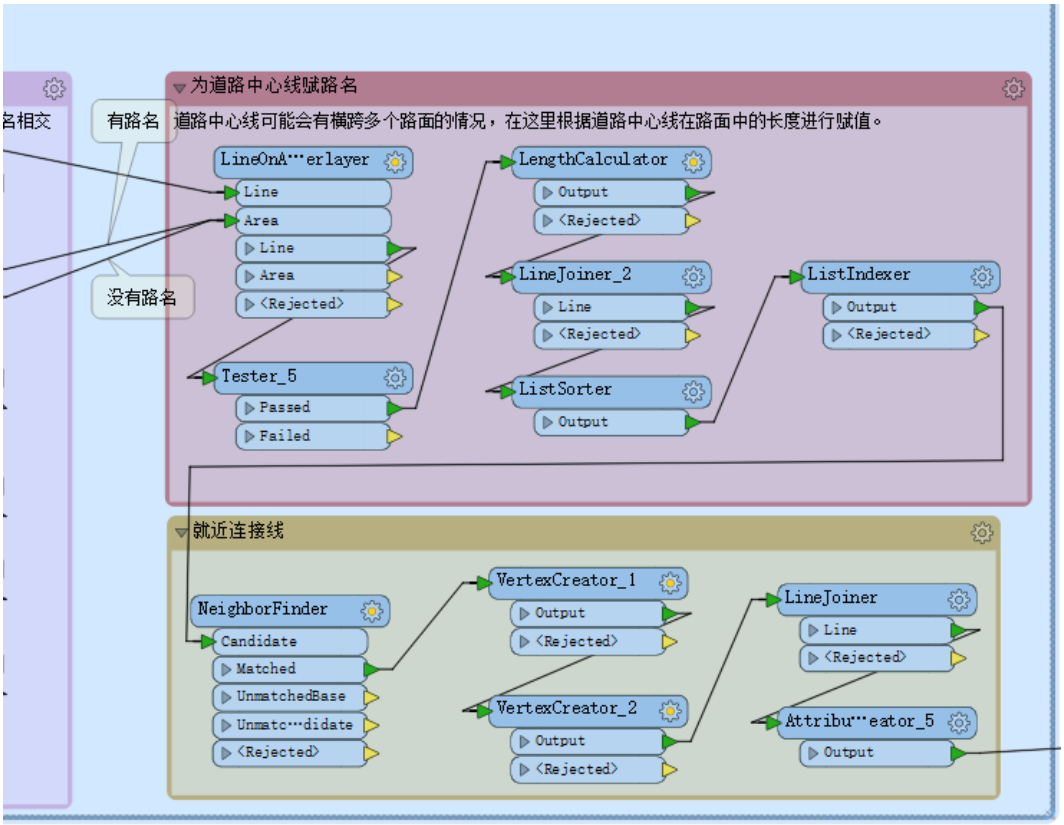
在 CAD 数据入库中，要素与要素之间经常会有微小的缝隙。如图二所示：



图（二）

这种数据之间存在缝隙的情况，在数据处理中非常的常见，当然，我也写过类似的文章。针对这种情况，解决方法也有很多，我介绍其中的一种，来展示 FME 在数据处理中有多么的灵活。

