
FME 在地理数据处理中的应用

1 前言

下载模板兴趣版的每一期作业，打开模板似乎每个人的作业中用到的转换器都不一样，但却都能很好的达到兴趣班作业的处理要求。FME 平台给与我们足够的自由，让我们可以通过对转换器进行适当的组合来解决工作中的实际问题。本文将通过几个示例进行阐述，请随我一起来探索 FME，让我们一起体会站在大神肩膀上解决问题的乐趣！

本文第二部分将介绍几个转换器的简单应用，第三部分通过一个实例来介绍 FME 在 CAD 数据转换 ArcGIS GDB 方面的应用。

2 简单示例

2.1 检查地址码跳号

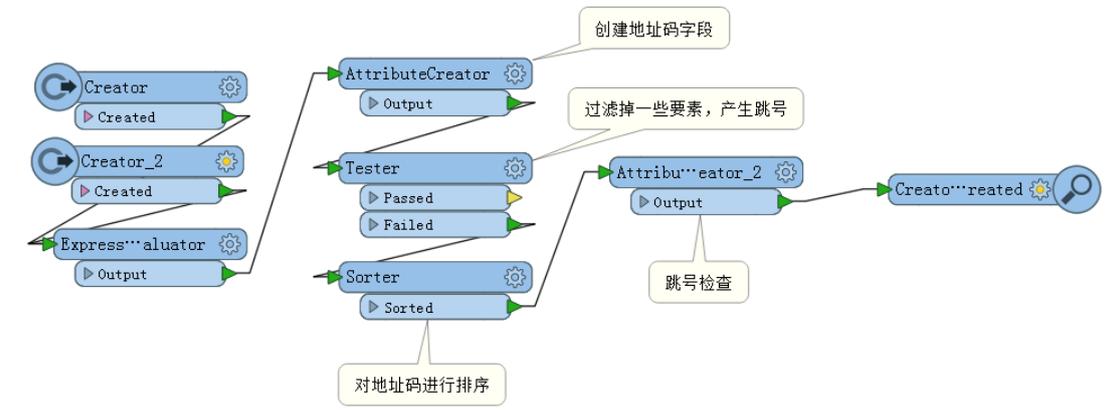
用途：

查找数据中不连续的要素（使用 AttributeCreator 转换器可以获取到临近要素的属性信息，在需要比较属性信息的应用场景下，这是非常使用的一个功能）。

思路：

使用 AttributeCreator 转换器的临近要素功能来实现核心功能；

示例模板：



具体实现方式：

- 对包含地址码数字的要素进行排序，排序方式选择升序；
- 启用临近要素，对每个要素保留前一个要素的属性信息；
- 使用条件赋值对数据进行检查；

2.2 使用正则表达式

使用正则表达式可以对文本进行清洗、过滤、结构化以及信息的提取，通过使用正则表达式技术，可以做到很多事情，其中有些是很酷的。在本文第三部分中的数据转换中，图框中的信息区分就是使用正则表达式技术完成的。

- 过滤

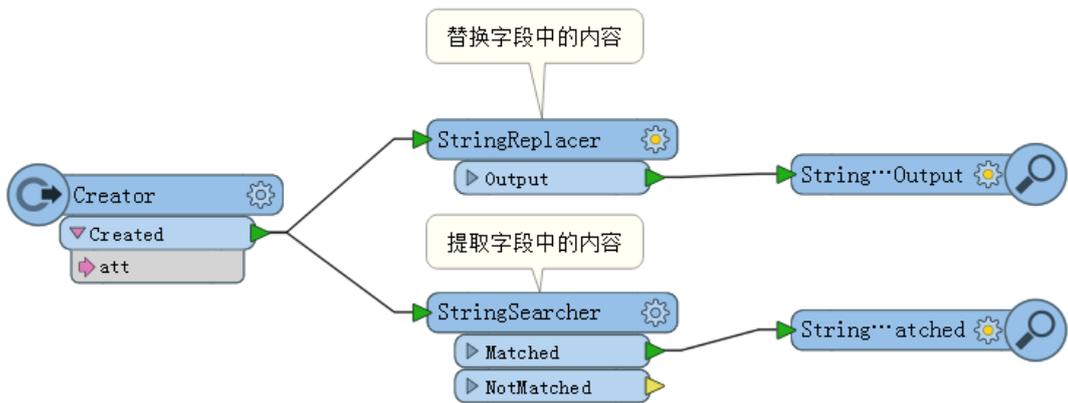
使用 Tester 可以对数据进行过滤，我们可以设置规则来进行一般的过滤，也可以通过对规则的组合来满足复杂的过滤。我们可以使用 Tester 转换器来做更酷的事情。