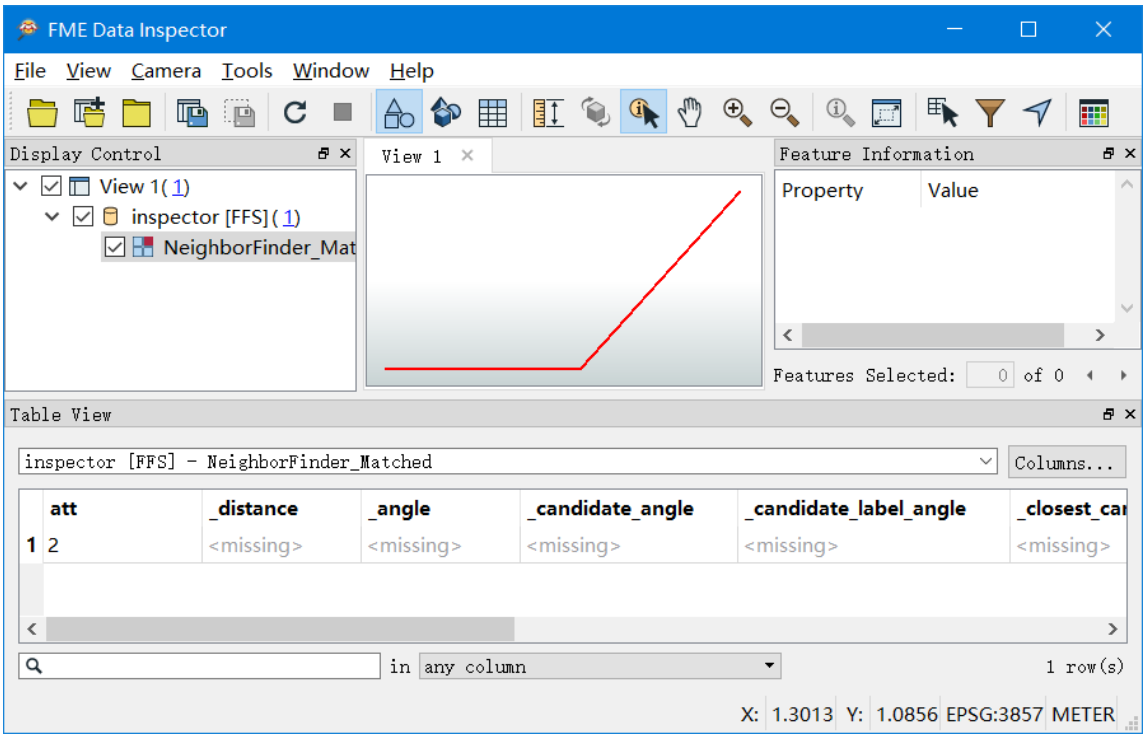
对这种数据，我采取了如图三所示的处理过程：



图（三）

将所有线连接起来：

使用 LineJoiner 转换器将所有的线连接起来后，结果如图四所示：



图（四）

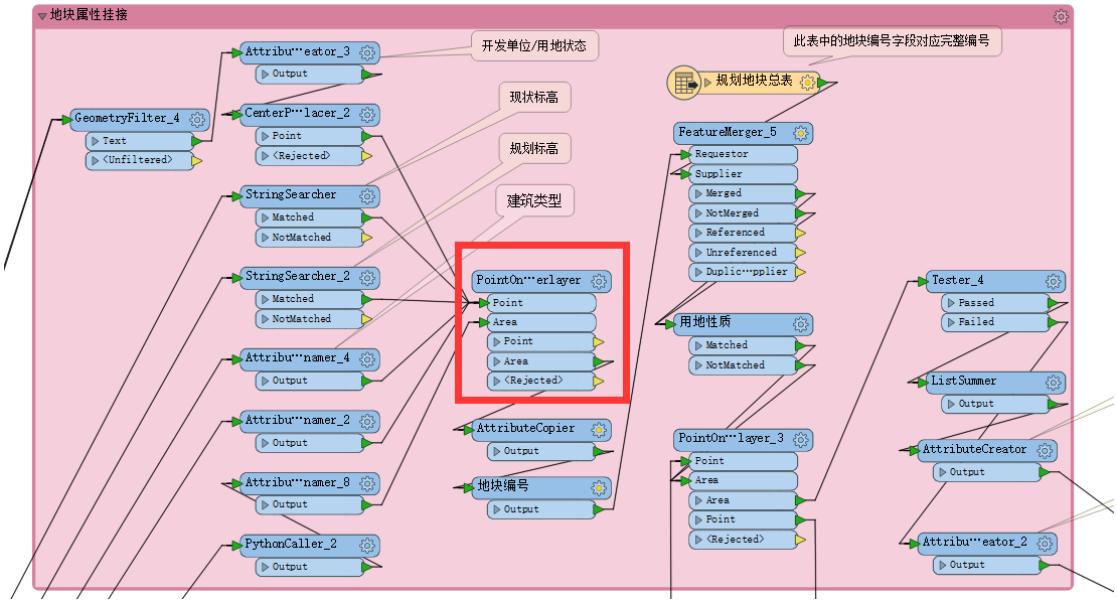
到这里为止我们已经成功的将中间有缝隙的线连接起来了，在处理过程中我们可以看到 FME 是多么的灵活。就这个小问题来说，解决方案就有好多套。同样的我们还有非常多的处理可以在 FME 中完成，比如：点与面的拓扑，线与面的拓扑，面之间的拓扑问题，有很多的拓扑问题在规则允许的情况下都可以在 FME 中自动的处理掉；点、线、面之间的相互转换；从二维数据生成三维；三维数据提取建筑物轮廓等一系列问题，都可以在 FME 中进行处理。就像本篇开始时说的那样，能用 FME 做什么，完全取决于使用者的想象力。

空间分析

与其他平台相比在 FME 中做空间分析是一件非常简单的事情，FME 的工作流的方式与转换器的出现，可以让使用者专注于空间分析本身，而不是代码之类的东西。有时候，处理与分析是分开进行的，

我们使用一个平台进行数据的处理，而分析就放在另一个平台中进行。但在 FME 中，只要你愿意，完全可以将处理的结果直接在模板中进行分析，同样的，我举一个 CAD-GIS 转换中的例子：

如图五所示：我们将各个注记层中的文字进行处理（处理过程就不展开了），处理后的结果直接按照空间关系赋值给地块，在进行点面叠加之前，我将所有不同的点层要素所带的属性命名到其对应的字段，这样处理之后，一个面可以通过一次空间分析获取来自多个点层的属性。



图（五）

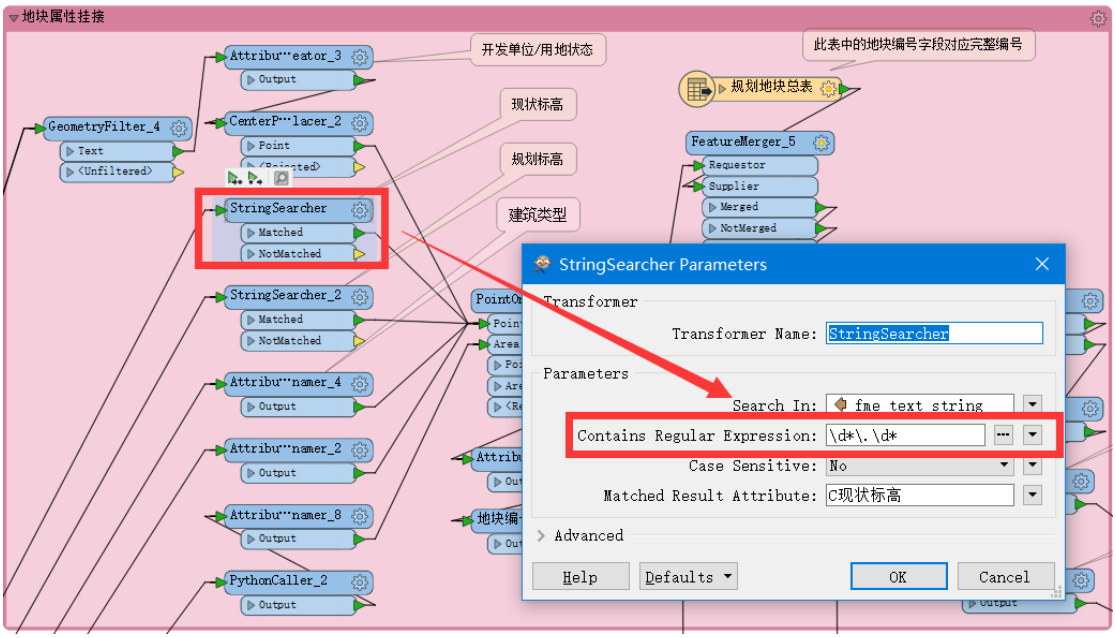
在 FME 中，空间分析类转换器有很多，通过这些转换器，我们可以很方便的实现属性的关联，几何之间的裁剪，缓冲区的生成，甚至是最短路径等空间分析，都有相关的转换器支持；不仅如此，通过 FME 还可以方便的实现矢量与栅格数据的叠加分析，通过 FME 的栅格相关的转换器，可以方便的实现从栅格中提取相关的信息进行分析，往年的博客大赛，我也写过相关的文章，这里不再赘述。

（二） 在自然语言处理中的应用

自然语言的处理与分析听起来是个很高端的行业，动不动就机器学习、人工智能。但是你知道吗？FME 在机器学习中也能占上一席之地，这其中的很大一部分原因，是因为 FME 对自然语言处理的支持与友好的扩展性。并且与代码相比，FME 作为 ETL 在界面上真的是非常友好。基于这个原因，在前不久与某国家重点 GIS 实验室合作，要对近 200 万条数据用机器学习的方式进行分词时，我就选择了 FME 来做监督数据。