举个例子：如果要筛选“我是 FME 中国技术交流群里的青，哈哈哈！”与“我是青，我在 FME 中国技术交流群，哦哦哦！”这样的两个字符串，并且我们将筛选规则确定为：字符串中同时含有 FME 和青这两个字符串，且 FME 在青之前；我们只需将规则写为‘FME.*青’即可对数据进行过滤。

b) 清洗

使用正则表达式可以轻松完成一些工作中的数据清洗工作，通过恰当的使用正则表达式，可以在很大程度上简化我们的工作。数据清洗其实就是将数据中不需要的部分去掉，保留有用部分；换个角度来看，其实提取出有用部分也是可以的。在 FME 中能完成这项工作的转换器有很多，在这里我着重介绍下

StringSearcher 转换器与 StringReplacer 转换器：



StringSearcher 转换器可以从数据中提取需要的部分，举个例子：假如有这样一个字符串：我最喜欢的 ETL 工具是 'FME'，并且我们将提取规则确定为：只提取出单引号之间的内容。要实现这样的功能只需要写这么一个表达式： $(?<=')\.(?<=')$ 。这个表达式的意思是：匹配引号中间的内容，但结果中并不会包含引号。

StringReplacer 转换器可以从数据中替换掉不需要的部分，举个例子：还是那个字符串，规则变更为替换掉单引号外面的部分，只保留单引号内的字符。要实现这样的功能，需要这样一个表达式： $.*(.+)'.*$ 。表达式的意思是，捕获引号中的字符到分组 1。

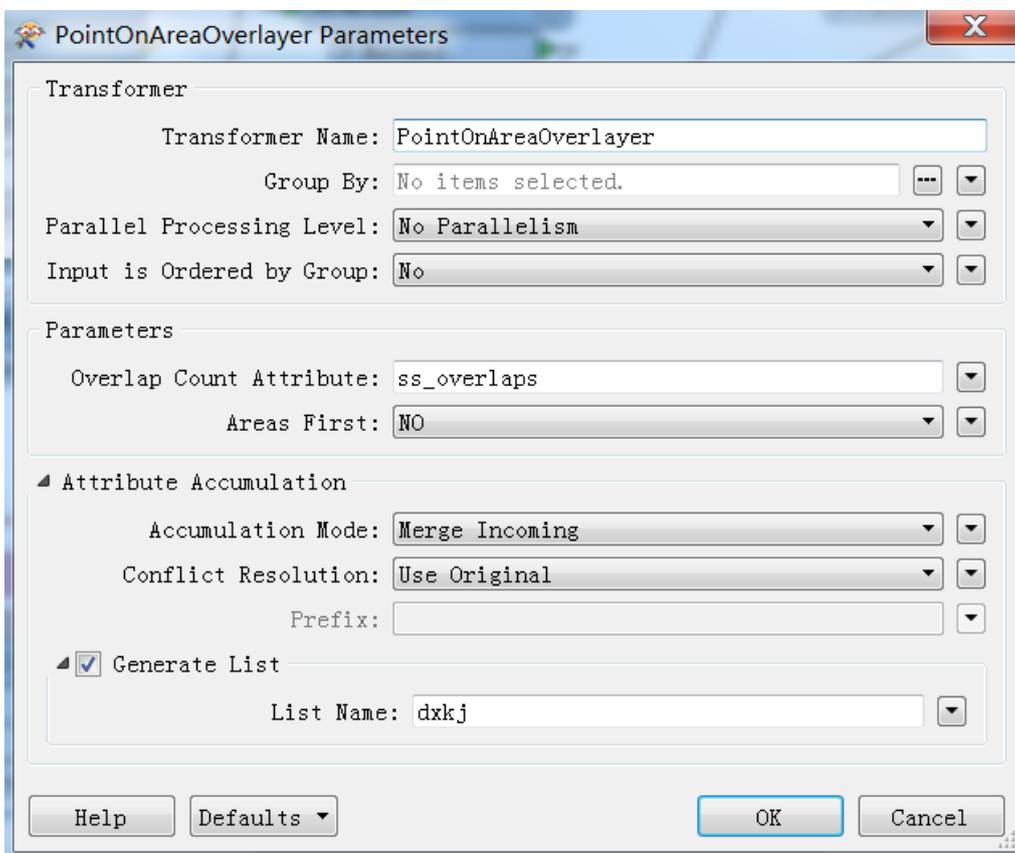
2.3 空间挂接

空间的挂接在 FME 中有许多转换器可以使用，基本上所有基于空间运算的转换器都可以传递属性，较为典型的有 **Clipper**、**PointOnAreaOverlay**、**SpatialFilter**、**LineOnAreaOverlay**

等转换器，需要注意的是要素在进入这些转换器前字段名称是否有冲突，如果有冲突则需要设置是否加前缀。

现已 PointOnAreaOverlayer 转换器为例进行说明。

PointOnAreaOverlayer 转换器接受点与面，经过这个转换器的处理，点要素会获取所在面要素的属性信息，相应的面要素也会获取对应的点要素属性信息。

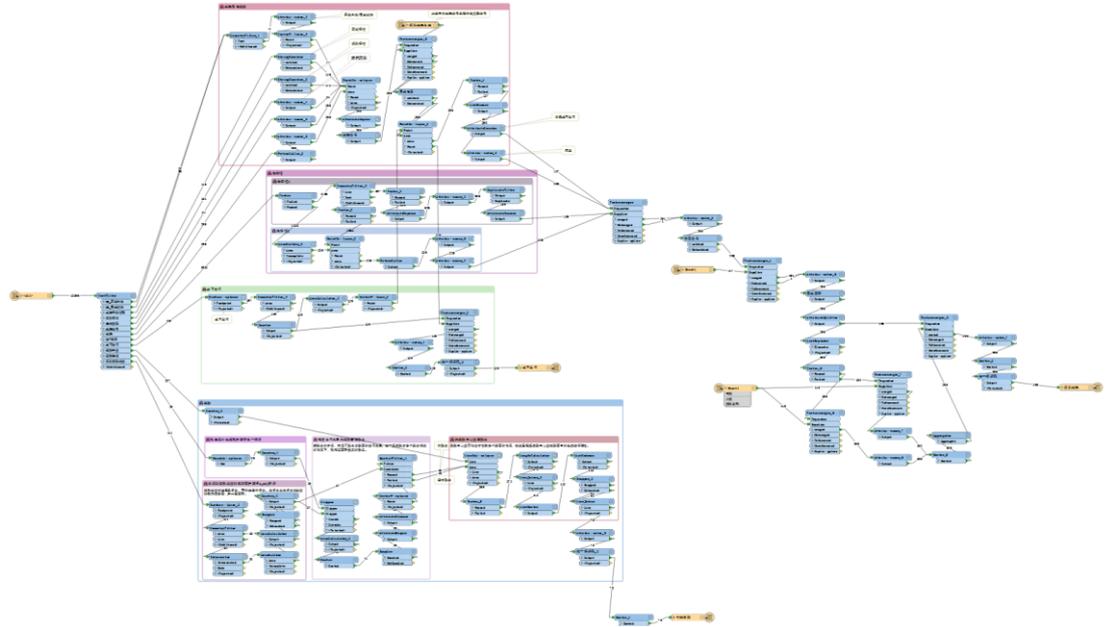


此图来自于本文第三部分所述模板

3 数据转换

数据的转换，是我在使用 FME 时一个主要的方向。一般情况下通常并不是简单的格式转换，一般情况下都需要做属性的挂接

(基于空间或者属性字段), 初步的检查 (比如文档第一部分中的检查地址码跳号), 数据质检, 数据融合等许多的数据综合处理。在本部分中, 将通过一个较为典型的例子来进行简单说明, 模板整体截图如下:



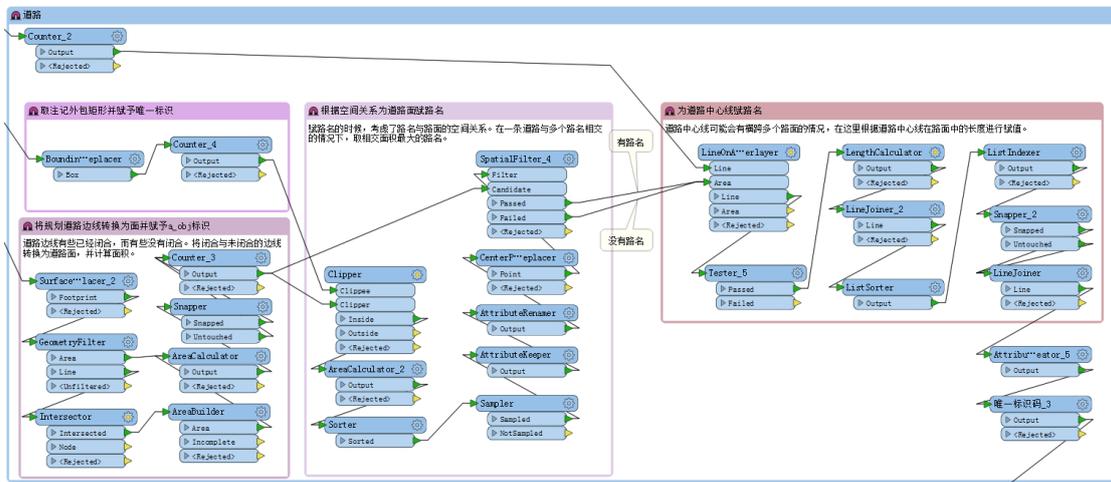
概述: 转换任务为 CAD 转 ArcGIS GDB 地理空间数据库, 首先要读入数据, 使用 TestFilter 对数据进行分流, 分出需要处理的图层, 以及这些层挂接属性所需要的数据层 (本文将着重介绍地块层、道路层、地下空间层), 经过分流的数据可以分别进行处理, 然后使用基于空间或者属性的方法进行相关的挂接。中间需要根据要求读入 Excel 相关的表进行属性的挂接与匹配, 最后将经过处理后的符合要求的数据进行输出。

由于模板整体比较复杂, 本文由于篇幅限制难以深入介绍, 所以将着重介绍以下几个方面:

3.2 道路层

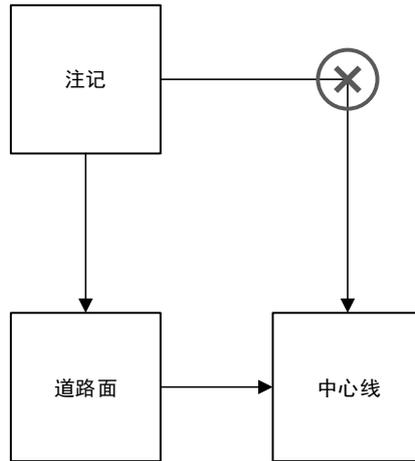
道路层需要挂接道路注记所蕴含的道路名称，并计算道路长度等信息，难点在于挂接道路名称，以下将着重介绍道路名称图层。