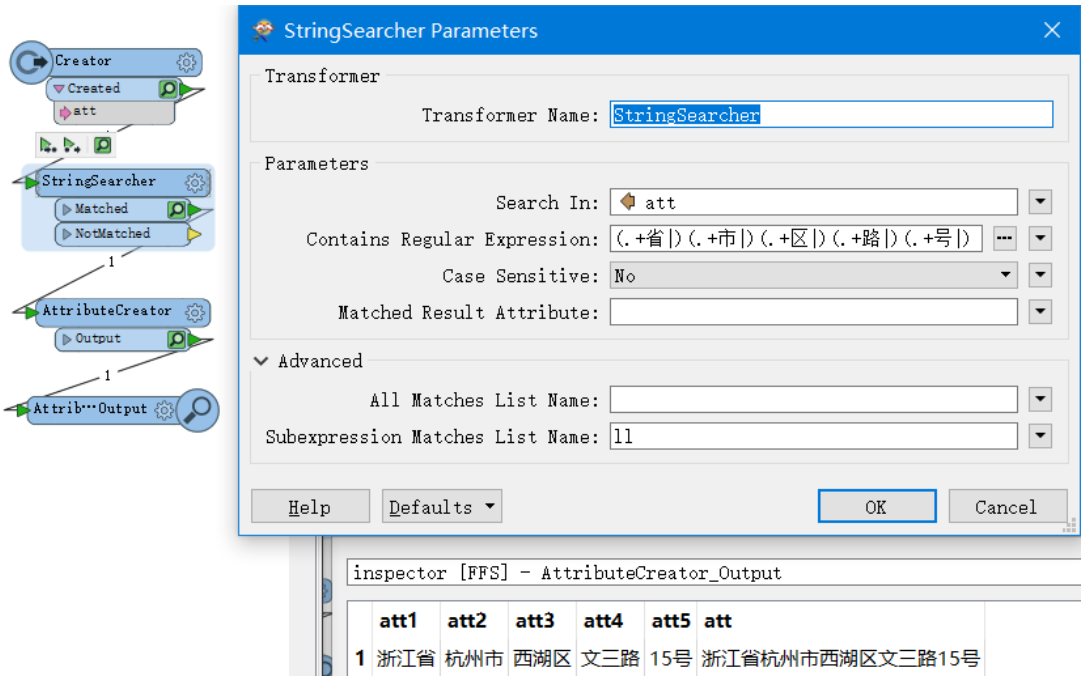
1）对正则的支持

在 FME 中使用正则，是一件非常简单的事情。以 CAD-GIS 的模板为例，使用正则可以方便的提取出数字，如图六所示：



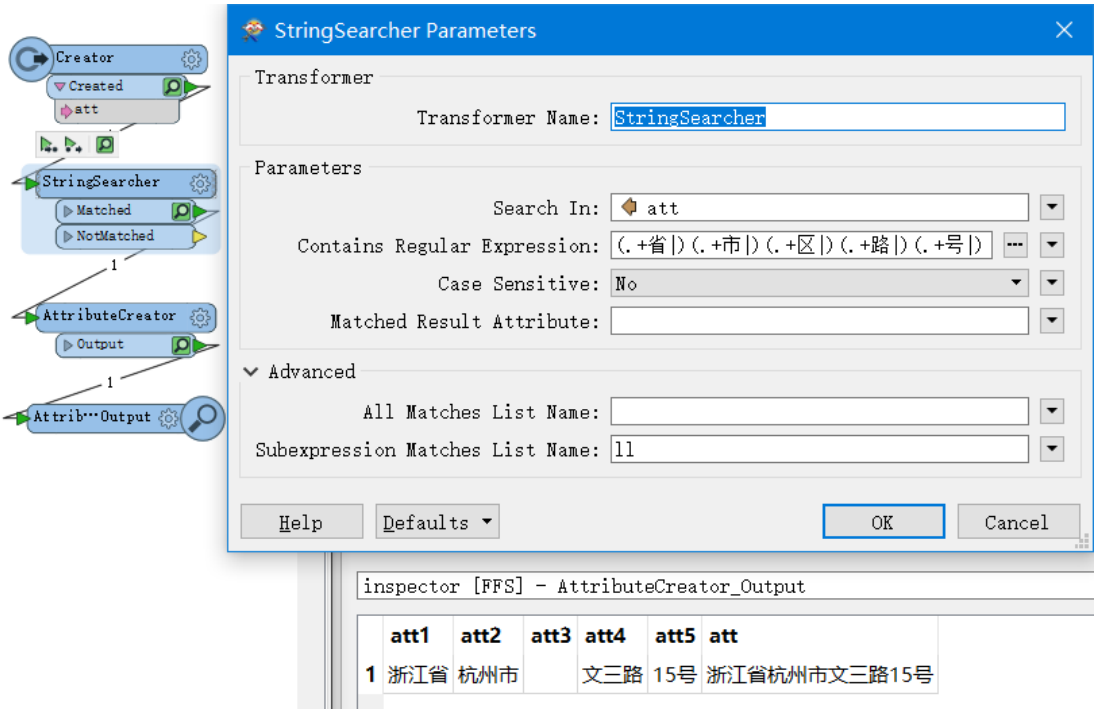
图（六）

只提取数字的话，不能满足数据处理的需求。大多时候需要对地址描述进行拆分，而这些拆分在 FME 中处理起来也非常方便，如图七所示：



图（七）

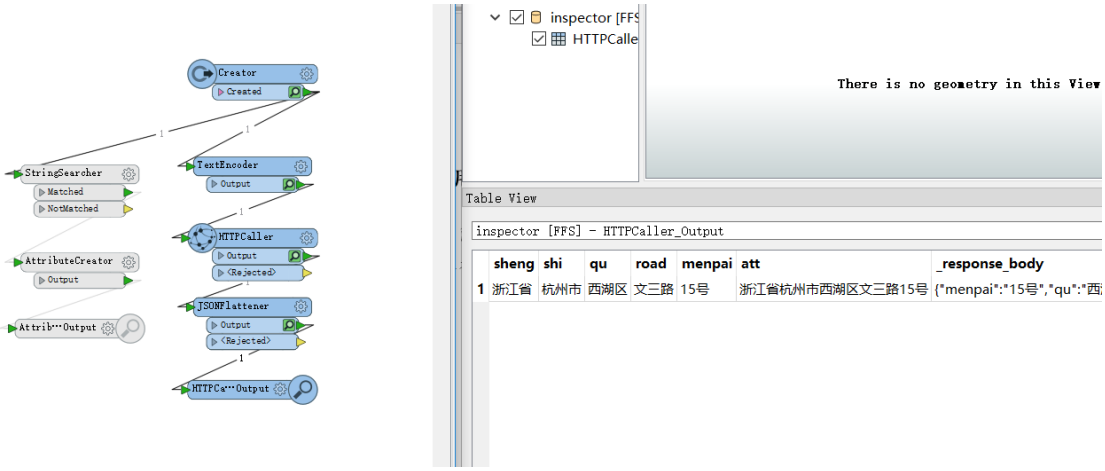
即使有些地址级别为空，也不会影响，拆分一样可以得到正确结果：



图（八）

2）调用服务

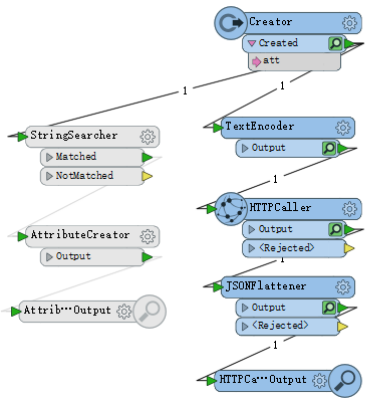
对于一些简单的字符串拆分，正则就够用了。但对于一些复杂的地址组合来讲，完全通过写正则进行拆分显然不够现实，这时候可以通过调用外部服务来进行字符串的解析。如图九所示：



图（九）

地址拆分服务与FME之间通过 json 传递数据，返回的结果非常友好，容易解析，并且可以根据自己需求，灵活的搭建自己的服务，比如说，还是刚才的地址，如果其中的地址搭配顺序变了，用正则写起来可能很麻烦，但是调用服务就完全没这回事了。如图十所示，att 字段里

包含的地址顺序全都乱了，而通过地址服务的解析，依然能够得到正确的结果：



图（十）

（三） 结语

到这这篇文章为止，我大概用了五年 FME 了。五年间，从不了解 FME，到 FME 成为工作中最亲密的伙伴，期间结识了很多数据处理方面的大牛，遇到问题时总有这些高手指点。现在我学会了如何使用 FME 等工具去解决问题，如果我在工作中遇到了要解决的问题，首选的技术方案就是 FME。

简单直白点：我喜欢 FME，喜欢敲击键盘出现的转换器，喜欢在软件中流动的数据，也喜欢与朋友探讨平台使用中遇到的问题。

最后谢谢安图，每年举办这种有意义的活动！