为道路赋名称也算是属性挂接的一种，但道路赋路名需要考虑的因素较多，整体截图：

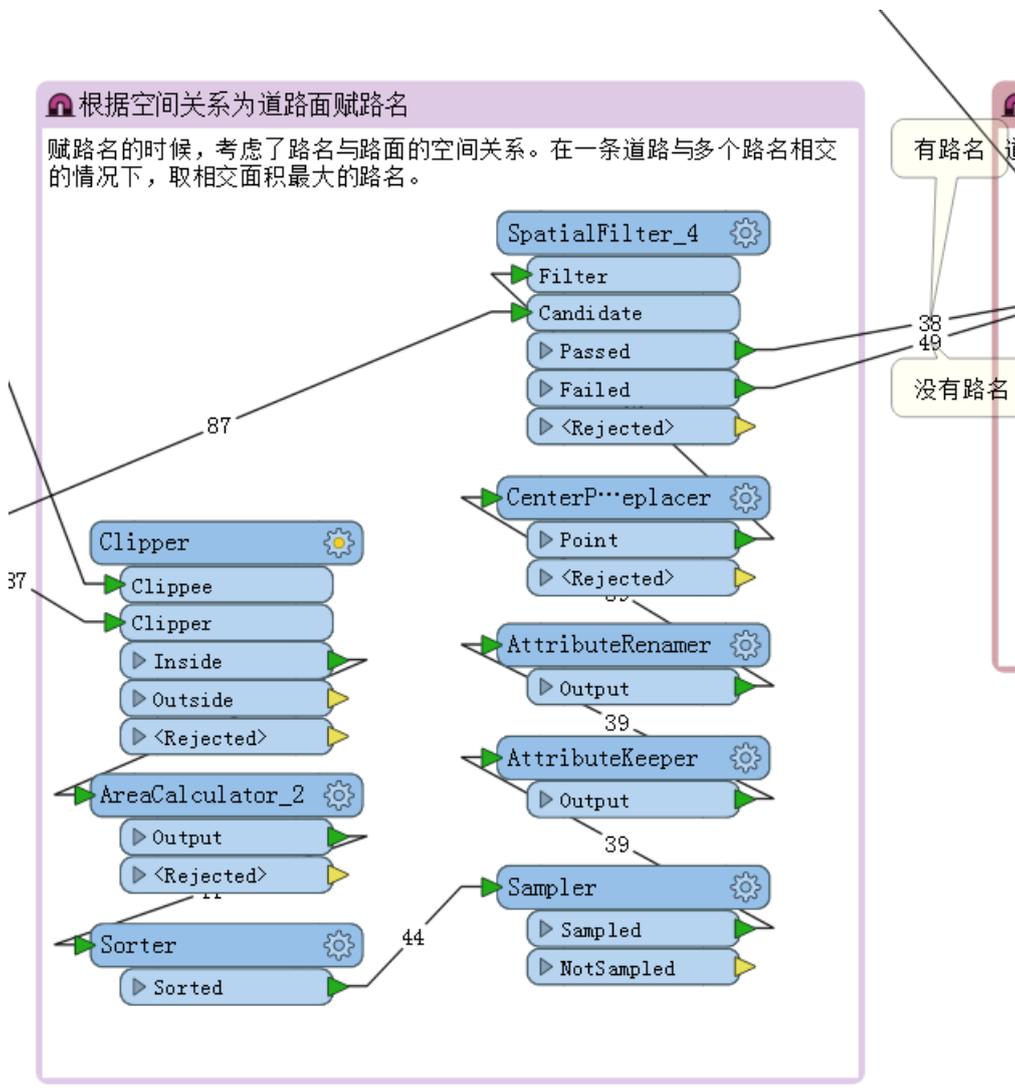


a) 为路面赋路名

采取如下图所示这种方式的原因是因为道路中心线可能会多个注记相交，不好去判断他们之间的关系，使用道路面就可以通过注记在道路面内的面积这种关系来判断注记与路面的关系。

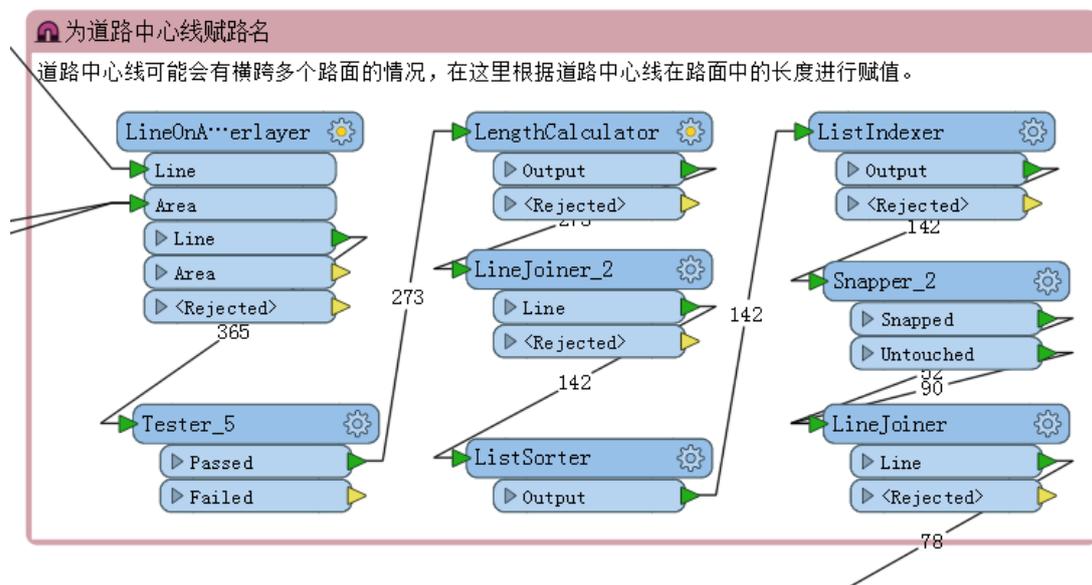


使用到的转换器截图如下所示



b) 为道路中心线赋路名

所用转换器截图：

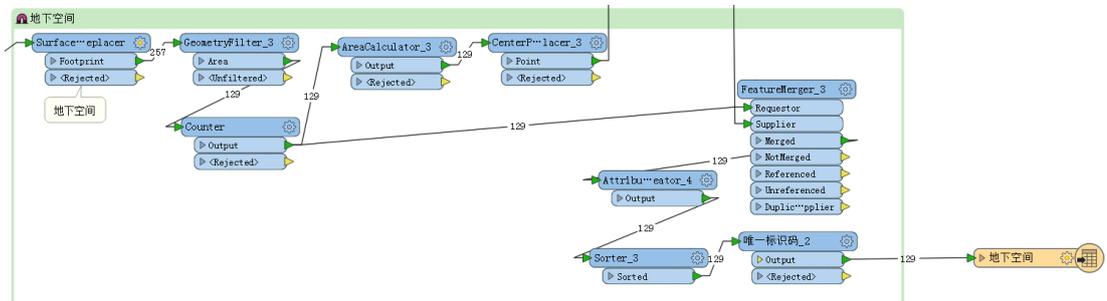


具体实现步骤为：

1. 根据线面的叠加关系筛选出面内线；
2. 计算每条线段落落在道路边线面中的长度；
3. 根据中心线唯一标识将中心线连接起来；
4. 根据步骤 2 中计算的长度进行排序；
5. 赋予中心线道路名称与道路面唯一标识（ListIndexer 转换器）；
6. 对道路中心线进行捕捉（中心线之间也会有微小距离，这点距离会影响后面的同一面内线连接，经过捕捉就可以规避这个问题）；
7. 根据道路面的唯一标识，将道路中心线连接为长线。

3.3 地下空间层

地下空间相对地块比较的简单，做法如下：



首先将地下空间的 Surface 投影在平面上，然后再进行几何过滤筛选出面状要素，赋予唯一编号，再计算面状要素的面积（地块需要挂接这个面积），然后对要素挂接来自地块的属性，并输出数据。

3.4 总结

数据在读取进来之后可以做各种处理，在本例中读取的是 cad 格式的数据，FME 支持的数据格式有好几百种，所有数据读取进来都成了 FME 可处理的要素。作为地信从业者，我接触到的大部分数据都是矢量数据，对矢量数据我们可以做空间相关的操作，比如点转线转面、面转点转线，这些操作在 FME 平台中都非常容易实现；我们还可以为矢量数据挂接来自其他数据源的属性（本文中的例子挂接了来自 Excel 的属性）。FME 对矢量数据中存在的一些质量问题也可以很好的修复，本文第三部分中的道路图层在

处理的时候就应用了捕捉，避免相邻线要素由于微小缝隙不能连接起来的问题，保证后续数据处理正确性。

4 结语

在工作中使用 FME 也有三四年了，在这中间随着对 FME 的了解越来越深入，能用 FME 来做的工作也越来越多，实现的功能也越来越强大，能达到的效果也是越来越好。在这期间，遇到了许多的问题与挫折，解决问题的时候结识了很多业界大牛，克服挫折的过程中也收获了许多。一路走来，从开始的看到那么多转换器一脸茫然，到现在能够自由的使用转换器构建工作空间让数据在模板中自由流动，FME 已经成为了我工作中最好的选择。

FME 作为一个可扩展的平台，能实现的功能是很多的，只要你有足够的想象力，这个平台都能帮你实现。本文所展示的几个例子只是九牛一毛，如果想要学习更多的关于 FME 的学习资料可以浏览 [FME 博客](#)、[FME 十分钟视频](#)，还可以加入 FME 中国技术交流群(群号 43814136 已满员),FME 中国技术交流群 2(438653305)。非常感谢世纪安图提供的平台。